Efficient GMM and Weighting Matrix under Misspecification

Byunghoon Kang*

October 15, 2026†

Abstract

This paper develops efficient GMM estimation under misspecified moment conditions. The influence function of the standard GMM estimator under misspecification depends on both the moment conditions and their Jacobian, motivating a new class of estimators based on augmented, recentered moment conditions. Standard GMM is a special case and generally suboptimal. By optimally weighting the augmented system, we obtain a misspecification-efficient (ME) estimator with the smallest asymptotic variance for the same GMM estimand. In linear models, its asymptotic variance takes the textbook form $(G'W^*G)^{-1}$, where W^* is the inverse variance of the residualized moments after projection on the Jacobian G. We propose a feasible bootstrap estimator, which is a misspecification-robust and efficient version of the Hall and Horowitz (1996) bootstrap GMM estimator. Finally, we establish uniform local asymptotic minimax bounds over a class of weighting matrices. We illustrate the proposed methods in simulation and two empirical examples.

Keywords: Generalized method of moments, misspecification, pseudo-true value, semiparametric efficiency, augmented moment conditions, bootstrap, instrumental variables

*Department of Economics, Lancaster University Management School. Email: b.kang1@lancaster.ac.uk.

†We thank seminar participants at Seoul National University and the James MacKinnon Conference at Aarhus University for helpful comments and suggestions.

1 Introduction

The generalized method of moments (GMM) offers a robust framework for estimation and inference, and is widely used in applied econometrics. When the model is over-identified, where the number of available moment conditions exceeds the number of parameters, the GMM estimator depends on the choice of weighting matrix. Under correct specification, where a unique true parameter satisfies the population moment conditions exactly equal to zero, the efficient weighting matrix is the inverse of the variance of the moments, which achieves the smallest asymptotic variance among GMM estimators (Hansen (1982), Chamberlain (1987)).

However, economic models are inherently structural approximations, and the overidentifying restriction test is often rejected due to various forms of misspecification; candidate instruments may be invalid, lag structures or functional forms may be misspecified or treatment effect heterogeneity implies that no parameter satisfies all moment equations exactly. A growing literature therefore studies the inference for the pseudo-true value, defined as the population minimizer of the objective function (e.g., White (1982), Hall and Inoue (2003), Müller (2013), Hansen and Lee (2021), and Andrews and Kwon (2024)).[1] Under misspecification, the asymptotic variance of the GMM estimator has different forms than the standard GMM variance estimator, and ignoring this can lead to inconsistent standard error estimation and asymptotically invalid confidence intervals for the pseudo-true values, which was also highlighted in Andrews, Chen, and Tecchio (2025).

More importantly, misspecification fundamentally changes the role of the weighting matrix, and has different implications than under the correct specification setup. The choice of the weighting matrix no longer merely affects the statistical efficiency of the estimator, but it also affects what is being estimated; different weighting matrices correspond to different target parameters.[2] This dual role of the weighting matrix under misspecification raises a number of questions that the standard GMM literature

[1] In the asset pricing literature, such as Kan and Robotti (2009) and Gospodinov, Kan, and Robotti (2014), the GMM pseudo-true value can be defined as a minimizer of the distance metric of Hansen and Jagannathan (1997).

[2] Pseudo-true values may differ from quantities of economic interest (Müller (2013), Andrews, Barnhard, and Carlson (2024), Andrews, Barahona, et al. (2025)). The main focus of this paper is on valid and efficient inference for a given GMM or two-stage least squares (TSLS) estimand, instead of identification and its causal interpretation; see, e.g., Kolesár (2013), Mogstad et al. (2021), Blandhol et al. (2025), and Słoczyński (2024) for the latter direction in the linear instrumental variables (IV) model under treatment effect heterogeneity.

does not directly address. Is the "standard" optimal weight, defined under correct specification, still efficient under misspecification? Does any other class of estimators exist with a smaller asymptotic variance than the standard GMM under misspecification for the same target? How should we think about asymptotic efficiency across different weighting matrices when each weighting matrix targets a different estimand? This paper addresses these challenging questions and provides theoretical and some empirical guidance for researchers using overidentified GMM models.

First, for a fixed weighting matrix W and the associated GMM estimand (pseudo-true value) θ_W, we observe that the influence function of the "standard" GMM estimator under misspecification can be expressed as a linear combination of augmented moment conditions, which stack the original moment functions and their Jacobian, with recentering. This motivates a new class of M-estimator based on the augmented moments that includes the "standard" misspecified-GMM as a special case. We then define the "misspecification-efficient" (ME) estimator as the estimator with the smallest asymptotic variance within this class, which is consistent for the *same* GMM pseudo-true value. By applying the optimal weighting, derived from the inverse of the variance matrix of the augmented moments, the proposed ME estimator captures the correlation between the moments and the Jacobian and extracts additional identifying information from the moment misspecification, whereas the standard GMM is suboptimal with particular linear combination of the available moments dictated by the usual GMM first-order condition.

The ME estimator is not directly feasible because it requires the population means of the original moments and the Jacobian for recentering, i.e., the degree of misspecification and the identification strength. However, our asymptotic result of the ME estimator is constructive in the sense that the variance of the ME estimator can be easily estimated and reported as an oracle efficiency bound; showing researchers what could be gained if these quantities were known. Under the linear model, the bound is $V_{ME}(W) = (G'W^*G)^{-1}$, where G is the Jacobian and $W^* = \Sigma_{11,2}^{-1}$ is the inverse of the variance of residualized moments after projection on the Jacobian, instead of the conventional efficient weighting matrix $W_0 = \Sigma_{11}^{-1}$ that is the inverse of the variance of the original moments. This bound is invariant to W under the linear model; different W target different θ_W, but do not affect how precisely they can be estimated. It also coincides with the textbook "efficient GMM" formula $(G'W^*G)^{-1}$, which offers a different angle on the well-known downward bias of the efficient-GMM

standard errors (Windmeijer (2005)); the formula is valid for the oracle ME estimator, but too "small" for the standard GMM estimator under misspecification when $W^* \approx W_0$, even in large samples. We therefore recommend reporting $V_{ME}(W)$ alongside the misspecification-robust standard errors (Andrews, Chen, and Tecchio (2025)) as a transparent efficiency measure that does not require committing to a specific W. Under homoskedasticity, the ratio of the GMM variance to the ME bound is approximately determined by the moment-Jacobian correlation and the J statistic scaled by the identification strength (see Remark 1). In general nonlinear models, $V_{ME}(W)$ varies with W.

Second, we propose two feasible bootstrap procedures. The bootstrap ME-GMM estimator is based on the GMM criterion function that recenters both the original moments and the Jacobian at their sample analogs evaluated at a preliminary GMM estimate, and uses the efficient weighting matrix of the augmented moments. We show that its bootstrap distribution approximates that of the oracle ME estimator, and it can be used to estimate the efficiency bounds. The proposed estimator can be viewed as a misspecification-robust and efficient version of the Hall and Horowitz (1996) recentered bootstrap, which is only valid under correct specification. We also consider the "double recentered" (DR) bootstrap, which jointly perturbs the moment conditions and the Jacobian as in the first-order condition of the misspecified GMM, and show that it mimics the limiting distribution of the standard GMM estimator regardless of model specification. Under correct specification, the additional Jacobian recentering is a higher-order correction and the DR bootstrap nests the original Hall and Horowitz (1996) bootstrap, and under misspecification it correctly approximates the Jacobian variation term that the Hall and Horowitz (1996) bootstrap omits. We also consider a repeated sample-splitting ME estimator in the spirit of Angrist and Krueger (1995) in Supplemental Appendix C.

Third, we develop uniform semiparametric efficiency bounds over a class of weighting matrices $W \in \mathcal{W}$. The notion of semiparametric efficiency needs to be defined differently under (global) misspecification, because θ_W changes with W, whereas the semiparametric efficiency framework typically assumes a single true parameter. We view each W as a different candidate model/design, and do not aim to choose an optimal W. For example, a more precisely estimated LATE (local average treatment effect) for one subpopulation is not necessarily better than a noisier LATE for another; they may answer different causal questions. We first derive the efficiency bound

for θ_W with a fixed W, and then establish uniform local asymptotic minimax bounds over $W \in \mathcal{W}$. These bounds are useful both for researchers who target a specific θ_W (e.g., TSLS), and for those who are agnostic about W, for example, when comparing one-step, two-step, and iterated GMM, each corresponds to a different pseudo-true value under misspecification.

We illustrate the proposed methods with Monte Carlo simulations and two empirical applications. The simulation results confirm that the oracle ME is meaningfully tighter, and the DR percentile CI delivers valid coverage, with lengths similar to the misspecification-robust Wald CI. In the returns to schooling examples of Card (1995) and Angrist and Krueger (1991), the ME efficiency bound is roughly 3-18% smaller than the misspecification-robust standard errors under the efficient weight $W^* = \Sigma_{11,2}^{-1}$ or the TSLS weight $(Z'Z)^{-1}$, and the gap is larger in Card (1995), where the moment-Jacobian correlation is larger. Overall, we find that the ME efficiency bound is most informative when (i) W is an efficient weight (W^* or W_0), rather than $W = I$ or other naive choices, and (ii) the identification strength is moderate, so that the Jacobian sampling variation is non-negligible.

Our paper complements a growing literature on sensitivity analysis and robust inference under misspecification (e.g., Andrews et al. (2017), Armstrong and Kolesár (2021), Bonhomme and Weidner (2022), Christensen and Connault (2023), and Armstrong et al. (2025)) and on the relationship between the pseudo-true and true values (e.g., Andrews, Barnhard, and Carlson (2024), Andrews, Barahona et al. (2025); see Andrews, Chen, and Tecchio (2025) for a review with recommendations for practitioners). While these papers focus on bias, sensitivity, or optimal inference, our paper analyses efficient estimation in misspecified GMM, and provides a feasible efficiency measure that researchers can report alongside robust standard errors. Our bootstrap procedures relate to the recentered GMM bootstrap (Hall and Horowitz (1996), Brown and Newey (2002)) and its misspecification-robust alternative (Lee (2014)), and to the use of Jacobian recentering and moment-Jacobian orthogonalization in the weak-identification literature (e.g., Kleibergen (2005), Dovonon and Gonçalves (2017), Lee and Liao (2018), Kleibergen and Zhan (2025)).

The remainder of the paper is organized as follows. Section 2 introduces the setup and Section 3 develops the ME estimator. Section 4 analyses the bootstrap procedures and Section 5 provides the uniform efficiency bounds. Section 6 reports the Monte Carlo evidence, Section 7 provides empirical results, and Section 8 concludes. The

Appendix includes all proofs of the results in the main paper. The Supplemental Appendix contains additional materials. Section A provides details on the linear IV example, Section B provides the worst-case variance bounds over misspecification and identification strength in the spirit of Conley, Hansen, and Rossi (2012). Sections C-F contain the split-sample ME estimator, additional simulations, the dynamic panel application of Arellano and Bond (1991), and the derivation of the decomposition of the ME efficiency gain, respectively.

1.1 Notation

$||A||$ denotes the spectral norm. $A \geq B$ denotes that $A - B$ is positive semidefinite for real symmetric matrices A and B. Let $o_p(\cdot)$ and $O_p(\cdot)$ denote the usual stochastic order symbols, and $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability, convergence in distribution, respectively. Superscript * denotes a probability or moment computed under the bootstrap distribution conditional on the original data set. Let $\xrightarrow{p^*}$ and $\xrightarrow{d^*}$ denote the convergence in probability, in probability, and the convergence in distribution, in probability, respectively. In addition, we write $\xi_n = o_{p^*}(1)$ if $\xi_n \xrightarrow{p^*} 0$ and $\xi_n = O_{p^*}(1)$ if ξ_n is bounded in probability, in probability.

2 The Setup

Suppose that we have data on observations $\{X_i\}_{i=1}^n$ randomly drawn from the unknown probability distribution F_0, with a data vector X. Let $g(X, \theta)$ be an $m \times 1$ vector of moment functions and θ be a $p \times 1$ parameter vector. The model is based on the moment conditions, and a popular estimator of θ in this setup is the generalized method of moments (GMM) estimator of Hansen (1982),

$$\widehat{\theta}_{GMM}(W) = \arg\min_{\theta} g_n(\theta)' W g_n(\theta) \tag{1}$$

where $g_n(\theta) = n^{-1} \sum_{i=1}^n g(X_i, \theta)$ and W is the weighting matrix. Under regularity conditions, the GMM estimator converges to the pseudo-true value θ_W, which is defined as a minimizer of the population objective function

$$\theta_W = \arg\min_{\theta} g(\theta)' W g(\theta) \tag{2}$$

where $g(\theta) = E[g_n(\theta)]$. Assuming global identification, θ_W satisfies the following just-identified equation from the first order condition (FOC) of (2);

$$G(\theta_W)'W g(\theta_W) = 0 \tag{3}$$

where $G(\theta) = E[\frac{\partial g(X_i,\theta)}{\partial \theta'}]$ is the population Jacobian. When the model is correctly specified, i.e., there exists a unique θ_0 that satisfies the original moment equations

$$g(\theta_0) = E[g_n(\theta_0)] = 0, \tag{4}$$

the pseudo-true value $\theta_W = \theta_0$ for all W, and it does not depend on W. Even if the moment condition is (globally) misspecified, i.e., $E[g_n(\theta)] \neq 0,\ \forall \theta \in \Theta$, the FOC condition (3) still holds. However, the GMM pseudo-true values generally depend on the choice of W under misspecification when the model is overidentified ($m > p$), as it satisfies the linear combinations of non-zero population moments $g(\theta)$ as in (3), so varying W changes the parameter being estimated.

In this paper, our objective is to provide valid inference procedures for the pseudo-true value θ_W. It is difficult to give general economic interpretation to the pseudo-true parameter values, and instead, researchers typically analyze and interpret them case-by-case. For example, in the linear IV model, moment misspecification can arise due to the treatment effect heterogeneity, and it is common to focus attention on the TSLS estimand, which corresponds to θ_W with $W = E[Z_i Z_i']^{-1}$ and the instrumental variables Z_i. In this setup, θ_W can be interpreted as a weighted LATE under appropriate assumptions; see Angrist and Imbens (1995).

The influence function of the "standard" GMM estimators under misspecification can be expressed in terms of augmented moment conditions, which include (i) the original moment restriction, (ii) their Jacobian, and (iii) with recentering. This result, while implicit in the form of the asymptotic variance of the misspecified-GMM in the literature, motivates the new class of estimators based on the augmented moments.

To see this, the GMM estimator has the following expression under misspecification with the standard regularity conditions (see the proof of Theorem 1 for details),

$$\sqrt{n}(\widehat{\theta}_{GMM}(W) - \theta_W) = -(A(W,\theta_W)\Gamma(\theta_W))^{-1} A(W,\theta_W) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\psi(X_i,\theta_W) - E[\psi(X_i,\theta_W)]\Big) + o_p(1), \tag{5}$$

where

$$\underbrace{\psi(X_i,\theta)}_{m(p+1)\times 1} = \begin{bmatrix} g(X_i,\theta) \\ vec[G(X_i,\theta)'] \end{bmatrix}, \quad G(X_i,\theta) = \frac{\partial g(X_i,\theta)}{\partial \theta'}, \quad F(X_i,\theta) = \frac{\partial vec[G(X_i,\theta)']}{\partial \theta'},$$

$$\underbrace{\Gamma(\theta)}_{m(p+1)\times p} = E[\frac{\partial \psi(X_i,\theta)}{\partial \theta'}] = \begin{bmatrix} G(\theta) \\ F(\theta) \end{bmatrix}, \qquad \underbrace{A(W,\theta)}_{p\times m(p+1)} = [\underbrace{G(\theta)'W}_{p\times m} \quad \underbrace{g(\theta)'W \otimes I_p}_{p\times mp}]. \tag{6}$$

We emphasize that the Jacobian term $G(X_i,\theta)$ already enters the influence function under misspecification without imposing extra moments. Based on (5), we can derive an asymptotic distribution of the GMM estimators under misspecification,

$$\sqrt{n}(\widehat{\theta}_{GMM}(W) - \theta_W) \xrightarrow{d} N(0, V_{GMM}(W,\theta_W)) \tag{7}$$

where

$$V_{GMM}(W,\theta) = (A(W,\theta)\Gamma(\theta))^{-1}A(W,\theta)\Sigma(\theta)A(W,\theta)'((A(W,\theta)\Gamma(\theta))')^{-1},$$

and $\Sigma(\theta) = E[\psi(X_i,\theta)\psi(X_i,\theta)'] - E[\psi(X_i,\theta)]E[\psi(X_i,\theta)]'$. The asymptotic distribution results of the misspecified-GMM in (7) are well-known in the literature; see Imbens (1997), Hall and Inoue (2003), Hansen and Lee (2021) for derivations of same result under misspecification for one-step/two-step GMM, and the iterated GMM estimator.[3]

3 Efficient Estimator under Misspecification

By inspection of the influence function in (5), we observe that the misspecified-GMM estimator is related to the M-estimator where the original moment functions and the Jacobian moments are "stacked" to form a vector of augmented moment conditions. We consider the following recentered moment function with augmented moments, $\widetilde{\psi}(X_i,\theta) := \widetilde{\psi}(X_i,\theta;W) = \psi(X_i,\theta) - E[\psi(X_i,\theta_W)]$, where $\psi(X_i,\theta)$ is de-

[3]One can generally interpret the overidentified GMM models as a just-identified GMM estimator with an augmented parameter vector (Newey and McFadden (1994)), and Imbens (1997) finds the connection with the misspecification-robust variance. However, they did not consider the explicit asymptotic variance form of the misspecified GMM estimator and the choice of weighting matrix, which are the main focus of this paper.

fined in (6), and θ_W is the GMM pseudo-true value in (2) for fixed W. By construction, the population augmented moments $E[\widetilde{\psi}(X_i, \theta)]$ are zero when θ is evaluated at θ_W.

We can then transform this to the just-identified moment $\Lambda E[\widetilde{\psi}(X_i, \theta_W)] = 0$ using the transform matrix $\Lambda = [\Lambda_1 \quad \Lambda_2]$, where Λ_1, Λ_2 are $p \times m$, and $p \times mp$ matrices, respectively. Then, we consider the following general class of *M-estimator* (or Z-estimator), $\widetilde{\theta}(\Lambda)$, to denote an M-estimator based on the sample analogs of augmented moment conditions

$$\Lambda \widetilde{\psi}_n(\widetilde{\theta}(\Lambda)) = 0 \tag{8}$$

where $\widetilde{\psi}_n(\theta) = \frac{1}{n}\sum_{i=1}^n \widetilde{\psi}(X_i, \theta)$. Alternative choices of Λ are associated with alternative estimators that are consistent for θ_W. This estimator is not feasible without estimation of $E[\psi(X_i, \theta_W)]$, but we postpone the discussion to focus first on the efficient choice of Λ and the asymptotic distribution of the oracle estimator.

3.1 Asymptotic Distribution of the M-Estimator with Augmented Moments

The following assumptions are used to show the asymptotic distribution of $\widetilde{\theta}(\Lambda)$.

Assumption 1. *1. The observations $\{X_i\}_{i=1}^n$ are independent and identically distributed (i.i.d.).*

2. For $W > 0$, $\theta_W = \arg\min_\theta g(\theta)' W g(\theta)$ is unique, and θ_W is in the interior of the compact parameter space $\Theta \subset \mathbb{R}^p$, where $g(\theta) \neq 0$ for all $\theta \in \Theta$.

3. $g(X, \theta)$ and $G(X, \theta)$ are continuous at each $\theta \in \Theta$ with probability one, and $E[\sup_\theta ||g(X_i, \theta)||] < \infty, E[\sup_\theta ||G(X_i, \theta)||] < \infty, E[\sup_\theta ||F(X_i, \theta)||] < \infty$.

4. $g(X, \theta)$ is twice continuously differentiable in θ with probability one, and $E[||g(X, \theta_W)||^2] < \infty$, $E[||G(X, \theta_W)||^2] < \infty$.

5. $\Lambda E[\widetilde{\psi}(X_i, \theta)] \neq 0$ for $\theta \neq \theta_W$. $\Lambda\Gamma(\theta_W)$ exists and is nonsingular.

Assumption 1.3 ensures uniform convergence of the $\widetilde{\psi}_n(\theta)$ and the population objective function $E[\widetilde{\psi}_n(\theta)]$ to show consistency. Assumption 1.4 is for the asymptotic normality of the normalized sum of $\widetilde{\psi}_n(\theta_W)$. The first part of Assumption 1.5 is an

identification condition. A necessary and sufficient condition for (local) identification is that $\Lambda\Gamma(\theta_W) = \Lambda_1 G(\theta_W) + \Lambda_2 F(\theta_W)$ has full column rank.[4] For example, when the moment functions are linear in θ, this reduces to the condition that $\Lambda\Gamma(\theta_W) = \Lambda_1 G$ has full column rank, where $G = E[G(X_i)]$ does not depend on θ. When Λ_1 takes the form of $\Lambda_1 = G'W$ with $W > 0$, it reduces to the standard GMM rank condition, which is implied by the global identification of θ_W in Assumption 1.2.

Theorem 1. *Under Assumption 1,*

$$\sqrt{n}(\widetilde{\theta}(\Lambda) - \theta_W) \xrightarrow{d} N(0, V_M(\Lambda, \theta_W)), \tag{9}$$

where

$$V_M(\Lambda, \theta_W) = (\Lambda\Gamma(\theta_W))^{-1}\Lambda\Sigma(\theta_W)\Lambda'((\Lambda\Gamma(\theta_W))')^{-1}$$

and $\Sigma(\theta) = E[\psi(X_i, \theta)\psi(X_i, \theta)'] - E[\psi(X_i, \theta)]E[\psi(X_i, \theta)]'$.

Theorem 1 shows that $\widetilde{\theta}(\Lambda)$ is consistent for the same GMM pseudo-true value θ_W, and the augmented-M estimator is a sufficiently general class of estimators to include the "standard" misspecified-GMM as a special case. When $\Lambda = A(W, \theta_W) = [G(\theta_W)'W \quad g(\theta_W)'W \otimes I_p]$ as in (5)-(6), $\widetilde{\theta}(\Lambda)$ has the same influence function representation as the original misspecified-GMM estimator $\widehat{\theta}_{GMM}(W)$ and has the same asymptotic distribution as in (7).

3.2 Efficient Choice of Λ and the Efficient Estimator under Misspecification

More importantly, Theorem 1 justifies the use of our new class of estimators, through the following corollaries, as we can consider an "efficient" choice of Λ to minimize the asymptotic variance $V_M(\Lambda, \theta_W)$. The optimal choice of Λ takes the following form[5]

$$\Lambda_W^* = \big(\frac{\partial}{\partial\theta'}E[\widetilde{\psi}(X_i, \theta)]\big)'|_{\theta=\theta_W} Var[\widetilde{\psi}(X_i, \theta_W)]^{-1} = \Gamma(\theta_W)'\Sigma(\theta_W)^{-1}, \tag{10}$$

[4] This is also related to the second-order identification condition because even if the Jacobian $G(\theta_W)$ is zero (rank deficient), the model can still be identified by nonzero $\Lambda_2 F(\theta_W)$.

[5] When $\Sigma(\theta_W)$ is singular, we can use the generalized inverse Σ^- and all the arguments below will still be valid. Σ is singular when some components of moments have an overlap, which occurs for example, when $g(X_i, \theta) = (X_i - \theta, (X_i - \theta)^2 - 1)'$ as $G(X_i, \theta) = (-1, -2(X_i - \theta))'$. In this case, the Jacobian carries no information beyond the original moments, but $(\Gamma'\Sigma^-\Gamma)$ may still be nonsingular.

by the standard theory of optimal estimating equations (Godambe (1960)) or semi-parametric efficiency of GMM.

With Λ_W^*, we use the optimal linear combinations of the augmented moments, while the specific choice $\Lambda = A(W, \theta_W)$ based on the FOC of the GMM objective is inefficient. When the estimator $\widetilde{\theta}(\Lambda)$ is constructed with Λ_W^* (or any consistent estimator of Λ_W^*), we label it as the *misspecification-efficient* (ME) estimator, $\widehat{\theta}_{ME}(W) = \widetilde{\theta}(\Lambda_W^*)$.

Corollary 1. *Suppose Assumption 1 holds for* $\Lambda = \Lambda_W^* = \Gamma(\theta_W)'\Sigma(\theta_W)^{-1}$. *Then, we have*

$$\sqrt{n}(\widehat{\theta}_{ME}(W) - \theta_W) \xrightarrow{d} N(0, V_{ME}(W)), \tag{11}$$

where $V_{ME}(W) = (\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Gamma(\theta_W))^{-1}$, *and the following holds for any* Λ *that satisfies Assumption 1,*

$$V_M(\Lambda, \theta_W) \geq V_{ME}(W).$$

Further, the asymptotic variance of the ME estimator has the following form,

$$V_{ME}(W) = (G(\theta_W)'\Sigma_{11}^{-1}G(\theta_W) + F_G(\theta_W)'\Sigma_{22,1}^{-1}F_G(\theta_W))^{-1}, \tag{12}$$

where $F_G(\theta_W) = F(\theta_W) - \Sigma_{21}\Sigma_{11}^{-1}G(\theta_W)$, $\Sigma_{22,1} = (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$, *and* $\Sigma(\theta_W) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$.

Corollary 1 shows that the misspecification-efficient estimator has the smallest asymptotic variance in the class of M-estimators that we consider with the augmented moment conditions $\widetilde{\psi}(X_i, \theta)$. Since misspecified-GMM estimator is the augmented-M estimator $\widetilde{\theta}(\Lambda)$ with $\Lambda = A(W, \theta_W) = [G(\theta_W)'W \quad g(\theta_W)'W \otimes I_p]$, Corollary 1 implies that

$$V_{GMM}(W, \theta_W) = V_M(A(W, \theta_W), \theta_W) \geq V_{ME}(W). \tag{13}$$

Corollary 1 also provides some insights on the "conventional" asymptotic variance of the efficient GMM under correct specification. The ME variance in (12) decomposes as the correct-specification efficient GMM variance contribution $G(\theta_W)'\Sigma_{11}^{-1}G(\theta_W)$ (evaluated at θ_W here instead of θ_0) and the curvature term contribution from the second derivative of the moments F after residualization $(F_G(\theta_W)'\Sigma_{22,1}^{-1}F_G(\theta_W))$, and we can deduce that

$$(G(\theta_W)'\Sigma_{11}^{-1}G(\theta_W))^{-1} \geq V_{ME}(W), \tag{14}$$

which implies that using the extra moment conditions from Jacobian does not worsen the asymptotic efficiency of estimators even under correct specification.[6]

When $F_G(\theta_W) = F(\theta_W) - \Sigma_{21}\Sigma_{11}^{-1}G(\theta_W) = 0$, we have $V_{ME}(W) = (G(\theta_W)'\Sigma_{11}^{-1}G(\theta_W))^{-1}$, so the ME estimator provides no efficiency gain. In the linear IV model, where $F(\theta_W) = 0$ and G does not depend on θ, this condition holds when $\Sigma_{12} = 0$, i.e., the covariance between the moment function $g(X_i, \theta)$ and the Jacobian $G(X_i)$ is zero. However, Σ_{12} (and thus $F_G(\theta_W)$) is generally nonzero both under correct specification and misspecification, implying that the ME estimator provides strictly positive efficiency gains in most empirically relevant settings.[7] Intuitively, variations of the Jacobian affect the distribution of the GMM estimator under misspecification, and thus improved efficiency can be achieved by imposing efficient weighting of the augmented moments accounting for the correlations between $g(X_i, \theta)$ and $G(X_i, \theta)$.

3.3 ME Estimator under the Linear Model

Under the linear model, we have further insights on the optimal choice of Λ_W^* and the asymptotic variance of the ME estimator. This will facilitate the discussions on the efficient weighting matrix under misspecification, compared with the "conventional" optimal weighting matrix that was defined under correct specification.

Example: Linear IV. Consider a linear instrumental variable regression $Y_i = X_i'\theta + \varepsilon_i$ where Y_i is an outcome variable, X_i is a potentially endogenous regressor, and Z_i is a $m \times 1$ vector of instrumental variables. Let Y, X, Z be $n \times 1, n \times p$, and $n \times m$ data matrices with row i equal to Y_i', X_i', Z_i', respectively. The moment function is $g(\cdot, \theta) = Z_i(Y_i - X_i'\theta)$, and we consider moment misspecification $E[Z_i(Y_i - X_i'\theta)] \neq 0$. With a $m \times m$ matrix $W > 0$, the standard GMM estimator is

$$\widehat{\theta}_{GMM}(W) = (X'ZWZ'X)^{-1}X'ZWZ'Y$$

[6]Under correct specification, $E[g(X_i, \theta_0)] = 0$, and $V_{GMM}(W, \theta_0)$ in (7) reduces to the classical form $(G'WG)^{-1}G'W\Sigma_{11}WG(G'WG)^{-1}$ evaluated at θ_0, where Σ_{11} is the $m \times m$ upper-left submatrix of $\Sigma(\theta_0)$. This is minimized by the "standard" optimal-weight matrix $W_0 = \Sigma_{11}^{-1}$, which gives $(G'\Sigma_{11}^{-1}G)^{-1}$.

[7]To see this, consider the simple linear IV model where $g(X_i, \theta) = Z_i(Y_i - X_i\theta)$, and θ is a scalar. Define $\varepsilon_i = Y_i - X_i\theta_W, \sigma_{\varepsilon X} = E[\varepsilon_i X_i | Z_i]$, $\sigma_{\varepsilon Z} = E[Z_i\varepsilon_i]$, and $X_i = \pi' Z_i + v_i, E[Z_i v_i] = 0$. Then, $\Sigma_{12} = -(E[\sigma_{\varepsilon X}Z_iZ_i'] - \sigma_{\varepsilon Z}E[X_iZ_i'])$. Under correct specification ($\sigma_{\varepsilon Z} = 0$), $\Sigma_{12} = -E[\sigma_{\varepsilon X}Z_iZ_i']$ is generally nonzero whenever there is endogeneity, $\sigma_{\varepsilon X} \neq 0$. Under misspecification ($\sigma_{\varepsilon Z} \neq 0$), Σ_{12} is also generally nonzero.

with the GMM pseudo-true value,

$$\theta_W = (E[X'Z]WE[Z'X])^{-1}(E[X'Z]WE[Z'Y]).$$

When $W = E[Z'Z]^{-1}$, θ_W is the TSLS estimand. Based on (8) and using the linearity of the moment condition, and the optimal Λ_W^* in Corollary 2 below, the ME has a closed-form as follows

$$\widehat{\theta}_{ME}(W) = (\frac{1}{n}X'Z\Sigma_{11,2}^{-1}Z'X)^{-1}\bigg(X'Z\Sigma_{11,2}^{-1}(\frac{1}{n}Z'Y - E[g(X_i,\theta_W)]) \\ - X'Z\Sigma_{11,2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}(\frac{1}{n}vec(X'Z) - E[vec(X_iZ_i')])\bigg) \quad (15)$$

where $\Sigma_{11,2} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$, $\Sigma(\theta_W) = \left[\begin{smallmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{smallmatrix}\right]$, with $\Sigma(\theta) = E[\left(\begin{smallmatrix}Z_iY_i - Z_iX_i'\theta\\ vec(X_iZ_i')\end{smallmatrix}\right)\left(\begin{smallmatrix}Z_iY_i - Z_iX_i'\theta\\ vec(X_iZ_i')\end{smallmatrix}\right)'] - E\left(\begin{smallmatrix}Z_iY_i - Z_iX_i'\theta\\ vec(X_iZ_i')\end{smallmatrix}\right)E\left(\begin{smallmatrix}Z_iY_i - Z_iX_i'\theta\\ vec(X_iZ_i')\end{smallmatrix}\right)'$. Sample analogs of $\Sigma(\theta_W)$ can also be used here. In the linear IV example, the ME estimator (15) can be viewed as efficient linear combinations of the recentered GMM/TSLS estimator and the Jacobian moment conditions.

Assuming that the data-generating process is $Y_i = \theta_0 X_i + Z_{2i} + \varepsilon_i$, $X_i = Z_{1i} + 2Z_{2i} + v_i$, $Z_i = (Z_{1i}, Z_{2i})' \overset{i.i.d.}{\sim} N(0, I_2)$, $(\varepsilon_i, v_i) \sim N(0, [1, 0.5; 0.5, 1])$ with $\theta_0 = 0$. When we consider the weight matrix $W = \left(\begin{smallmatrix}1 & \rho\\ \rho & 1\end{smallmatrix}\right)$ for $\rho \in (-1, 1)$, the GMM pseudo-true value is $\theta_W(\rho) = \frac{\rho+2}{5+4\rho}$, which is a strictly decreasing function of $\rho \in (-1, 1)$, and there is a one-to-one mapping from ρ to $\theta_W(\rho)$. By using the formula above, the asymptotic variance of the GMM estimator as a function of ρ is:

$$V_{GMM}(\rho) = \frac{2(56\rho^4 + 131\rho^3 + 210\rho^2 + 232\rho + 100)}{(5+4\rho)^4}.$$

The variance of the ME estimator can be calculated as $V_{ME} = \frac{969}{6116} \approx 0.1584$ (which does not depend on ρ), and it can be verified that $V_{GMM}(\rho) > V_{ME}$ for all ρ. At $\rho = 0$ (TSLS), the variance simplifies to $V_{GMM}(0) = 0.32$, and the efficiency gain is over 50% against TSLS. The efficiency gain $(1 - \frac{V_{ME}}{V_{GMM}(\rho)})$ grows as ρ decreases; for example, it is 73% and 88% for $\rho = -0.5$ and -0.9, respectively.[8]

[8]Although we do not report the results here for brevity, we also considered a general nonlinear instrumental variables (NLIV) model, where the moment function is $g(X_i, \theta) = Z_i(Y_i - \exp(X_i'\theta))$. This model has been considered in count data models with endogenous regressors (e.g., Windmeijer and Santos Silva (1997)). Using numerical methods, we verified that $V_{GMM} > V_{ME}$, and V_{ME} varies with W in the nonlinear model.

When the model is linear, i.e., $G(X_i) = \frac{\partial g(X_i,\theta)}{\partial \theta'}$ does not depend on θ, and $F(X_i,\theta) = \frac{\partial vec[G(X_i,\theta)']}{\partial \theta'} = 0$, we can further simplify the form of Λ_W^* and the asymptotic variance of the ME estimator.

Corollary 2. *If $g(X,\theta)$ is linear in θ, we have*

$$\Lambda_W^* = [G'\Sigma_{11,2}^{-1} \quad -G'\Sigma_{11,2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}] \tag{16}$$

where $G = E[G(X_i)], \Sigma_{11,2} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$, $\Sigma(\theta_W) = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}$. The asymptotic variance of the ME estimator is

$$V_{ME}(W) = (G'\Sigma_{11,2}^{-1}G)^{-1}, \tag{17}$$

and $\Sigma_{11,2}$ is invariant to θ_W.

Under the linear model, Corollary 2 shows that the asymptotic variance of the ME estimator is $(G'\Sigma_{11,2}^{-1}G)^{-1}$, where $\Sigma_{11,2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. We refer to $W^* = \Sigma_{11,2}^{-1}$ as the *misspecification-efficient* weighting matrix, in analogy with the conventional (correct-specification) efficient weighting matrix $W_0 = \Sigma_{11}^{-1}$. While W_0 is the inverse of the variance of the original moments $g(X_i,\theta)$, $W^* = \Sigma_{11,2}^{-1}(\theta)$ is the inverse of the variance of $g(X_i,\theta) - \Sigma_{12}\Sigma_{22}^{-1}vec(G(X_i)')$, which is the residualized moment after the projection onto the Jacobian. Even under correct specification, the ME bound $(G'\Sigma_{11,2}^{-1}G)^{-1}$ is strictly smaller than the asymptotic variance of the GMM estimator with the conventional efficient weighting matrix W_0 when $\Sigma_{12} \neq 0$ as we discussed.

Corollary 2 shows that the conventional efficient GMM variance formula $(G'WG)^{-1}$ achieves the efficiency bounds of the oracle-ME estimator, when $W^* = \Sigma_{11,2}^{-1}(\theta_W)$. The caveat is that the GMM estimator itself with the choice W^* does not achieve the oracle ME bound as shown in Corollary 1, $V_{GMM}(W^*,\theta_W) \geq (G'\Sigma_{11,2}^{-1}G)^{-1}$ under misspecification. It follows that the conventional formula for efficient GMM is too "small", even in large samples, as it will be valid for the ME estimator, but not for the standard GMM estimator. Researchers have found that the standard errors of the efficient GMM are often severely downward biased.[9] Our results show that

[9]Many researchers thus routinely use the finite-sample correction of Windmeijer (2005). Hwang

reporting the conventional efficient GMM variance formula in the linear GMM will still be useful as an oracle efficiency bound; showing researchers what could be gained by imposing moments from the Jacobian $G(X_i, \theta)$, but should not be used directly for inference with the standard GMM itself under misspecification. When $\Sigma_{12} = 0$, $W^* = W_0 = \Sigma_{11}^{-1}$, and the misspecification-efficient weighting matrix reduces to the conventional one.

We can easily report the oracle ME bound, $V_{ME}(W) = (G'\Sigma_{11,2}^{-1}G)^{-1}$, as an efficiency bound alongside conventional/misspecification-robust standard errors with no significant cost. We can consistently estimate $V_{ME}(W) = (G'\Sigma_{11,2}^{-1}G)^{-1}$ regardless of the model specification based on the sample estimator

$$\widehat{\Sigma}_{11,2} = \frac{1}{n}\sum_i (r_i(\widehat{\theta}) - \bar{r}_n(\widehat{\theta}))(r_i(\widehat{\theta}) - \bar{r}_n(\widehat{\theta}))'$$

with $r_i(\theta) = g_i(\theta) - \widehat{\Sigma}_{12}\widehat{\Sigma}_{22}^{-1} vec(G_i(\theta)')$, $\bar{r}_n(\theta) = 1/n \sum r_i(\theta)$ and the preliminary GMM estimator $\widehat{\theta} = \widehat{\theta}_{GMM}(W)$.[10]

Furthermore, Corollary 2 shows that under the linear model, $V_{ME} = (G'\Sigma_{11,2}^{-1}G)^{-1}$ is the same for all different pseudo-true values θ_W, due to the invariance property of $\Sigma_{11,2}$. This implies that misspecification affects what is being estimated, but not how precisely it can be estimated. This is because the linear GMM model has a "location-like" structure in the sense that $g(X_i, \theta) = g(X_i, 0) + G(X_i)\theta$, so shifting θ moves the moment vector along the Jacobian direction. The invariance property holds because $\Sigma_{11,2}$ is the variance of the residual from the population regression of $g(X_i, \theta)$ on $vec(G(X_i)')$, so all the θ-dependent parts are projected out. Practically, this implies that V_{ME} in the linear model can be consistently estimated from any preliminary GMM without committing to a particular weight matrix.

In general nonlinear settings, above arguments fail to hold as the Jacobian $G(X_i, \theta)$, and $\Sigma_{11,2}(\theta)$ change with θ. However, the distribution of the oracle-ME estimator $\widehat{\theta}_{ME}(W)$ and the oracle bounds $V_{ME}(W)$ can be approximated by the bootstrap, and it can be viewed as a misspecification-robust and efficient version of the Hall and Horowitz (1996) recentered bootstrap; see Section 4.

---

et al. (2022) further show, under the linear model, that misspecification-robust GMM variance will not only be valid under misspecification, but also work as a finite-sample correction even if the model is correctly specified.

[10] Under misspecification, we must use this centered covariance estimator because the uncentered covariance estimator $\frac{1}{n}\sum_i r_i(\widehat{\theta})r_i(\widehat{\theta})'$ is not consistent for $\Sigma_{11,2}$.

Remark 1 (Decomposition of the ME efficiency gain)**.** *Consider the simple linear IV model, $Y_i = X_i\theta_0 + Z_i'\eta + \varepsilon_i, X_i = Z_i'\pi + v_i$, where θ_0 is a scalar with m excluded instruments Z_i. Suppose (ε_i, v_i) are conditionally homoskedastic given Z_i, and the first-stage and the direct effects of instruments are local, $\pi = O(n^{-1/2}), \eta = O(n^{-1/2})$. Then, we show (in Supplemental Appendix F) that we have the following decomposition for $W \propto Q_{ZZ}^{-1}, Q_{ZZ} = E[Z_i Z_i']$ (which covers TSLS, W_0, W^* under homoskedasticity)*

$$\frac{V_{GMM}(W, \theta_W)}{V_{ME}} - 1 \approx \frac{\rho^2}{1-\rho^2} + \frac{J_n}{\mu_n^2(1-\rho^2)} \tag{18}$$

where ρ is the correlation between $\varepsilon_W = Y_i - X_i\theta_W$ and X_i, $J_n = ng(\theta_W)'\Sigma_{11}^{-1}g(\theta_W)$, and $\mu_n^2 = n\pi' Q_{ZZ}\pi/\sigma_v^2$. We can report this measure approximating μ_n^2 by mF with the number of excluded instruments m and the first-stage F statistic, and J_n by the reported Hansen J statistics. So, the ME efficiency gain depends on the moment-Jacobian correlation and also on the degree of misspecification relative to the strength of identification, rather than the degree of misspecification alone. We also show that the misspecification-robust SE exceeds the conventional SE by the factor $\sqrt{1 + J_n/(mF)}$, while the ME bound lies below the conventional SE by the factor $\sqrt{1-\rho^2}$. See Supplemental Appendix F for detailed derivations and Section 7 for the empirical results and how the decomposition in (18) *can be helpful in these examples.*

4 Misspecification-Efficient Bootstrap GMM Estimators

We first consider a related class of estimators, misspecification-efficient (ME) GMM based on the augmented moment conditions. We then propose the bootstrap GMM estimator, which can be considered as a misspecification-robust and efficient version of the Hall and Horowitz (1996) recentered bootstrap GMM estimator.

The ME-GMM estimator, $\widehat{\theta}_{ME-GMM}(\Delta)$, solves

$$\widehat{\theta}_{ME-GMM}(\Delta) = \arg\min_{\theta} \left(n^{-1} \sum_{i=1}^{n} \widetilde{\psi}(X_i, \theta) \right)' \Delta \left(n^{-1} \sum_{i=1}^{n} \widetilde{\psi}(X_i, \theta) \right) \tag{19}$$

where $\Delta > 0$ is a fixed weighting matrix, and $\widetilde{\psi}(X_i, \theta) = \begin{pmatrix} g(X_i, \theta) - E[g(X_i, \theta_W)] \\ vec(G(X_i, \theta)') - E[vec(G(X_i, \theta_W)')] \end{pmatrix}$.

Under the same assumptions (Assumption 1.2-1.4), we can easily show that the ME-GMM with the efficient weighting matrix $\Delta = E[\widetilde{\psi}(X_i, \theta_W)\widetilde{\psi}(X_i, \theta_W)']^{-1} = \Sigma(\theta_W)^{-1}$ is asymptotically equivalent to the ME estimator, $\widehat{\theta}_{ME}(W)$. We note that Lee and Liao (2018) consider the same GMM problem under correct specification ($E[g(X_i, \theta_W)] = 0$) and a singular Jacobian ($E[G(X_i, \theta_W)] = 0$), where the recentering terms are known to be zero.

Here, we consider the bootstrap GMM estimator which mimics (19) with the efficient weighting matrix $\Delta = \Sigma(\theta_W)^{-1}$. Let $\{X_i^*\}_{i=1}^n$ be the bootstrap samples, drawn independently from the original sample $\{X_i\}_{i=1}^n$ with replacement. With recentering, the bootstrap ME-GMM estimator solves

$$\widehat{\theta}^*_{ME}(W) = \arg\min_{\theta} \left(n^{-1} \sum_{i=1}^{n} \widetilde{\psi}^*(X_i^*, \theta) \right)' \Delta^* \left(n^{-1} \sum_{i=1}^{n} \widetilde{\psi}^*(X_i^*, \theta) \right), \tag{20}$$

where

$$\widetilde{\psi}^*(X_i^*, \theta) = \begin{pmatrix} g(X_i^*, \theta) - g_n(\widehat{\theta}_W) \\ vec(G(X_i^*, \theta)') - vec(G_n(\widehat{\theta}_W)') \end{pmatrix},$$

$$g_n(\theta) = n^{-1} \sum_{i=1}^{n} g(X_i, \theta), \quad G_n(\theta) = n^{-1} \sum_{i=1}^{n} G(X_i, \theta),$$

$$\Delta^* = \left(E^*[\psi(X_i^*, \widehat{\theta}_W)\psi(X_i^*, \widehat{\theta}_W)'] - E^*[\psi(X_i^*, \widehat{\theta}_W)]E^*[\psi(X_i^*, \widehat{\theta}_W)]' \right)^{-1}, \tag{21}$$

and $\widehat{\theta}_W = \widehat{\theta}_{GMM}(W)$ is the GMM estimator. The recentering by $g_n(\widehat{\theta}_W), vec(G_n(\widehat{\theta}_W))$ ensures $E^*[g(X_i^*, \widehat{\theta}_W) - g_n(\widehat{\theta}_W)] = 0$, and the bootstrap moments $\widetilde{\psi}^*(X_i^*, \theta)$ have mean zero at $\widehat{\theta}_W$ conditional on the data. Δ^* is a symmetric positive definite weighting matrix that is a bootstrap sample version of the efficient weighting matrix $\Delta = \Sigma(\theta_W)^{-1}$.

The proposed bootstrap estimator utilizes the recentering of the Jacobian estimator to ensure that the bootstrap expectation of the recentered Jacobian moments evaluated at the estimator is zero, which is essential to correctly mimic the asymptotic distribution under misspecification. This "double recentering" idea has also been considered in weak-identification literature to accurately reflect the behavior of the original Jacobian (e.g., Dovonon and Gonçalves (2017)), yet under correct specification.

The standard nonparametric bootstrap without recentering provides first-order valid inference under correct specification (Hahn (1996)), however, the standard bootstrap does not achieve an asymptotic refinement in overidentified models. This is because the bootstrap sample mean of the moment function evaluated at the estimator, $E^*[g(X_i^*, \widehat{\theta}_W)] = g_n(\widehat{\theta}_W)$, is not necessarily equal to zero, and the recentering has been considered important to achieve higher-order improvements under correct specification (Hall and Horowitz (1996), Brown and Newey (2002)). Hall and Horowitz (1996, HH hereafter) proposed a recentered bootstrap GMM estimator,

$$\widehat{\theta}^*_{HH} = \arg\min_{\theta} \left(n^{-1} \sum_{i=1}^{n} g(X_i^*, \theta) - g_n(\widehat{\theta}_W) \right)' W_0^* \left(n^{-1} \sum_{i=1}^{n} g(X_i^*, \theta) - g_n(\widehat{\theta}_W) \right) \quad (22)$$

where W_0^* is the bootstrap version of the efficient weighting matrix under correct specification, $W_0 = \Sigma_{11}(\theta)^{-1}$. $\widehat{\theta}^*_{HH}$ can be considered as a special case of the proposed bootstrap ME-GMM estimator $\widehat{\theta}^*_{ME}$ by ignoring the Jacobian augmented parts and only recentering the original sample moments, i.e., $\Delta_{11}^* = W_0^*, \Delta_{12}^* = 0$, and $\Delta_{22}^* = 0$.

The conventional GMM and the HH bootstrap estimator would have different asymptotic distributions under misspecification because the Hall-Horowitz recentered bootstrap always creates a correctly specified bootstrap world, regardless of whether the actual model is misspecified or not. The Hall-Horowitz bootstrap eliminates the additional source of Jacobian sampling variability that arises under misspecification, while the proposed ME bootstrap can consider these variations, and furthermore it considers more efficient combinations of the sampling uncertainty for the original and Jacobian moments. In the linear IV model, the bootstrap ME-GMM estimator has a closed form and is an efficient linear combination of the HH recentered bootstrap TSLS estimator and the recentered Jacobian; see Supplemental Appendix A.

We show below that the asymptotic distribution of the bootstrap estimator $\widehat{\theta}^*_{ME}(W)$ is identical to that of the original ME estimator $\widehat{\theta}_{ME}(W)$. We consider the following assumptions for the bootstrap validity. Define bootstrap sample mean of moments and Jacobian as $g_n^*(\theta) = n^{-1} \sum_{i=1}^n g(X_i^*, \theta), G_n^*(\theta) = n^{-1} \sum_{i=1}^n G(X_i^*, \theta)$.

Assumption 2. *1.* $\widehat{\theta}^*_{ME}(W) \xrightarrow{p^*} \theta_W$.

2. $\sup_{\theta \in \Theta} \|g_n^*(\theta) - g_n(\theta)\| \xrightarrow{p^*} 0$, *and* $\sup_{\theta \in \Theta} \|G_n^*(\theta) - G_n(\theta)\| \xrightarrow{p^*} 0$.

3. $\sqrt{n}\big(g_n^*(\widehat{\theta}_W) - g_n(\widehat{\theta}_W),\ \mathrm{vec}(G_n^*(\widehat{\theta}_W)' - G_n(\widehat{\theta}_W)')\big) \xrightarrow{d^*} N(0, \Sigma(\theta_W))$, *where* $\widehat{\theta}_W =$

$\widehat{\theta}_{GMM}(W)$, *and* $\Sigma(\theta_W)$ *is the same covariance matrix as in Theorem 1.*

These conditions are standard assumptions under i.i.d. sampling; see, e.g., Giné and Zinn (1990), and Hahn (1996). Assumption 2.1 on the bootstrap consistency is a high-level assumption, which can be verified from more primitive conditions using the standard bootstrap consistency of the extremum estimators, combined with the uniform convergence conditions in Assumption 2.2. Assumption 2.3 is implied by Assumptions 1.1-1.4 under the i.i.d. setup.

Theorem 2. *Suppose that Assumptions 1.1-1.4 hold. Further, we assume that* $\theta = \theta_W$ *is a unique solution to* $E[\widetilde{\psi}(X_i, \theta)] = 0$*, and* Δ *is positive definite. Then, we have*

$$\sqrt{n}(\widehat{\theta}_{ME-GMM}(\Delta) - \theta_W) \xrightarrow{d} N(0, (\Gamma(\theta_W)'\Delta\Gamma(\theta_W))^{-1}\Gamma(\theta_W)'\Delta\Sigma(\theta_W)\Delta\Gamma(\theta_W)(\Gamma(\theta_W)'\Delta\Gamma(\theta_W))^{-1}).$$

When $\Delta = \Sigma(\theta_W)^{-1}$*, the asymptotic variance reduces to* $(\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Gamma(\theta_W))^{-1}$*, provided it is non-singular. In addition, suppose that Assumption 2 holds, then*

$$\sqrt{n}(\widehat{\theta}^*_{ME}(W) - \widehat{\theta}_{GMM}(W)) \xrightarrow{d^*} N(0, (\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Gamma(\theta_W))^{-1}).$$

Theorem 2 shows that the asymptotic distribution of the bootstrap estimator $\sqrt{n}(\widehat{\theta}^*_{ME}(W) - \widehat{\theta}_{GMM}(W))$ is equal to that of $\sqrt{n}(\widehat{\theta}_{ME}(W) - \theta_W)$. Theorem 2 allows us to approximate the sampling variation of the oracle-ME estimator $\widehat{\theta}_{ME}(W)$, and the theorem justifies the use of the bootstrap to obtain the minimax bounds that will be established in Section 5 (Theorem 3).

4.1 Double-Recentered Bootstrap Estimator

We note, however, that Theorem 2 does not guarantee the validity of the percentile bootstrap CI based on $\widehat{\theta}^*_{ME}(W)$ because it is centered at $\widehat{\theta}_{GMM}(W)$, which has a different asymptotic distribution. Using the double-recentering idea from the above section, we propose a double-recentered (DR) bootstrap that jointly perturbs the moment conditions and the Jacobian. The DR bootstrap correctly approximates the asymptotic distribution of the GMM estimator regardless of specification status, nesting the Hall and Horowitz (1996) bootstrap as a special case under correct specification.

Lee (2014) shows that the standard nonparametric GMM bootstrap with the analytic misspecification-robust variance estimator works regardless of misspecification

unlike Hall and Horowitz (1996), and argues that the recentering can be detrimental. In our paper, however, we clarify this argument that additional considerations of re-centered Jacobian can achieve asymptotic validity under misspecification. While the standard GMM bootstrap estimator without recentering will be valid combined with the analytic misspecification-robust SE, the Hall and Horowitz (1996) bootstrap does not achieve asymptotic validity under misspecification because it *only* recenters the original moment functions.[11]

We define the double-recentered bootstrap estimator $\widehat{\theta}^*_{DR}$ as the solution to the bootstrap analog of the augmented estimating equation (8) with the particular choice of $\Lambda = A(W, \theta_W) = [G(\theta_W)'W,\ g(\theta_W)'W \otimes I_p]$:

$$A_n(W, \widehat{\theta}_W)\widetilde{\psi}^*_n(\widehat{\theta}^*_{DR}) = 0, \tag{23}$$

where

$$\widetilde{\psi}^*_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \widetilde{\psi}^*(X^*_i, \theta) = \begin{pmatrix} g^*_n(\theta) - g_n(\widehat{\theta}_W) \\ vec(G^*_n(\theta)') - vec(G_n(\widehat{\theta}_W)') \end{pmatrix},$$

and

$$A_n(W, \widehat{\theta}_W) = \big[G_n(\widehat{\theta}_W)'W\ ,\ \ g_n(\widehat{\theta}_W)'W \otimes I_p\big].$$

While the HH bootstrap only recenters the original moments g^*_n at $g_n(\widehat{\theta}_W)$, the DR bootstrap recenters the full augmented vector ψ^*_n at its sample counterpart ψ_n, including Jacobian estimation variability, appropriately weighted by $\Lambda_n = A_n(W, \widehat{\theta}_W)$ to mimic the distribution of $\widehat{\theta}_W$.[12]

For the implementation, we first compute the original sample GMM estimator $\widehat{\theta}_W$, and calculate $A_n(W, \widehat{\theta}_W)$, which requires $g_n(\widehat{\theta}_W)$ and $G_n(\widehat{\theta}_W)$. Then, for each bootstrap replication $b = 1, \ldots, B$:

1. Draw $\{X^*_i\}_{i=1}^n$ with replacement;

[11] Since the main focus of the paper is on the efficiency results for θ_W and minimax bounds, we only considered the GMM estimator with a fixed W. Thus, we do not need to consider variations from the estimated weight matrix, which was typically considered for the construction of the misspecification-robust variance in the literature. Although it is beyond the scope of this paper, the idea of the "double recentering" in this paper can be easily extended to the "triple recentering" with the estimated W_n (one-step) or the $W_n(\theta)$ (two-step) case by augmenting the weighting matrix into the $\psi(\cdot)$.

[12] We note that the DR bootstrap estimator is different from the bootstrap M-estimator based on the just-identified moment functions with an augmented parameter as in Imbens (1997). One can show that the bootstrap M-estimator based on the just-identified augmented moment functions corresponds to the standard GMM bootstrap estimator without recentering.

2. With $A_n(W, \widehat{\theta}_W)$, solve (23) for $\widehat{\theta}^*_{DR}$ via Newton–Raphson initialized at $\widehat{\theta}_W$. For the linear model, $\widehat{\theta}^*_{DR}$ has a closed form (see Supplemental Appendix A);

3. Confidence intervals are constructed as standard percentile or percentile-t intervals based on $t^* = \frac{\widehat{\theta}^*_{DR} - \widehat{\theta}_W}{s.e.(\widehat{\theta}^*_{DR})}$ from the B bootstrap draws.

Corollary 3. *Suppose that Assumptions 1.1-1.5, and Assumption 2 hold (replacing consistency Assumption 2.1 with $\widehat{\theta}^*_{DR}$). Then,*

$$\sqrt{n}(\widehat{\theta}^*_{DR} - \widehat{\theta}_{GMM}(W)) \xrightarrow{d^*} N(0, V_{GMM}(W, \theta_W)),$$

where $V_{GMM}(W, \theta)$ is defined in (7).

We recommend $\widehat{\theta}^*_{DR}$ and the percentile confidence interval for valid inference for θ_W. In addition, we recommend reporting separately $\widehat{V}_{ME}(W)$ and $\sup_{W \in \mathcal{W}} \widehat{V}_{ME}(W)$ based on the recentered bootstrap ME estimator $\widehat{\theta}^*_{ME}$ or using analytical ME estimates with the standard GMM point estimate $\widehat{\theta}_{GMM}(W)$ as evidence of the attainable oracle bound by showing researchers what could be gained if a feasible ME point estimate were available with known misspecification and identification. This provides honest bounds that acknowledge the researcher doesn't know which W (and hence θ_W) is "right."

5 Semiparametric Efficiency under Misspecification

Based on the results from the previous sections, we derive the uniform (over $W \in \mathcal{W}$) asymptotic minimax bounds of the misspecified moment condition models. We can generally view the misspecified GMM model as a semiparametric model such that we consider a family of probability distributions of the observed data X, indexed by the finite-dimensional parameter of interest θ for some weighting matrix W. In practice, researchers can choose different weighting matrices $W \in \mathcal{W}$ based on the application they consider, where $\mathcal{W}$ denotes a class of $m \times m$ positive definite matrices.

However, the notion of semiparametric efficiency and optimality needs to be defined differently under (global) misspecification. As noted earlier, the fundamental difficulty here is that the pseudo-true value θ_W changes with $W \in \mathcal{W}$ in our misspecified moment condition models. The semiparametric efficiency framework typically

assumes the existence of a well-defined "true" parameter, and the efficiency bound is computed at this target parameter. For example, the GMM estimator is semi-parametrically efficient under correct specification (Chamberlain (1987)) for $\theta = \theta_0$ under the moment conditions $E[g(X_i, \theta)] = 0$ with an "efficient" weighting matrix $W = \Sigma_{11}(\theta_0)^{-1}$, the inverse of the variance matrix of the original moments.

We may compare two or more different weighting matrices, and may want to choose a "better" W and pseudo-true θ_W than the others, based on the pseudo-distance measure such as J statistics, or based on the goodness-of-fit measures or the information criteria (e.g., Rivers and Vuong (2002), Marmer and Otsu (2012)). We can also consider the optimal weights $W^\dagger = \arg\min_W V_{GMM}(W, \theta_W)$ that minimize the asymptotic variance of misspecified-GMM in (7) or $W^\dagger = \arg\min_W V_{ME}(W)$ that minimizes the asymptotic variance of misspecification-efficient GMM in (11). However, as noted in Hall and Pelletier (2011), economic theory dictates an appropriate choice of the weighting matrix in some cases and the researcher may choose W based on which θ_W they care about substantively, and in the absence of these economic considerations, the choice of W and the relative ranking over $\mathcal{W}$ can be arbitrary.

In this paper, we treat $\mathcal{W}$ as fixed, but in practice $\mathcal{W}$ can be chosen by the researchers. For example, researchers may focus only on the identity matrix and diagonal matrices due to computational costs, or may impose some restrictions to obtain particular causal interpretations.

Example: Linear IV (continued). Suppose we focus on a scalar θ ($p = 1$), instruments Z consist of m single instruments $Z_1, \cdots Z_m$. Then, the GMM pseudo-true value can be written as a linear combination of the single-instrument IV estimands $\theta_j = E[Z_j'Y]/E[Z_j'X]$,

$$\theta_W = \frac{E[X'Z]WE[Z'Y]}{E[X'Z]WE[Z'X]} = \sum_{j=1}^{m} w_j \theta_j$$

where the weight w_j for the j-th instrument is given by:

$$w_j = \frac{\pi' W e_j e_j' \pi}{\pi' W \pi} = \frac{(\sum_{k=1}^m W_{jk}\pi_k)\pi_j}{\pi' W \pi}$$

with the j-th unit vector e_j and $\pi = E[Z'X]$. The weights sum to one, but w_j is negative if π_j and $\sum_{k=1}^m W_{jk}\pi_k$ have opposite signs. When $W = E[Z'Z]^{-1}$, θ_W is

the TSLS estimand and Angrist and Imbens (1995) and Kolesár (2013) characterize θ_W as a weighted average of LATE under monotonicity conditions (see also Andrews (2019)). However, the TSLS weights are not guaranteed to be positive. Researchers may restrict the weighting matrices to the class with positive weights,

$$\mathcal{W}^+ = \{W \in \mathcal{W} : \pi' W e_j e_j' \pi \geq 0 \text{ for all } j = 1, \cdots, m, \pi = E[Z'X]\},$$

which preserves the sign of the instrument strength, $\text{sign}\left((W\pi)_j\right) = \text{sign}(\pi_j)$, and includes any diagonal positive definite matrix (e.g., identity, or $\text{diag}((E[Z'Z])^{-1})$). The TSLS weighting matrix is in this class only under strict conditions when we have covariates (e.g., Blandhol et al. (2025) and Słoczyński (2024)). Under treatment effect heterogeneity, $W \in \mathcal{W}^+$ guarantees that θ_W lies between the minimum and maximum of the single-instrument LATE estimands θ_j, and the bounds below are uniform over this class of causal estimands.

5.1 Uniform Asymptotic Minimax Bounds

To formally state our results, we first define the class of distributions and the parameter spaces we consider. We also provide more specific definitions of correct/misspecified models.

Suppose that we observe an i.i.d. sample $\{X_i\}_{i=1}^n$ drawn from the true (unknown) probability distribution F_0. Let $\mathcal{F}$ be the space of all probability distributions and define $\mathcal{F}^C$ as follows;

$$\begin{aligned}\mathcal{F}^C \;=\; \{&(F,\theta) \in \mathcal{F} \times \Theta : (i)\ \theta \text{ is in an open set } \Theta \subset \mathbb{R}^p \text{ such that } g(X,\beta) \text{ and } \partial g(X,\beta)/\partial\beta' \\ &\text{are continuous for } (X,\beta) \in \mathcal{X} \times \Theta; (ii)\ E_F g(X,\theta) = 0; \\ &(iii)\ E_F[g(X,\theta)g(X,\theta)'] \text{ is positive-definite}; (iv)\ rank\ E_F[\partial g(X,\theta)/\partial\theta'] = p\},\end{aligned} \tag{24}$$

i.e., $\mathcal{F}^C$ is the set of probability distributions F, such that for some $\theta \in \Theta$, (F,θ) satisfies the regularity and the moment conditions in $\mathcal{F}^C$. Correct specification assumes $(F_0,\theta_0) \in \mathcal{F}^C$ such that the true probability distribution F_0 with unique parameter value θ_0 satisfies conditions above. Chamberlain (1987) shows that $\widehat{\theta}_{GMM}(\Sigma_{11}(\theta_0)^{-1})$ achieves semiparametric efficiency bounds in the sense of Hájek's (1972) local asymptotic minimaxity for the neighborhood of $(F_0,\theta_0) \in \mathcal{F}^C$.

In this paper, we consider the misspecified moment condition models and the

pseudo-true values associated with the choice of weighting matrix W. We define that the moment function $g(\cdot)$ is misspecified if

$$(F_0, \theta) \notin \mathcal{F}^C \text{ for all } \theta \in \Theta.$$

Then, for any $W \in \mathcal{W}$, where $\mathcal{W}$ is a set of positive definite matrices, we can define the pseudo-true values as the minimizer of a goodness-of-fit measure of moment condition $g(\cdot)$ to the true data distributions F_0 using the weighted-norm associated with the weighting matrix W;

$$\theta_W = \arg\min_{\theta} E_{F_0} g(X_i, \theta)' W E_{F_0} g(X_i, \theta).$$

To consider the uniform minimax bounds, we define the extended class of models $\mathcal{G}$ as the set of tuples, $\mathcal{G} = \{(W, (F, \theta)) : W \in \mathcal{W}, (F, \theta) \in \mathcal{F}^W\}$, where $\mathcal{F}^W$ is the space of probability distributions that is consistent with θ_W, i.e., the set of probability distributions F, such that for some $\theta \in \Theta$, (F, θ) satisfies the following regularity conditions,

$$\begin{aligned}
\mathcal{F}^W \;=\; \{(F,\theta) \in \mathcal{F} \times \Theta : (i)\;\; & \theta \text{ is in an open set } \Theta \subset \mathbb{R}^p \text{ such that } \psi(X,\beta) \text{ and } \partial\psi(X,\beta)/\partial\beta' \\
& \text{are continuous for } (X,\beta) \in \mathcal{X} \times \Theta; (ii)\;\; E_F\tilde{\psi}(X,\theta) = 0, \text{ where} \\
& \tilde{\psi}(X,\theta) = \psi(X,\theta) - E_{F_0}[\psi(X,\theta_W)], \theta_W = \arg\min_{\theta} E_{F_0} g(X_i,\theta)' W E_{F_0} g(X_i,\theta), \qquad (25) \\
& \text{assuming it exists and unique; } (iii)\;\; E_F[\tilde{\psi}(X,\theta)\tilde{\psi}(X,\theta)'] \text{ is positive-definite;} \\
& (iv)\;\; rank\;\; E_F[\partial\psi(X,\theta)/\partial\theta'] = p\}.
\end{aligned}$$

Under the same identification assumption (Assumption 1), we have $(F_0, \theta_W) \in \mathcal{F}^W$. Condition (i) requires that the original function $g(X, \theta)$ is twice continuously differentiable. Condition (iv) requires that $\Gamma(\theta) = [G(\theta); F(\theta)]$ has rank p. This is a weaker condition than standard GMM identification (condition (iv) in $\mathcal{F}^C$) because it allows identification to come from the curvature ($F(\theta)$) in addition to the first-order derivative ($G(\theta)$). $\mathcal{F}^W$ is defined through a recentering that depends on the pseudo-true value θ_W (and thus implicitly on F_0), and this is innocuous for the local asymptotic minimax analysis; it concerns the minimax risk over neighborhoods of (F_0, θ_W), and within such a neighborhood, the recentering is fixed.

We introduce further notation on a neighborhood of the probability distribution

and the loss function $l(\cdot)$. Let Π be the set of all probability measures F defined on the Borel sets in a Euclidean space with the support of a probability measure $\mathcal{X}$. Given an $F \in \Pi$, the basic neighborhoods of F are sets of the form:

$$\left\{ Q \in \Pi : \left| \int f_j dQ - \int f_j dF \right| < \epsilon_j, j = 1, \ldots, k \right\}$$

where $\epsilon_j > 0$, k is some positive integer, and $f_j : \mathcal{X} \to \mathbb{R}$ are measurable functions such that $\int |f_j| dF < \infty$. An arbitrary neighborhood of F is formed by taking unions of sets of this form. Let $\mathcal{L}$ be the class of loss functions $l \in \mathcal{L}$ that satisfy the following conditions for all $u, v \in \mathbb{R}$: (i) $l(u) = l(|u|)$, (ii) $|u| \leq |v|$ implies $l(u) \leq l(v)$, (iii) $\int_{-\infty}^{\infty} l(u) \exp\left(-\frac{1}{2}\lambda u^2\right) du < \infty$ for all $\lambda > 0$, (iv) $l(0) = 0$. Many standard loss functions satisfy these conditions, for example, quadratic $l(u) = u^2$, absolute function $l(u) = |u|$, and polynomial loss $l(u) = |u|^p$ for $p > 0$.

The next theorem shows the semiparametric efficiency bounds in the sense of Hájek's (1972) local asymptotic minimaxity for the neighborhood of $(W, (F_0, \theta_W)) \in \mathcal{G} = \{(W, (F, \theta)) : W \in \mathcal{W}, (F, \theta) \in \mathcal{F}^W\}$.

Theorem 3. *Suppose that (F_0, θ_W) satisfies the conditions in $\mathcal{F}^W$ for all $W \in \mathcal{W}$. Let Δ^W be any neighborhood of (F_0, θ_W), and let Γ^W be the subset of Δ^W such that the conditions in $\mathcal{F}^W$ are satisfied for all $(F, \theta) \in \Gamma^W$. Let $\mathcal{G}^\Gamma = \{(W, (F, \theta)) : W \in \mathcal{W}, (F, \theta) \in \Gamma^W\}$. For any pair $(F, \theta) \in \Gamma^W$, let θ_1 be the first component of θ and let $T_n(X_1, \ldots, X_n)$ be any (measurable) estimator for θ_1. Then for any loss function $l(\cdot) \in \mathcal{L}$, we have*

$$\liminf_{n \to \infty} \sup_{(W,(F,\theta)) \in \mathcal{G}^\Gamma} E_F \left\{ l\left[\sqrt{n}(T_n - \theta_1) \right] \right\} \geq \int_{-\infty}^{\infty} l(\sigma u) d\Phi(u),$$

where $\Phi(u)$ is the cumulative distribution function of the standard normal distribution, $\sigma^2 = \sup_W \sigma_W^2 < \infty$, σ_W^2 is the (1,1) element of $V_{ME}(W) = (\Gamma(\theta_W)' \Sigma(\theta_W)^{-1} \Gamma(\theta_W))^{-1}$.

Theorem 3 states that the asymptotic variance of any semiparametric estimator under the model class $\mathcal{G} = \{(W, (F, \theta)) : W \in \mathcal{W}, (F, \theta) \in \mathcal{F}^W\}$ is no smaller than the bounds with the least favorable weighting matrix W that maximizes risk over the class $\mathcal{W}$. This bound can be achieved by the worst-case variance of the ME estimator $\widehat{\theta}_{ME}(W), W \in \mathcal{W}$, which can be feasibly approximated by the analytic formula or the bootstrap method. Since V_{ME} doesn't depend on W in the linear IV model, the ME

estimator eliminates the question of which W is most efficient among $W \in \mathcal{W}$, and the uniformity can be obtained with no extra cost.

The bound in Theorem 3 is based on the model class $\mathcal{F}^W$, where the recentering part $E_{F_0}[\psi(X, \theta_W)]$ is held fixed. One can straightforwardly show that the misspecified-GMM estimator achieves a different semiparametric efficiency bound $V_{GMM}(W, \theta_W)$ in a class of models that restricts F only through the definition of θ_W (by replacing moment condition (ii) in $\mathcal{F}^C$ (equation (24)) with the local first order condition $E_F[\partial g(X,\theta)/\partial\theta']'W E_F[g(X,\theta)] = 0$); see also Mukhin (2019). Equivalently, one can use the augmented parameter space approach as in Imbens (1997) with the regularity (smoothness) conditions to obtain the same efficiency bound $V_{GMM}(W, \theta_W)$. However, as shown in Corollary 1, the optimal linear combinations of the augmented moments $\Lambda^*_W = \Gamma(\theta_W)'\Sigma(\theta_W)^{-1}$ exploit the correlation between the moments and the Jacobian, which the standard GMM combination $\Lambda = [G(\theta_W)'W \quad g(\theta_W)'W \otimes I_p]$, dictated by the FOC of the GMM objective, ignores. Using the moment vector $\widetilde{\psi}(X, \theta)$ provides an over-identified system for θ, exploiting more information from the degree of misspecification and the identification strength.

6 Monte Carlo Simulation

We compare the finite-sample performance of the standard GMM and the proposed misspecification-efficient estimator in a linear IV model. The data generating process is

$$Y_i = X_i\theta + Z_i'\eta + \varepsilon_i, \qquad X_i = Z_i'\pi + v_i, \tag{26}$$

with $Z_i = (Z_{1i}, Z_{2i})' \sim_{iid} N(0, I_2)$ and $(\varepsilon_i, v_i) \sim N(0, [1, 0.5; 0.5, 1])$. We set $\theta = 1$, and vary the "degree of misspecification" through the direct effect $\eta = (0, \delta)'$ with $\delta \in \{0, 0.5, 1, 2\}$; $\delta = 0$ is the correctly specified benchmark, $\delta > 0$ introduces an exclusion violation through Z_2. The sample sizes are $n \in \{200, 500, 1000\}$. We set the first-stage coefficient $\pi = (\tilde{\pi}, 2\tilde{\pi})'$ so that the scaled concentration parameter (μ^2/m) is 50 or 10, corresponding to moderately strong and weak instrument setups for $n = 200$. The pseudo-true value θ_W is W-dependent and moves away from $\theta = 1$ as δ changes. We consider $W = I, \widehat{\Sigma}^{-1}_{11,2}$, and the moderately strong instrument case here in the main paper, and the weak instrument design results, together with additional results under the conventional optimal weighting $W = \widehat{\Sigma}^{-1}_{11}$ are reported

in Supplemental Appendix D. For each estimator, we report the standard deviation (SD), empirical 95% coverage (Cov), and median CI length (Len) over $2,000$ Monte Carlo replications, using $B = 1,000$ bootstrap draws. SD is normalized relative to the SD of standard GMM with the correctly specified benchmark ($n = 200, \delta = 0$).

Table 1 reports results for the one-step ($W = I$) and the two-step GMM ($W = \widehat{\Sigma}_{11,2}^{-1}$). For each (n, δ), we report the following estimators and standard errors (SE) with Wald CI: (i) standard GMM estimator with the conventional SE; (ii) GMM with the misspecification-robust SE; (iii) the oracle ME estimator with efficient $V_{ME}(W)$. We also consider the following bootstrap estimators: (iv) HH (Hall and Horowitz (1996)) percentile bootstrap CI; and (v) the DR (double-recentered) percentile bootstrap CI proposed in Section 4.1. Since the oracle ME results are identical under the two weighting matrices, we report them only once.

As the degree of misspecification (δ) increases, we observe several consistent patterns across DGPs; 1) the SD of GMM increases, 2) the Wald CI of GMM with conventional SE undercovers, while the misspecification-robust SE restores coverage by correctly capturing the SD of GMM, and 3) the oracle ME dominates in SD and CI length across all designs, while retaining coverage (coverage 94-96% across all $\{n, \delta\}$), consistent with Corollaries 1 and 2. Note that the SD of standard GMM grows substantially with δ, whereas the SD of oracle ME grows only modestly (e.g., SD of GMM $0.43 \to 1.38$, SD of ME $0.40 \to 0.79$, when δ changes from 0 to 2, $n = 1000$); the oracle ME estimator takes into account increases in variance through optimal Jacobian recentering rather than letting the asymptotic variance inflate. The efficiency gains of the oracle ME over GMM ($1 - V_{ME}/V_{GMM}$, $n = 1000$) are 25%, 56%, 67% for $W = I$, and 14%, 50%, 63% for $W = \widehat{\Sigma}_{11,2}^{-1}$, when $\delta = 0.5, 1, 2$, respectively. We also observe that the efficient weighting choice $W = \widehat{\Sigma}_{11,2}^{-1}$ reduces the SD of GMM.

Furthermore, the DR percentile CI has coverage close to the nominal level in all designs, while the HH bootstrap exhibits coverage distortions under misspecification. Under $W = \widehat{\Sigma}_{11,2}^{-1}$, the weight does not depend on the preliminary estimator in the linear model, and the corresponding Wald CI with misspecification-robust SE is close to the nominal level, and comparable to the DR percentile CI. By contrast, under the standard optimal weighting matrix $W = \widehat{\Sigma}_{11}^{-1}$ (Table D.1 in Supplemental Appendix D), the weight depends on the preliminary estimator, and the corresponding Wald CI under-covers (0.90-0.92 for $\delta = 1, 2$), whereas the DR CI remains close to the nominal level.

The simulation results confirm the efficiency of the oracle ME estimator, and overall suggest that the DR bootstrap provides a valid procedure, and a reasonable alternative to the standard GMM with analytic misspecification-robust standard errors.

7 Empirical Examples

We illustrate our proposed methods with two main empirical applications: the returns to schooling examples of Card (1995) and Angrist and Krueger (1991). We also consider the dynamic panel data model of the employment equation of Arellano and Bond (1991) in Supplemental Appendix E.

7.1 Card (1995)

We replicate the well-known college-proximity IV analysis in Card (1995), who uses the National Longitudinal Survey of Young Men for 1976. The sample size is $n = 3010$. The outcome is log wage and the regressors include education and some demographic/geographic variables. The instruments include two binary indicators for living near a four-year public and private college, respectively.

Table 2 compares several estimators across four specifications and three weighting choices: $W = I$, $W = (Z'Z)^{-1}$ (TSLS), and $W = \widehat{\Sigma}_{11,2}^{-1}$ (efficient).[13] The first column (1) corresponds to the baseline specification with experience, experience squared, an indicator for Black, indicators for residence in an SMSA in 1976, and residence in the South in 1976. Column (3) treats experience and its square as endogenous, adding age and age squared to the instrument set. These specifications correspond to Card (1995, Table 3). In Columns (2) and (4), we additionally control for more covariates including cubic and quartic experience, and all pairwise interactions between baseline covariates, motivated by the rich-covariates condition of Blandhol et al. (2025), required for the TSLS estimand to admit a weakly causal interpretation.

Table 2 shows that GMM point estimates are very similar across specifications and W (0.159-0.176 in four columns with three W). The conventional and misspecification-robust SEs are very similar in this example (within about 2%). The Hansen J test

[13] In both applications we consider here, we also considered the conventional optimal weight $W = \widehat{\Sigma}_{11}^{-1}$ for each specification. The results are almost identical and thus we do not report them in the paper.

Table 1: Monte Carlo results under $W = I$ and $W = \widehat{\Sigma}_{11,2}^{-1}$

	$\delta = 0$			$\delta = 0.5$			$\delta = 1$			$\delta = 2$		
Estimator	SD	Cov	Len	SD	Cov	Len	SD	Cov	Len	SD	Cov	Len
Weight matrix: $W = I$												
$n = 200$												
GMM + Conv. SE	1.00	0.945	0.386	1.01	0.916	0.351	1.62	0.867	0.476	3.10	0.869	0.907
GMM + Robust SE	1.00	0.946	0.390	1.01	0.953	0.400	1.62	0.940	0.612	3.10	0.951	1.203
HH Bootstrap	1.00	0.940	0.395	0.90	0.910	0.359	1.25	0.866	0.488	2.41	0.875	0.930
DR Bootstrap	1.03	0.943	0.404	1.04	0.951	0.410	1.62	0.940	0.629	3.19	0.953	1.246
$n = 500$												
GMM + Conv. SE	0.61	0.956	0.246	0.65	0.910	0.223	1.02	0.862	0.304	1.92	0.871	0.584
GMM + Robust SE	0.61	0.956	0.247	0.65	0.945	0.256	1.02	0.940	0.392	1.92	0.952	0.764
HH Bootstrap	0.62	0.953	0.249	0.56	0.909	0.226	0.77	0.859	0.307	1.49	0.870	0.590
DR Bootstrap	0.63	0.955	0.249	0.65	0.943	0.257	1.00	0.938	0.397	1.96	0.952	0.775
$n = 1000$												
GMM + Conv. SE	0.43	0.955	0.175	0.45	0.912	0.159	0.71	0.863	0.217	1.38	0.859	0.416
GMM + Robust SE	0.43	0.955	0.175	0.45	0.950	0.181	0.71	0.945	0.278	1.38	0.948	0.542
HH Bootstrap	0.44	0.957	0.175	0.40	0.913	0.159	0.54	0.861	0.217	1.05	0.862	0.418
DR Bootstrap	0.44	0.959	0.176	0.45	0.947	0.183	0.70	0.946	0.279	1.37	0.947	0.546
Weight matrix: $W = \widehat{\Sigma}_{11,2}^{-1}$												
$n = 200$												
GMM + Conv. SE	1.00	0.943	0.384	0.94	0.940	0.347	1.48	0.882	0.460	2.95	0.855	0.841
GMM + Robust SE	1.00	0.944	0.388	0.94	0.951	0.374	1.48	0.938	0.557	2.95	0.943	1.116
HH Bootstrap	1.00	0.941	0.393	0.90	0.935	0.355	1.20	0.880	0.472	2.22	0.853	0.862
DR Bootstrap	1.02	0.942	0.401	1.02	0.957	0.402	1.54	0.954	0.599	3.00	0.944	1.168
$n = 500$												
GMM + Conv. SE	0.61	0.953	0.245	0.61	0.933	0.223	0.94	0.876	0.295	1.84	0.856	0.541
GMM + Robust SE	0.61	0.954	0.246	0.61	0.950	0.239	0.94	0.936	0.360	1.84	0.948	0.724
HH Bootstrap	0.62	0.955	0.248	0.56	0.933	0.225	0.75	0.877	0.297	1.38	0.854	0.545
DR Bootstrap	0.63	0.956	0.249	0.64	0.957	0.253	0.95	0.950	0.378	1.85	0.947	0.731
$n = 1000$												
GMM + Conv. SE	0.43	0.953	0.174	0.42	0.932	0.158	0.66	0.883	0.210	1.31	0.855	0.385
GMM + Robust SE	0.43	0.953	0.175	0.42	0.950	0.170	0.66	0.945	0.256	1.31	0.950	0.513
HH Bootstrap	0.44	0.954	0.175	0.40	0.933	0.159	0.53	0.881	0.211	0.97	0.852	0.385
DR Bootstrap	0.44	0.955	0.176	0.45	0.962	0.179	0.67	0.949	0.267	1.30	0.946	0.517
Oracle ME												
$n = 200$	0.93	0.944	0.367	0.87	0.949	0.349	1.08	0.945	0.422	1.78	0.942	0.695
$n = 500$	0.58	0.945	0.232	0.56	0.945	0.221	0.69	0.944	0.267	1.09	0.952	0.440
$n = 1000$	0.40	0.958	0.164	0.39	0.948	0.156	0.47	0.950	0.189	0.79	0.951	0.311

Notes. We consider (i) standard GMM estimator with the conventional SE; (ii) GMM with the misspecification-robust SE; (iii) Hall and Horowitz (1996) (HH) percentile bootstrap; and (iv) the Double-Recentered percentile bootstrap (DR), under $W = I$ and $W = \widehat{\Sigma}_{11,2}^{-1}$, respectively. We also report the oracle ME estimator with efficient SE, $V_{ME}(W)$, whose results are identical under the two weighting matrices (Corollary 2). For GMM and Oracle ME, "SD" is the MC standard deviation of the point estimate across MC draws; for HH and DR bootstrap, "SD" is the average bootstrap standard deviation across MC draws. "SD" is reported relative to the correctly-specified benchmark SD (GMM with $\delta = 0, W = I$, $n = 200$), so values < 1 (> 1) indicate higher (lower) precision than this benchmark. "Cov" is empirical 95% coverage; "Len" is median CI length. Based on 2,000 Monte Carlo replications, $B = 1{,}000$ bootstrap draws.

does not reject in any specifications (p-values ≥ 0.35); however, misspecification can still exist even when the J test fails to reject. Although the ME efficiency bound is essentially similar across specifications (0.0343-0.0366), it is 9-19% smaller than the conventional/misspecification-robust SE. The ME-GMM bootstrap SD tracks the bound closely, and the DR percentile CI is comparable to the Wald CI with misspecification-robust SEs across specifications.

As in Mogstad and Torgovitsky (2024), we find that a Ramsey (1969) RESET test rejects the null of rich-covariates (testing whether $E[Z|X]$ is linear in X) in basic specifications (1) and (3), rejects at 5% but not at 1% in Column (2) ($p = 0.041$), and the condition holds in Column (4) with saturated exogenous controls. Since the point estimates are similar across all specifications, the same caveat as in Mogstad and Torgovitsky (2024) applies: two estimates can be very similar even when one corresponds to a non-negatively weighted causal estimand and the other does not.

7.2 Angrist and Krueger (1991)

Now, we consider the quarter-of-birth instruments example of Angrist and Krueger (1991). We use the US Census 1980 extract of men born 1930-1939 ($n = 329,509$). The outcome is the log wage and the endogenous variable is education with exogenous controls (year-of-birth, race, SMSA, marital status, region, state). The instruments are quarter-of-birth (QOB) with the interactions. In Table 3, Column (1) uses the baseline QOB dummies ($m = 3$); (2) uses QOB×YOB interactions ($m = 30$); (3) adds QOB×State interactions ($m = 180$) which may exhibit many and many-weak instrument issues as noted in Bound et al. (1995). These specifications correspond to Angrist and Krueger (1991, Table V Columns 2 and 4 and Table VII Column 8). For each case, we consider the basic controls and the flexible specification that adds all interactions between controls.

The results in Table 3 parallel Table 2 in the Card (1995) example. The misspecification-robust SE only marginally exceeds the conventional SE, and the gap is larger for many instrument specifications. The ME-GMM bootstrap SD tracks the ME efficiency bound closely, in line with the simulation evidence. The ME efficiency bound is substantially smaller than the conventional/misspecification-robust SE when $W = I$. When we use TSLS weights ($(Z'Z)^{-1}$) and efficient weights ($\widehat{\Sigma}_{11,2}^{-1}$), the ME efficiency bounds are close to the conventional SE, but still 3-18% smaller than the

Table 2: Returns to schooling: Card (1995)

	exp/exp^2 as exogenous		exp/exp^2 as endogenous	
	(1)	(2)	(3)	(4)
ME efficiency bound	(0.0366)	(0.0347)	(0.0347)	(0.0343)
Panel A: $W = I$ (identity)				
GMM point	0.1763	0.1734	0.1674	0.1663
Conv. SE	(0.0448)	(0.0412)	(0.0430)	(0.0418)
Misspec-robust SE	(0.0445)	(0.0410)	(0.0427)	(0.0416)
ME-GMM Boot SD	(0.0375)	(0.0358)	(0.0372)	(0.0374)
DR Boot 95% CI	[0.096,0.289]	[0.104,0.269]	[0.092,0.291]	[0.099,0.292]
Panel B: $W = (Z'Z)^{-1}$ (TSLS)				
GMM point	0.1611	0.1620	0.1597	0.1616
Conv. SE	(0.0405)	(0.0382)	(0.0408)	(0.0404)
Misspec-robust SE	(0.0412)	(0.0386)	(0.0415)	(0.0408)
ME-GMM Boot SD	(0.0382)	(0.0357)	(0.0368)	(0.0370)
DR Boot 95% CI	[0.090,0.259]	[0.091,0.258]	[0.090,0.282]	[0.096,0.273]
Panel C: $W = \widehat{\Sigma}_{11,2}^{-1}$ (efficient)				
GMM point	0.1606	0.1619	0.1594	0.1615
Conv. SE	(0.0404)	(0.0381)	(0.0408)	(0.0404)
Misspec-robust SE	(0.0413)	(0.0387)	(0.0417)	(0.0407)
ME-GMM Boot SD	(0.0377)	(0.0361)	(0.0376)	(0.0361)
DR Boot 95% CI	[0.089,0.265]	[0.091,0.254]	[0.090,0.275]	[0.094,0.281]
Panel D: Diagnostics				
First-stage F	13.5	15.7	8.6	9.0
J-test p-value	0.351	0.410	0.461	0.592
RESET p-value	0.000	0.041	0.000	-
# excluded IVs	2	2	4	4
# exogenous controls	6	18	4	8

Notes. We report returns to education coefficients with standard errors or bootstrap SD (in parentheses). Column (1) includes the baseline controls with indicators for Black, residence in an SMSA in 1976, residence in the South in 1976, and experience and experience squared. Column (2) additionally controls for cubic and quartic terms of experience, and all interactions between covariates. Columns (3) and (4) treat experience and its square as endogenous, adding age and age squared to the instrument sets. Column (3) includes baseline specifications and Column (4) adds pairwise interactions. Specifications correspond to Card (1995, Table 3). $n = 3010$. Bootstrap size: $B = 2,000$. The ME efficiency bound is the oracle ME standard error $\widehat{V}_{ME}(W)^{1/2}/\sqrt{n}$ with $V_{ME}(W) = (G'\Sigma_{11,2}^{-1}G)^{-1}$ (Corollary 2), which does not depend on W. DR bootstrap percentile 95% CI in Section 4.1 is reported in brackets.

misspecification-robust SE as predicted by our theory. By Remark 1, the ratio of the misspecification-robust SE to the ME bound is approximately $\sqrt{(1 + J/(mF))/(1 - \rho^2)}$. The moment-Jacobian correlation ρ^2 is small in this application, and the gap increases with the number of instruments as $J/(mF)$ increases because F falls (weak instruments).

Adding interaction terms barely changes the GMM point estimates. However, the GMM point estimate under $W = I$ is systematically higher (0.088-0.113) than under TSLS or efficient weighting (0.079-0.099). Furthermore, the DR bootstrap CI is close to the Wald CIs with the misspecification-robust SE in most cases. The RESET test rejects the rich-covariates condition of Blandhol et al. (2025) in all specifications, except (1) for the basic instruments. ME-GMM inference still applies to θ_W, but the causal interpretation of the estimates in Table 3 may be fragile in that case.

8 Conclusion

This paper develops the theory of efficient GMM estimation under moment misspecification. The key insight is that the influence function of the standard misspecified-GMM estimator depends both on the original moment conditions and their Jacobian. By optimally weighting this augmented system, we propose a misspecification-efficient estimator that achieves a smaller asymptotic variance than standard GMM for the same pseudo-true value. We also establish semiparametric efficiency bounds uniform over a class of weighting matrices, and propose a bootstrap implementation based on double recentering of both moment and Jacobian conditions. We illustrate our theory in two cross-sectional IV applications, and find that the efficiency loss is small to moderate, with the TSLS and the efficient weights.

Several directions remain for future work. First, extending the theory to dependent data with HAC-type weighting matrices would significantly broaden the applicability of the ME estimator to time-series settings. Second, we only consider the one-step GMM with a fixed W. Under misspecification, the pseudo-true value changes with each step of the two-step or iterated GMM, whose asymptotic distributions also differ. Existing papers typically use the "standard" optimal weighting matrix W_0, which we show in this paper to be suboptimal under misspecification; it would be interesting to analyze the behavior of the two-step/iterated GMM based on $W^* = \Sigma_{11,2}^{-1}(\theta)$ considered in this paper. Third, the relationship between the ME esti-

Table 3: Returns to schooling: Angrist and Krueger (1991)

	(1) QOB (m=3)		(2) + QOB × YOB (m=30)		(3) + QOB × State (m=180)	
	basic	flexible	basic	flexible	basic	flexible
ME efficiency bound	(0.0204)	(0.0207)	(0.0164)	(0.0166)	(0.0095)	(0.0096)
Panel A: $W = I$ (identity)						
GMM point	0.1127	0.1116	0.0919	0.0904	0.0906	0.0881
Conv. SE	(0.0250)	(0.0252)	(0.0193)	(0.0194)	(0.0148)	(0.0149)
Misspec-robust SE	(0.0252)	(0.0254)	(0.0210)	(0.0211)	(0.0162)	(0.0163)
ME-GMM Boot SD	(0.0204)	(0.0202)	(0.0151)	(0.0154)	(0.0080)	(0.0080)
DR Boot 95% CI	[0.063,0.166]	[0.062,0.168]	[0.051,0.135]	[0.050,0.134]	[0.058,0.124]	[0.057,0.121]
Panel B: $W = (Z'Z)^{-1}$ (TSLS)						
GMM point	0.0990	0.0981	0.0806	0.0792	0.0831	0.0809
Conv. SE	(0.0207)	(0.0210)	(0.0165)	(0.0166)	(0.0098)	(0.0098)
Misspec-robust SE	(0.0210)	(0.0213)	(0.0178)	(0.0179)	(0.0115)	(0.0114)
ME-GMM Boot SD	(0.0202)	(0.0210)	(0.0151)	(0.0151)	(0.0081)	(0.0079)
DR Boot 95% CI	[0.058,0.142]	[0.058,0.144]	[0.045,0.117]	[0.042,0.114]	[0.061,0.106]	[0.058,0.103]
Panel C: $W = \widehat{\Sigma}_{11,2}^{-1}$ (efficient)						
GMM point	0.0991	0.0982	0.0821	0.0807	0.0836	0.0814
Conv. SE	(0.0207)	(0.0210)	(0.0165)	(0.0166)	(0.0096)	(0.0096)
Misspec-robust SE	(0.0210)	(0.0213)	(0.0177)	(0.0178)	(0.0116)	(0.0113)
ME-GMM Boot SD	(0.0201)	(0.0205)	(0.0149)	(0.0150)	(0.0081)	(0.0081)
DR Boot 95% CI	[0.060,0.141]	[0.056,0.141]	[0.044,0.117]	[0.044,0.116]	[0.061,0.106]	[0.059,0.103]
Panel D: Diagnostics						
First-stage F	30.5		4.75		2.43	
Hansen J p-value	0.310	0.348	0.821	0.838	0.774	0.808
RESET p-value	0.831	0.870	0.000	0.000	0.000	0.000

Notes. We report returns to education coefficients with standard errors or bootstrap SD (in parentheses). Controls include year-of-birth, race, SMSA, marital status, region, and state dummies. Column (1) uses the baseline QOB dummies ($m = 3$); (2) uses QOB×YOB interactions ($m = 30$), (3) adds QOB×State interactions ($m = 180$). "Basic" includes baseline specifications and "flexible" adds all pairwise dummy interactions. Specifications correspond to Angrist and Krueger (1991, Table V Columns 2 and 4 and Table VII Column 8). $n = 329,509$. Bootstrap size: $B = 2,000$. The ME efficiency bound is the oracle ME standard error $\widehat{V}_{ME}(W)^{1/2}/\sqrt{n}$ with $V_{ME}(W) = (G'\Sigma_{11,2}^{-1}G)^{-1}$ (Corollary 2), which does not depend on W in the linear model. DR bootstrap percentile 95% CI in Section 4.1 is reported in brackets.

mator and generalized empirical likelihood (GEL) estimators under misspecification deserves further investigation. Under misspecification, these estimators have different pseudo-true values and different asymptotic distributions, so they are generally difficult to compare in a unified framework. Fourth, extending the framework to accommodate weak identification, where the Jacobian $G(\theta_W)$ is near-singular, and the behavior of the GMM estimators may be nonstandard (e.g., Stock and Wright (2000), Andrews and Cheng (2012), Dovonon and Renault (2013)), would be important given the connection to the literature (e.g., Lee and Liao (2018) and Kleibergen (2005)) that already exploits the Jacobian moment conditions. Finally, developing practical guidance on the choice of the weighting matrix class $\mathcal{W}$ would provide applied researchers with concrete recommendations for implementation.

Appendix: Proofs

Proof of Theorem 1

Let $G(X_i,\theta) = \partial g(X_i,\theta)/\partial\theta'$, $G_n(\theta) = n^{-1}\sum_{i=1}^n G(X_i,\theta)$, $G(\theta) = E[G_n(\theta)]$. Also, let $F(X_i,\theta) = \frac{\partial}{\partial\theta'}vec(G(X_i,\theta)')$, $F_n(\theta) = n^{-1}\sum_{i=1}^n F(X_i,\theta)$ and $F(\theta) = E[F_n(\theta)]$. Using the first-order Taylor expansion from the FOC of the GMM estimator,

$$G_n(\widehat{\theta}_{GMM}(W))'W g_n(\widehat{\theta}_{GMM}(W)) = 0,$$

we have

$$0 = G_n(\theta_W)'W g_n(\theta_W) + H_n(\widetilde{\theta})(\widehat{\theta}_{GMM}(W) - \theta_W)$$

where

$$H_n(\theta) = G_n(\theta)'W G_n(\theta) + (g_n(\theta)'W \otimes I_p)F_n(\theta)$$

and $\widetilde{\theta}$ lies between $\widehat{\theta}_{GMM}(W)$ and θ_W. By rearranging, we have

$$\begin{aligned}
\widehat{\theta}_{GMM}(W) - \theta_W &= -H_n(\widetilde{\theta})^{-1}G_n(\theta_W)'W g_n(\theta_W) \\
&= -H_n(\widetilde{\theta})^{-1}[G_n(\theta_W)'W(g_n(\theta_W) - g(\theta_W)) + (G_n(\theta_W) - G(\theta_W))'W g(\theta_W)] \\
&= -H_n(\widetilde{\theta})^{-1}[G_n(\theta_W)'W(g_n(\theta_W) - g(\theta_W)) + (g(\theta_W)'W \otimes I_p)vec((G_n(\theta_W) - G(\theta_W))'
\end{aligned}$$

where we use the FOC of the population GMM, $G(\theta_W)'W g(\theta_W) = 0$, in the second equation, and use the properties of the Kronecker product, $vec(AB) = (B'\otimes I)vec(A)$.

Under Assumption 1, which is the standard regularity conditions for the GMM estimators, we have $\widehat{\theta}_{GMM}(W) \xrightarrow{p} \theta_W, G_n(\theta_W) \xrightarrow{p} G(\theta_W), H_n(\widetilde{\theta}) \xrightarrow{p} H(\theta_W)$, where $H(\theta) = G(\theta)'WG(\theta) + (g(\theta)'W \otimes I_p)F(\theta) = A(W,\theta)\Gamma(\theta)$. Thus,

$$\begin{aligned}\sqrt{n}(\widehat{\theta}_{GMM}(W) - \theta_W) &= -H(\theta_W)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{G(\theta_W)'W[g(X_i,\theta_W) - Eg(X_i,\theta_W)] \\ &\quad +(g(\theta_W)'W \otimes I_p)vec[G(X_i,\theta_W)' - EG(X_i,\theta_W)']\} + o_p(1) \\ &= -(A(W,\theta_W)\Gamma(\theta_W))^{-1}A(W,\theta_W)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\psi(X_i,\theta_W) - E[\psi(X_i,\theta_W)]) + o_p(1).\end{aligned}$$

Then, the asymptotic distribution of the GMM estimators in (7) follows by CLT and CMT under Assumption 1.

We next focus on the $\widetilde{\theta}(\Lambda)$. A consistency result for $\widetilde{\theta}(\Lambda) \xrightarrow{p} \theta_W$ holds under Assumption 1. Then, by the Taylor expansion of (8) around θ_W, we have

$$0 = \Lambda\widetilde{\psi}_n(\theta_W) + \Lambda\Gamma_n(\overline{\theta})(\widetilde{\theta}(\Lambda) - \theta_W)$$

where

$$\Gamma_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta'}\psi(X_i,\theta) = \begin{bmatrix} G_n(\theta) \\ F_n(\theta)\end{bmatrix}$$

and $\overline{\theta}$ lies between $\widetilde{\theta}(\Lambda)$ and θ_W. Then, under Assumption 1,

$$\widetilde{\theta}(\Lambda) - \theta_W = -(\Lambda\Gamma(\theta_W))^{-1}\Lambda[\psi_n(\theta_W) - E[\psi(X_i,\theta_W)]] + o_p(n^{-1/2})$$

where $\psi_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\psi(X_i,\theta)$, and the asymptotic distribution in Theorem 1 holds by CLT. □

Proof of Corollary 1

With the choice of $\Lambda = \Lambda_W^*$ in (10), the asymptotic distribution of the ME estimator in (11) follows by Theorem 1 as the asymptotic variance of M-estimator reduces to

$$\begin{aligned}V_M(\Lambda_W^*,\theta_W) &= (\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Gamma(\theta_W))^{-1}\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Sigma(\theta_W)\Sigma(\theta_W)^{-1}\Gamma(\theta_W) \\ &\quad ((\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Gamma(\theta_W))')^{-1} \\ &= (\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Gamma(\theta_W))^{-1}.\end{aligned}$$

We now show that $V_M(\Lambda, \theta_W) - (\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Gamma(\theta_W))^{-1}$ is positive semidefinite. We first note that the following projection matrix is positive semidefinite,

$$I - \Sigma(\theta_W)^{-1/2}\Gamma(\theta_W)(\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Gamma(\theta_W))^{-1}\Gamma(\theta_W)'\Sigma(\theta_W)^{-1/2} \geq 0.$$

We also note that if a $L \times L$ square matrix P is positive semidefinite, then HPH' is also positive semidefinite for any $K \times L$ matrix H. So, with $H = (\Lambda\Gamma(\theta_W))^{-1}\Lambda\Sigma(\theta_W)^{1/2}$, we have

$$(\Lambda\Gamma(\theta_W))^{-1}\Lambda\Sigma(\theta_W)^{1/2}\big(I - \Sigma(\theta_W)^{-1/2}\Gamma(\theta_W)(\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Gamma(\theta_W))^{-1}\Gamma(\theta_W)'\Sigma(\theta_W)^{-1/2}\big)$$
$$\Sigma(\theta_W)^{1/2}\Lambda'(\Gamma(\theta_W)'\Lambda')^{-1} \geq 0.$$

This implies

$$V_M(\Lambda, \theta_W) - V_M(\Lambda^*_W, \theta_W) = (\Lambda\Gamma(\theta_W))^{-1}\Lambda\Sigma(\theta_W)\Lambda'((\Lambda\Gamma(\theta_W))')^{-1} - (\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Gamma(\theta_W))^{-1} \geq 0.$$

For the last part of Corollary 1, we have

$$\Sigma(\theta_W)^{-1} = \begin{bmatrix} \Sigma_{11,2}^{-1} & -\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22,1}^{-1} \\ -\Sigma_{22,1}^{-1}\Sigma_{21}\Sigma_{11}^{-1} & \Sigma_{22,1}^{-1} \end{bmatrix}$$

by the inverting partitioned matrices, where $\Sigma_{11,2} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$ and $\Sigma_{22,1} = (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$, and $\Sigma(\theta_W) = \begin{bmatrix} \underbrace{\Sigma_{11}}_{m\times m} & \underbrace{\Sigma_{12}}_{m\times mp} \\ \underbrace{\Sigma_{21}}_{mp\times m} & \underbrace{\Sigma_{22}}_{mp\times mp} \end{bmatrix}$.

Then,

$$\Lambda^*_W = \Gamma(\theta_W)'\Sigma(\theta_W)^{-1} = [G(\theta_W)'\Sigma_{11,2}^{-1} - F(\theta_W)'\Sigma_{22,1}^{-1}\Sigma_{21}\Sigma_{11}^{-1} \quad -G(\theta_W)'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22,1}^{-1} + F(\theta_W)'\Sigma_{22,1}^{-1}],$$

and

$$\begin{aligned}\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Gamma(\theta_W) &= G(\theta_W)'\Sigma_{11,2}^{-1}G(\theta_W) - F(\theta_W)'\Sigma_{22,1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}G(\theta_W) \\ &\quad -G(\theta_W)'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22,1}^{-1}F(\theta_W) + F(\theta_W)'\Sigma_{22,1}^{-1}F(\theta_W) \quad (27)\end{aligned}$$

By the following Woodbury matrix identity for nonsingular matrices A, C, and

$A - BCB'$,

$$(A - BCB')^{-1} = A^{-1} - A^{-1}B(B'A^{-1}B - C^{-1})^{-1}B'A^{-1},$$

we have

$$\Sigma_{11,2}^{-1} = \Sigma_{11}^{-1} - \Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - \Sigma_{22})^{-1}\Sigma_{21}\Sigma_{11}^{-1} = \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22,1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}$$

where the last equality uses the definition of $\Sigma_{22,1}$. Thus, (27) can be written as

$$G(\theta_W)'\Sigma_{11}^{-1}G(\theta_W) + (F(\theta_W) - \Sigma_{21}\Sigma_{11}^{-1}G(\theta_W))'\Sigma_{22,1}^{-1}(F(\theta_W) - \Sigma_{21}\Sigma_{11}^{-1}G(\theta_W)).$$

This completes the proof. □

Proof of Corollary 2

When the model is linear, so that the Jacobian $G(\theta) = G$ does not depend on θ, $F(\theta) = 0$,

$$\Lambda_W^* = [G'\Sigma_{11,2}^{-1} \quad -G'\Sigma_{11,2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}],$$

by Corollary 1 and $\Sigma_{11,2}^{-1}\Sigma_{12}\Sigma_{22}^{-1} = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22,1}^{-1}$. Thus the equation (27) becomes

$$\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Gamma(\theta_W) \;=\; G'\Sigma_{11,2}^{-1}G.$$

Next, we show that $\Sigma_{11,2}(\theta_W)$ doesn't depend on θ_W. Fix an arbitrary $\theta_0 \in \mathbb{R}^p$ and let $\delta = \theta - \theta_0$. Since the model is linear, $g(X_i, \theta) = g(X_i, \theta_0) + G(X_i)\delta$, we have

$$\psi(X_i, \theta) = \psi_i(\theta) = \begin{pmatrix} g(X_i, \theta_0) + G(X_i)\delta \\ vec(G(X_i)') \end{pmatrix} = \begin{pmatrix} I_m & C \\ 0 & I_{mp} \end{pmatrix} \psi_i(\theta_0)$$

where $C = (\delta' \otimes I_m)\, K_{p,m}$ and $K_{p,m}$ is the $mp \times mp$ commutation matrix that satisfies $vec(G(X_i)) = K_{p,m}\, vec(G(X_i)')$ (Abadir and Magnus (2005)). $L(\theta) = \left(\begin{smallmatrix} I_m & C \\ 0 & I_{mp} \end{smallmatrix}\right)$ is block upper-triangular (with identity diagonal blocks) and depends on θ only through C.

Since $\psi_i(\theta) = L(\theta)\psi_i(\theta_0)$, we have $\Sigma(\theta) = L(\theta)\Sigma(\theta_0)L(\theta)'$, and expanding this gives

$$\Sigma_{11}(\theta) = \Sigma_{11}(\theta_0) + C\Sigma_{21}(\theta_0) + \Sigma_{12}(\theta_0)C' + C\Sigma_{22}C',$$

$$\Sigma_{12}(\theta) = \Sigma_{12}(\theta_0) + C\Sigma_{22},$$
$$\Sigma_{22}(\theta) = \Sigma_{22},$$
$$\Sigma_{12}(\theta)\Sigma_{22}^{-1}\Sigma_{21}(\theta) = \Sigma_{12}(\theta_0)\Sigma_{22}^{-1}\Sigma_{21}(\theta_0) + \Sigma_{12}(\theta_0)C' + C\Sigma_{21}(\theta_0) + C\Sigma_{22}C'.$$

Thus,

$$\Sigma_{11,2}(\theta) = \Sigma_{11}(\theta_0) - \Sigma_{12}(\theta_0)\Sigma_{22}^{-1}\Sigma_{21}(\theta_0),$$

and all θ-dependent terms cancel out. Thus, $\Sigma_{11,2}(\theta_W)$ does not depend on W. This completes the proof. □

Proof of Theorem 2

It is straightforward to show that the ME-GMM estimator is asymptotically equivalent to the augmented M-estimator $\widetilde{\theta}(\Lambda)$ with $\Lambda = \Gamma(\theta_W)'\Delta$. Then, the first result in Theorem 2 follows from Theorem 1, since Assumption 1.5 holds by the assumptions of Theorem 2.

The bootstrap ME-GMM estimator $\widehat{\theta}^*_{ME}$ in (20) satisfies the first-order condition,

$$\Gamma_n^*(\widehat{\theta}^*_{ME})'\Delta^*\widetilde{\psi}_n^*(\widehat{\theta}^*_{ME}) = 0$$

where

$$\widetilde{\psi}_n^*(\theta) = \frac{1}{n}\sum_{i=1}^n \widetilde{\psi}^*(X_i^*, \theta) = \begin{pmatrix} g_n^*(\theta) - g_n(\widehat{\theta}_W) \\ vec(G_n^*(\theta)') - vec(G_n(\widehat{\theta}_W)') \end{pmatrix},$$

and $\Gamma_n^*(\theta) = \frac{1}{n}\sum_{i=1}^n \frac{\partial\widetilde{\psi}^*(X_i^*,\theta)}{\partial\theta'} = \frac{1}{n}\sum_{i=1}^n \frac{\partial\psi(X_i^*,\theta)}{\partial\theta'}$. Note that

$$\begin{aligned}\Delta^* &= \left(E^*[\psi(X_i^*, \widehat{\theta}_W)\psi(X_i^*, \widehat{\theta}_W)'] - E^*[\psi(X_i^*, \widehat{\theta}_W)]E^*[\psi(X_i^*, \widehat{\theta}_W)]'\right)^{-1} \\ &= \left(\frac{1}{n}\sum_{i=1}^n \psi(X_i, \widehat{\theta}_W)\psi(X_i, \widehat{\theta}_W)' - (\frac{1}{n}\sum_{i=1}^n \psi(X_i, \widehat{\theta}_W))(\frac{1}{n}\sum_{i=1}^n \psi(X_i, \widehat{\theta}_W))'\right)^{-1}\end{aligned}$$

depends only on the original sample quantities $\widehat{\theta}_W = \widehat{\theta}_{GMM}(W)$.

Expanding $\widetilde{\psi}_n^*(\widehat{\theta}^*_{ME})$ around $\widehat{\theta}_W$, multiplying through by $\sqrt{n}$, and solving gives

$$\sqrt{n}(\widehat{\theta}^*_{ME} - \widehat{\theta}_W) = -(\Gamma_n^*(\widehat{\theta}^*_{ME})'\Delta^*\Gamma_n^*(\tilde{\theta}^*))^{-1}\Gamma_n^*(\widehat{\theta}^*_{ME})'\Delta^*\sqrt{n}\widetilde{\psi}_n^*(\widehat{\theta}_W) \qquad (28)$$

for some $\tilde{\theta}^*$ between $\widehat{\theta}^*_{ME}$ and $\widehat{\theta}_W$. Under Assumption 2, $\tilde{\theta}^* \xrightarrow{p^*} \theta_W$, and thus $\Gamma^*_n(\widehat{\theta}^*_{ME}) \xrightarrow{p^*} \Gamma(\theta_W)$ by the bootstrap WLLN (Assumption 2.2), and WLLN by Assumption 1. Furthermore, we have $\Delta^* = \Sigma_n(\widehat{\theta}_W)^{-1} \xrightarrow{p^*} \Sigma(\theta_W)^{-1}$. So that,

$$(\Gamma^*_n(\widehat{\theta}^*_{ME})'\Delta^*\Gamma^*_n(\tilde{\theta}^*))^{-1}\Gamma^*_n(\widehat{\theta}^*_{ME})'\Delta^* \xrightarrow{p^*} (\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Gamma(\theta_W))^{-1}\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}$$

Next, we consider $\sqrt{n}\widetilde{\psi}^*_n(\widehat{\theta}_W)$ in (28),

$$\sqrt{n}\widetilde{\psi}^*_n(\widehat{\theta}_W) = \sqrt{n}\begin{pmatrix} g^*_n(\widehat{\theta}_W) - g_n(\widehat{\theta}_W) \\ vec(G^*_n(\widehat{\theta}_W)') - vec(G_n(\widehat{\theta}_W)') \end{pmatrix}. \tag{29}$$

Note that each of the two terms has conditional mean zero given $\{X_i\}_{i=1}^n$, since $E^*[g^*_n(\theta)] = g_n(\theta)$, $E^*[G^*_n(\theta)] = G_n(\theta)$. Then, the triangular array CLT under Assumption 2 gives $\sqrt{n}\widetilde{\psi}^*_n(\widehat{\theta}_W) \xrightarrow{d^*} N(0, \Sigma(\theta_W))$. The conclusion then follows by the Slutsky's theorem,

$$\begin{aligned} \sqrt{n}(\widehat{\theta}^*_{ME}(W) - \widehat{\theta}_W) &\xrightarrow{d^*} (\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Gamma(\theta_W))^{-1}\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}N(0, \Sigma(\theta_W)) \\ &= N(0, (\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Gamma(\theta_W))^{-1}). \quad \square \end{aligned}$$

Proof of Corollary 3

The DR estimator $\widehat{\theta}^*_{DR}$ in (23) satisfies the following;

$$A_n(W, \widehat{\theta}_W)\widetilde{\psi}^*_n(\widehat{\theta}^*_{DR}) = 0$$

with $A_n(W, \widehat{\theta}_W)$ instead of $\Gamma^*_n(\widehat{\theta}^*_{ME})'\Delta^*$ as in the $\widehat{\theta}^*_{ME}$ case. Then, it can be similarly shown as in the proof of Theorem 2, that the DR estimator satisfies;

$$\begin{aligned} \sqrt{n}(\widehat{\theta}^*_{DR} - \widehat{\theta}_W) &= -(A(W, \theta_W)\Gamma(\theta_W))^{-1}A(W, \theta_W)\sqrt{n}\widetilde{\psi}^*_n(\widehat{\theta}_W) + o_{p^*}(1) \\ &\xrightarrow{d^*} N(0, V_{GMM}(W, \theta_W)) \end{aligned}$$

which is the same asymptotic distribution as the GMM estimator, with asymptotic variance $V_{GMM}(W, \theta)$ in (7).

$\square$

Proof of Theorem 3

By construction, for any finite n, the following holds,

$$\begin{aligned}\sup_{(W,(F,\theta))\in\mathcal{G}^\Gamma} E_F\left\{l\left[\sqrt{n}(T_n-\theta_1)\right]\right\} &= \sup_{W\in\mathcal{W}}\sup_{(F,\theta)\in\Gamma^W} E_F\left\{l\left[\sqrt{n}(T_n-\theta_1)\right]\right\} \\ &\geq \sup_{(F,\theta)\in\Gamma^{\widetilde{W}}} E_F\left\{l\left[\sqrt{n}(T_n-\theta_1)\right]\right\} \qquad (30)\end{aligned}$$

for any arbitrary element $\widetilde{W}\in\mathcal{W}$. Since the inequality (30) holds for all n, we have that

$$\liminf_{n\to\infty}\sup_{(W,(F,\theta))\in\mathcal{G}^\Gamma} E_F\left\{l\left[\sqrt{n}(T_n-\theta_1)\right]\right\} \geq \liminf_{n\to\infty}\sup_{(F,\theta)\in\Gamma^{\widetilde{W}}} E_F\left\{l\left[\sqrt{n}(T_n-\theta_1)\right]\right\} \qquad (31)$$

Now, we fix $W=\widetilde{W}\in\mathcal{W}$. Then, the remaining proof similarly follows the argument of Chamberlain (1987). Conditions (i)-(iv) of $\mathcal{F}^W$ are exactly condition (C0) of Chamberlain (1987, Section 2) for the moment functions $\widetilde{\psi}(X,\theta)$. Then, we can construct the multinomial distribution G_0 which is arbitrarily close to (F_0,θ_W) that satisfies the conditions in $\mathcal{F}^W$, and $(G_0,\theta_W)\in\Gamma^W$. To be more specific, by Lemma 3 of Chamberlain (1987), there exists a probability measure G_0 whose support is a finite set in the support of F_0 with probabilities $p=(p_1,\cdots,p_r)$ on the support points, which satisfies the moment condition $\sum_{j=1}^r p_j\psi(X_j,\theta_W)=E_{F_0}[\psi(X_i,\theta_W)]$. Then, we apply the local asymptotic minimax bounds for the multinomial distribution (Lemma 2 in Chamberlain (1987)).

By Theorem 2 of Chamberlain (1987) for the moment function $\widetilde{\psi}(X,\theta)$ and the pairs (F,θ) that satisfy the regularity conditions the space of probability distributions $\mathcal{F}^{\widetilde{W}}$, (31) is bounded below by the following,

$$\int_{-\infty}^{\infty} l(\sigma_{\widetilde{W}}u)d\Phi(u),$$

which is the expected loss of $\mathcal{N}(0,\sigma^2_{\widetilde{W}})$, where $\sigma^2_{\widetilde{W}}$ is the (1,1) element of

$$\begin{aligned}&\left(\left[E_{F_0}\frac{\partial\widetilde{\psi}(X,\theta_{\widetilde{W}})}{\partial\theta'}\right]'\left[E_{F_0}\widetilde{\psi}(X,\theta_{\widetilde{W}})\widetilde{\psi}'(X,\theta_{\widetilde{W}})\right]^{-1}\left[E_{F_0}\frac{\partial\widetilde{\psi}(X,\theta_{\widetilde{W}})}{\partial\theta'}\right]\right)^{-1} \\ =\ &(\Gamma(\theta_{\widetilde{W}})'\Sigma(\theta_{\widetilde{W}})^{-1}\Gamma(\theta_{\widetilde{W}}))^{-1}.\end{aligned}$$

Since this bound holds for any choice of $\widetilde{W} \in \mathcal{W}$, we have

$$\liminf_{n\to\infty} \sup_{(W,(F,\theta))\in\mathcal{G}^\Gamma} E_F\left\{l\left[\sqrt{n}(T_n-\theta_1)\right]\right\} \geq \sup_W \int_{-\infty}^{\infty} l(\sigma_W u)d\Phi(u) = \int_{-\infty}^{\infty} l(\sup_W \sigma_W u)d\Phi(u)$$

where the last equality holds by the symmetry, monotonicity of the $l(\cdot)$, and $0 < \sigma_W \leq \sup_W \sigma_W < \infty$, and by the Monotone Convergence Theorem. □

References

Abadir, K. M., and Magnus, J. R. (2005). Matrix Algebra. *Cambridge University Press*.

Andrews, D. W. K., and Cheng, X. (2012). Estimation and inference with weak, semi-strong and strong identification. *Econometrica*, 80, 2153-2211.

Andrews, D. W. K., and Kwon, S. (2024). Misspecified Moment Inequality Models: Inference and Diagnostics. *The Review of Economic Studies*, 91(1), 45-76.

Andrews, I. (2019). On the Structure of IV Estimands. *Journal of Econometrics*, 211 (1), 294-307.

Andrews, I., Barahona, N., Gentzkow, M., Rambachan, A., and Shapiro, J. M. (2025). Structural estimation under misspecification: Theory and implications for practice. *Quarterly Journal of Economics*, 140(3):1801-1855.

Andrews, I., Barnhard, H., and Carlson, J. (2024). True and pseudo-true parameters. *Working Paper*.

Andrews, I., Chen, J., and Tecchio, O. (2025). The Purpose of an Estimator is What it Does. *Econometric Society 2025 World Congress Volume*.

Andrews, I., Gentzkow, M., and Shapiro, J. M. (2017). Measuring the sensitivity of parameter estimates to estimation moments. *Quarterly Journal of Economics*, 132(4), 1553-1592.

Angrist, J. D., and Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity, *Journal of the American Statistical Association*, 90, 431-442.

Angrist, J. D., and Krueger, A. B. (1991). Does Compulsory School Attendance Affect Schooling and Earnings?, *The Quarterly Journal of Economics*, 106(4), 979-1014.

Angrist, J. D., and Krueger, A. B. (1995). Split-sample instrumental variables estimates of the return to schooling. *Journal of Business and Economic Statistics*, 13(2), 225-235.

Arellano, M., and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58(2), 277-297.

Armstrong, T. B., and Kolesár, M. (2021). Sensitivity analysis using approximate moment condition models. *Quantitative Economics*, 12(3), 705-740.

Armstrong, T. B., Kline, P., and Sun, L. (2025). Adapting to Misspecification. *Econometrica*, 93 (6), 1981-2005.

Blandhol, C., Bonney, J., Mogstad, M., and Torgovitsky, A. (2025). When is TSLS Actually LATE?. *Working Paper*

Bonhomme, S., and Weidner, M. (2022). Minimizing sensitivity to model misspecification. *Quantitative Economics*, 13(3), 907-954.

Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90 (430), 443-450.

Brown, B. W., and Newey, W. K. (2002). Generalized method of moments, efficient bootstrapping, and improved inference. *Journal of Business and Economic Statistics*, 20(4), 507-517.

Card, D. (1995). Using Geographic Variation in College Proximity to Estimate the Return to Schooling, *in Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, ed. by L. N. Christofides, K. E. Grant, and R. Swidinsky, Toronto: University of Toronto Press, 201-222.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3), 305-334.

Christensen, T., and Connault, B. (2023). Counterfactual sensitivity and robustness. *Econometrica*, 91(1), 263-298.

Conley, T.G., Hansen, C.B., and Rossi, P.E. (2012). Plausibly exogenous, *The Review of Economics and Statistics*, 94(1), 260-272.

Dovonon, P, and Gonçalves, S. (2017). Bootstrapping the GMM overidentification test under first-order underidentification. *Journal of Econometrics*, 201, 43-71.

Dovonon, P., and Renault, E. (2013). Testing for common conditionally heteroskedastic factors. *Econometrica*, 81, 2561-2586.

Giné, E., and Zinn, J. (1990). Bootstrapping general empirical measures, *The Annals of Probability*, 18(2), 851-869.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, 31(4), 1208-1211.

Gospodinov, N., Kan, R., and Robotti, C. (2014). Misspecification-robust inference in linear asset-pricing models with possibly nonlinear pricing kernels. *Review of Financial Studies*, 27(7), 2139-2170.

Hahn, J. (1996). A note on bootstrapping generalized method of moments estimators. *Econometric Theory*, 12(1), 187-197.

Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. In Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, *University of California Press*, 175-194.

Hall, A. R., and Inoue, A. (2003). The large sample behavior of the generalized method of moments estimator in misspecified models. *Journal of Econometrics*, 114(2), 361-394.

Hall, A. R., and Pelletier, D. (2011). Nonnested testing in models estimated via generalized method of moments. *Econometric Theory*, 27(2), 443-456.

Hall, P., and Horowitz, J. L. (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica*, 64(4), 891-916.

Hansen, B.E., and Lee, S. (2021). Inference for Iterated GMM Under Misspecification. *Econometrica*, 89, 1419-1447

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4), 1029-1054.

Hansen, L. P., and Jagannathan, R. (1997). Assessing specification errors in stochastic discount factor models. *Journal of Finance*, 52(2), 557-590.

Hwang, J., Kang, B., and Lee, S. (2022). A Doubly Corrected Robust Variance Estimator for Linear GMM. *Journal of Econometrics*, 229(2), 276-298.

Imbens, G. W. (1997). One-step estimators for over-identified generalized method of moments models. *Review of Economic Studies*, 64(3), 359-383.

Kan, R., and Robotti, C. (2009). Model Comparison Using the Hansen-Jagannathan Distance. *The Review of Financial Studies*, 22(9), 3449-3490.

Kleibergen, F. (2005). Testing Parameters in GMM Without Assuming that They Are Identified. *Econometrica*, 73(4), 1103-1123.

Kleibergen, F., and Zhan, Z. (2025). Double robust inference for continuous updating GMM, *Quantitative Economics*, 16(1), 295-327.

Kolesár, M. (2013). Estimation in an Instrumental Variables Model With Treatment Effect Heterogeneity, *Working Paper*.

Lee, J.H., and Liao, Z. (2018). On Standard Inference for GMM with Local Identification Failure of Known Forms. *Econometric Theory*, 34, 790-814.

Lee, S. (2014). Asymptotic refinements of a misspecification-robust bootstrap for generalized method of moments estimators, *Journal of Econometrics*, 178(3), 398-413.

Marmer, V., and Otsu, T. (2012). Optimal comparison of misspecified moment restriction models under a chosen measure of fit. *Journal of Econometrics*, 170(2), 538-550.

Mogstad, M., and Torgovitsky, A. (2024). Instrumental Variables with Unobserved Heterogeneity in Treatment Effects. In *Handbook of Labor Economics*, Volume 5, forthcoming.

Mogstad, M., Torgovitsky, A., and Walters, C. (2021). The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables. *American*

Economic Review, 111(11), 3663-3698.

Mukhin, Y. (2019). Sensitivity of Regular Estimators. *Working Paper*.

Müller, U. K. (2013). Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix. *Econometrica*, 81(5), 1805-1849.

Newey, W. K., and McFadden, D. (1994). Large sample estimation and hypothesis testing. *In Handbook of Econometrics*, Vol. 4, 2111-2245.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society: Series B*, 31(2), 350-371.

Rivers, D., and Vuong, Q. (2002). Model selection tests for nonlinear dynamic models. *The Econometrics Journal*, 5(1), 1-39.

Słoczyński, T. (2024). When Should We (Not) Interpret Linear IV Estimands as LATE?, *Working Paper*.

Stock, J.H., and Wright, J.H. (2000). GMM with weak identification, *Econometrica*, 68, 1055-1096.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1), 1-25.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators, *Journal of Econometrics*, 126(1), 25-51.

Windmeijer, F., and Santos Silva, J.M.C. (1997). Endogeneity in count data models: An application to demand for health care, *Journal of Applied Econometrics*, 12(3), 281-294.