

Efficient GMM and Weighting Matrix under Misspecification

Byunghoon Kang*

May 7, 2026

Abstract

This paper develops efficient GMM estimation when the moment conditions are misspecified. We observe that the influence function of the standard GMM estimator under misspecification depends on both the original moment conditions and their Jacobian, motivating a new class of estimators based on augmented moment conditions with recentering. The standard GMM estimator is a special case within this class, 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 pseudo-true value. In linear models, the asymptotic variance of ME estimator reduces to the textbook efficient-GMM variance formula $(G'W^*G)^{-1}$, where W^* is the inverse of the variance of residualized moments after projection on the Jacobian G . We consider a feasible double-recentered bootstrap estimator, which can be considered as a misspecification-robust and efficient version of Hall and Horowitz (1996) recentered bootstrap GMM estimator, and also consider a split-sample ME estimator. Finally, we establish uniform local asymptotic minimax bounds over a class of weighting matrices. We illustrate the proposed methods in simulation and empirical examples.

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

JEL classification: C12, C13, C14, C26.

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 weight matrix. Under the assumption of the correct model specification when there exists a unique true parameter that satisfies the population moment conditions exactly equal to zero, the efficient weighting matrix is the inverse of the variance-covariance matrix of the moments. This choice achieves the smallest asymptotic variance among standard GMM estimators (Hansen (1982), Chamberlain (1987)), and researchers often use this efficient weight in practice,

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

for example, an inverse of the variance of the instruments in the instrumental variables (IV) setup, or a Heteroskedasticity and Autocorrelation Consistent (HAC) matrix estimator.

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 are misspecified or treatment effect heterogeneity makes no parameter satisfies all moment equations exactly. The presence of misspecification has significant implications for the properties of estimators, potentially affecting asymptotic distribution, and consequently, the validity of standard inference procedures. Recognizing this challenge, a growing literature studies the inference for the pseudo-true value, defined as the population minimizer of the objective functions, including Hall and Inoue (2003), Müller (2013), Hansen and Lee (2021), and Andrews and Kwon (2024) among others.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 Techhio (2025).

More importantly, misspecification fundamentally changes the role of the weighting matrix, and has different implications than 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 corresponds to different target parameters.² This dual role of the weighting matrix under misspecification raises a number of questions that the standard GMM literature 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 of all, by focusing on the fixed weighting matrix W and the associated pseudo-true parameter θ_W , we consider the class of M-estimator where the original moment functions and the Jacobian moments are “stacked” to form a vector of augmented moment conditions with recentering. This is based on the novel observation that the “standard” GMM estimators under misspecification can be

¹Defining the pseudo-true value as minimizers of population criterion has a long tradition in econometrics, including White (1982), Hall and Inoue (2003), Müller (2013), Hansen and Lee (2021), etc. 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).

²We are well aware that pseudo-true values, in general, may be possibly different from quantities of economic interest (Müller (2013), Andrews, Barnhard, and Carlson (2024)). The main focus of this paper is not on the identification and the causal interpretation of the widely used GMM/TSLs estimators, but on the valid and improved inference for the same GMM/TSLs estimand under moment misspecification. We refer the readers to other influential papers (and references therein) such as Kolesár (2013), Mogstad et al. (2021), Blandhol et al. (2025), and Słoczyński (2024) for advances in this direction in the linear IV model under treatment effect heterogeneity. See also Andrews, Barahona, et al. (2025) for the potentially misspecified general structural model.

expressed as an influence function of the augmented moments. The proposed M-estimator is a general class of estimators that include the “standard” misspecified-GMM as a special case. We then define the “efficient estimator” within this class, which we label as a “misspecification-efficient” (ME) estimator, that has the smallest asymptotic variance that 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; while the standard misspecified-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 its construction requires the population re-centering of the original and Jacobian moments. However, our asymptotic result of the oracle-ME estimator is constructive in the sense that the variance of ME estimator can be easily estimated (without knowing these unknown quantities) and reported as a measure of efficiency frontier; showing researchers what could be gained by imposing an additional information from the misspecification and identification strength. Furthermore, our results provide new insights into the conventional (correct-specification) optimal weighting matrix and variance of the efficient GMM. Under the linear model, the asymptotic variance of the ME estimator simplifies to $V_{ME}(W) = (G'\Sigma_{11,2}^{-1}G)^{-1}$, where G is Jacobian, $\Sigma_{11,2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, and $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ is the variance of the augmented moments. The weighting matrix $W^* = \Sigma_{11,2}^{-1}$ is the inverse of the variance of residualized moment after projection on the Jacobian, while the conventional (correct-specification) efficient weighting matrix $W_0 = \Sigma_{11}^{-1}$ is the inverse of the variance of the original moments.

This variance expression in linear models has two notable features. First, $V_{ME}(W)$ is invariant to W ; under misspecification with linear moments, different W targets different θ_W , but does not affect how precisely it can be estimated. Second, it coincides with the textbook “efficient GMM” formula $(G'W^*G)^{-1}$ with $W^* = \Sigma_{11,2}^{-1}$, and our result provide a different angle to the well-known downward bias of the efficient-GMM standard errors. In practice, researchers have found that the standard error of the two-step efficient GMM is often severely downward biased, motivating the finite-sample correction of Windmeijer (2005). We show that, with the choice of $W^* = \Sigma_{11,2}^{-1}$ in the linear GMM, the “conventional” efficient GMM formula will be valid for the oracle ME estimator, but it is too “small” for the standard GMM estimator under misspecification, even in large samples. This further echoes the use of the misspecification-robust standard errors (SE), instead of the conventional formula (Andrews, Chen and Techhio (2025)).³ We also recommend to report the efficiency frontier $V_{ME}(W) = (G'\Sigma_{11,2}^{-1}G)^{-1}$ alongside with the misspecification-robust standard errors as a transparent efficiency measure, without need to commit to a specific W and specific target parameter θ_W ; perhaps it was reported anyway with an incorrect labelling of “conventional-correct specification” SE of the two-step efficient GMM when $\Sigma_{11,2} \approx \Sigma_{11}$. In general nonlinear model, however, above arguments fail to hold, and $V_{ME}(W)$ varies with W and no longer have a

³Hwang, Kang, and Lee (2022) further show that, in linear models, the misspecification-robust SE also works as a finite-sample corrections even under correct specification.

textbook formula.

We propose three feasible procedures for the oracle-ME estimator. The first is a bootstrap ME-GMM estimator 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, with the efficient choice $W^* = \Sigma_{11,2}^{-1}$. We show that the asymptotic distribution of the bootstrap ME-GMM estimator is equal to those of the oracle ME estimator, which allows us to approximate the sampling variation of the oracle-ME estimator and the efficiency bounds considered in this paper. The proposed bootstrap estimator can be considered as a misspecification-robust and misspecification-efficient version of Hall and Horowitz (1996) recentered bootstrap GMM estimator, which is only valid under correct specification. We also consider a closely related “double recentered” (DR) GMM bootstrap that jointly perturbs the moment conditions and the Jacobian similarly to the FOC of the misspecified GMM, and show that the bootstrap distribution mimics the limiting distribution of the standard GMM estimators regardless of model specification. Under correct specification the additional Jacobian recentering correction is higher-order and the DR bootstrap nests the original Hall and Horowitz (1996) procedure as a special case; under misspecification it correctly approximates the Jacobian variation term that the Hall and Horowitz (1996) bootstrap omits. Finally, we also develop a repeated sample-splitting ME estimator, in the spirit of Angrist and Krueger (1995), that uses a held-out half of the sample to estimate the recentering quantities. In the linear model with $W = \Sigma_{11,2}^{-1}$, the moment recentering term drops out by the first-order condition, so only the Jacobian recentering term needs to be estimated.

The recentering has been considered important when applying the bootstrap in the over-identified GMM model to achieve higher-order improvements under correct specification (Hall and Horowitz (1996) and Brown and Newey (2002)). In the linear IV example, the proposed bootstrap estimators can be seen as a linear combinations of the HH recentered GMM/TSLs estimator and the Jacobian estimator, where the recentering of the Jacobian estimator ensure that the bootstrap expected Jacobian matrix evaluated the estimator is zero. Lee (2014) argues that recentering in the standard nonparametric GMM bootstrap can be detrimental and is not even needed if we use the analytic misspecification-robust variance estimator. In our paper, however, we clarify this argument that additional considerations of recentered Jacobian can also achieve asymptotic validity under misspecification. This “double recentering” has been also considered in weak-identification literature to accurately mimic the behaviour of the original Jacobian under weak identification (e.g., Dovonon and Goncalves (2017), Lee and Liao (2018)), yet under correct specification.

In this paper, we view the choice of W as a different candidate models/designs, and we do not evaluate the performances of different weighting matrix W and do not aim to choose the optimal W (and thus θ_W) under certain statistical criteria. The weighting matrix W and the candidate class of $\mathcal{W}$ can be possibly chosen by the researchers a priori, but in general, it may be difficult for researchers view one candidate pseudo-true value better than another, even if it is based on statistically well-defined criteria. For example, we cannot tell that a more precisely estimated LATE (local average treatment effect) for one subpopulation is better than a noisier LATE for different

subpopulation; they may answer different causal questions.

In this paper, we develop uniform semiparametric efficiency bounds over $W \in \mathcal{W}$ with the augmented system, which exhausts all additional information from the direction of misspecification and Jacobian, taking into account the correlation between them. We want to highlight that the notion of semiparametric efficiency would be defined differently under (global) misspecification. The pseudo-true value θ_W changes with W under misspecification, but the semiparametric efficiency framework typically assumes a true-parameter and the efficiency bound is computed at the well-defined single target parameter. We first focus on fixed W and consider the semiparametric efficiency bounds for θ_W , and consider uniform bounds over $W \in \mathcal{W}$.

Our uniform local asymptotic minimax bounds will be not only useful for a researcher who chooses W based on target a specific economic parameter and consider only θ_W , but also useful for a researcher who is completely agnostic to the choice of $W \in \mathcal{W}$. Many applied researchers may just pick one W (e.g., TSLS), however, applied researcher who runs GMM, already considers choosing W over a class of different weighting matrices $\mathcal{W}$ routinely; one-step $W = I$ or $(Z'Z)^{-1}$, two-step, and iterated until convergence. Under misspecification each iteration changes the pseudo-true value. So when results differ across one-step, two-step, and iterated GMM, it can be a direct evidence of sensitivity to W .⁴

We consider Monte Carlo simulations and three empirical applications to illustrate the proposed methods. The simulation results confirm that the oracle ME is meaningfully tighter, and the DR percentile CI delivers valid coverage, with similar or slightly shorter lengths than the misspecification-robust CI. Repeated sample splitting estimator also performs well, with slightly over coverage and larger lengths as expected. In both of Card (1995) and Angrist and Krueger (1991) returns to schooling examples, the ME efficiency frontier is roughly 10%-20% smaller than the conventional and misspecification-robust standard errors when using efficient weighting matrix $W^* = \Sigma_{11,2}^{-1}$ or $W_0 = \Sigma_{11}^{-1}$. From extensive simulations, and all of the three applications, we found that the ME efficiency bound is most informative when (i) W is efficient weight (either W_0 or W^*) and far from $W = I$ or other naive choices; (ii) identification strength is moderate, so that Jacobian sampling variation is non-negligible. The identity matrix ($W = I$) is a poor default in our simulations and empirical examples, as the estimates are very different from using efficient weightings and the gap between conventional/robust SE with the oracle ME bound is large (20-40% in Angrist and Krueger (1991) example). Under W^* with moderate identification strength, the gap is typically below 10%, and the misspecification-robust SE is close to the efficiency frontier. Overall, we find that the ME efficiency bounds can be useful, and the proposed DR bootstrap and sample-split procedures provide reasonable alternatives to the standard GMM with analytic misspecification-robust standard errors.

Our paper complements a growing literature on sensitivity analysis and robust inference under misspecification, such as Andrew, Gentzkow and Shapiro (2017), Armstrong and Kolesár (2021),

⁴Another example is Kan and Robotti (2009) and Gospodinov, Kan, and Robotti (2014) who explicitly study inference under misspecification using Hansen-Jagannathan (1997)'s distance measure, and W denotes different asset portfolios in that context.

Bonhomme and Weidner (2022), and Christensen and Connault (2023), among many others. More recent work focuses on the relationship between the pseudo-true value and the true value under misspecification, e.g., Andrews, Barnhard, and Carlson (2024), Andrews, Barahona et al. (2025), and also see Andrews, Chen, and Techhio (2025) for a summary of recent developments in the literature with some constructive recommendations for practitioners. While these papers focus on bias, sensitivity, or optimal inference, this paper analyse efficient estimation in the GMM setup under misspecification, and we provide a feasible efficiency measure that researchers can report alongside robust-standard errors.

There are some limitations to our paper. First, we do not consider two-step and the iterated GMM estimators, and we focus only on the one-step GMM estimator. The textbook motivation of going beyond the one-step GMM estimator is to achieve efficiency, and existing papers (e.g., Hall and Inoue (2003), Hwang, Kang and Lee (2022), and Hansen and Lee (2021)) uses the “standard” optimal weighting-matrix W_0 , which we show in this paper suboptimal under misspecification. It is not clear why we use the “standard” optimal weighting matrix in each iteration under misspecification. It would be interesting to see the behavior of the two-step/iterated GMM using other weighting matrix schemes, for example, $W^* = \Sigma_{11,2}^{-1}(\theta)$ considered in our paper.

Second, our results are constructed under a standard full rank assumption of the Jacobian, and we do not consider the weak-identification settings, where the behavior of the GMM estimators may be nonstandard (e.g., Stock and Wright (2000), Andrews and Cheng (2012), Dovonon and Renault (2013)). While, in this paper, using a Jacobian of the moment condition comes from the influence function of the GMM under misspecification, there are some papers utilize a Jacobian estimation in the weak identification or the lack of first-order local identification in GMM literature. Kleibergen (2005) uses Jacobian estimator that is asymptotically uncorrelated with the original sample moments to develop test statistic that is valid without assuming identification. Lee and Liao (2018) uses zero Jacobian moment conditions, which holds when the rank condition fails to hold, as extra moment conditions in addition to the original moment conditions. Extension of our approach that allow for weak identification would be important, but more challenging. We leave these for future research.

The remainder of the paper is organized as follows. Section 2 introduces misspecified GMM setup. Section 3 develops the augmented moment framework and the ME estimator, and derives its asymptotic distribution. Section 4 analyses the feasible bootstrap estimators - ME-GMM and the double recentered (DR) bootstrap procedure. Section 5 considers split-sample ME estimator as an alternative. Section 6 provides the uniform semiparametric efficiency bounds over classes of weighting matrices. Section 7 reports the Monte Carlo evidence and Section 8 shows illustrative empirical examples for the proposed methods. Section 9 concludes. Appendix A includes all proofs of the results in the main paper. Appendix B considers worst-case variance bounds for sensitivity analysis over misspecification and identification strength in the spirit of Conley, Hansen, and Rossi (2012). Appendix C reports additional simulations. Appendix D contains the dynamic-panel application.

1.1 Notation

$\|A\|$ denotes the spectral norm. $A \geq B$ denotes that the $A - B$ is positive semi-definite for real symmetric matrix 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.

We introduce further notations 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 $\mathbb{R}^m$ with the support of a probability measure $\mathcal{X}$. Given an $F \in \Pi$, the basic neighborhoods of F are sets of the form:

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

where $\epsilon_j > 0$, q is some positive integer, and $f_j : \mathcal{X} \rightarrow \mathbb{R}$ 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. We next define a class of loss function $l \in \mathcal{L}$, if $l(\cdot) \in \mathcal{L}$ satisfies the following conditions, if for all $u, v \in \mathbb{R}$: (i) $l(u) = l(|u|)$, (ii) $|u| \leq |v|$ implies $l(u) \leq l(v)$, (iii) $\int_{-\infty}^{\infty} l(u) \exp(-\frac{1}{2}\lambda u^2) du < \infty$ for $\lambda > 0$, (iv) $l(0) = 0$. Many standard loss function satisfy these conditions, for example, quadratic $l(u) = u^2$, absolute function $l(u) = |u|$, and polynomial loss $l(u) = |u|^p$ for $p > 0$.

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),

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

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

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

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

$$G(\theta_W)'Wg(\theta_W) = 0 \quad (3)$$

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

$$g(\theta_0) = E[g_n(\theta_0)] = 0, \quad (4)$$

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), and that the varying W changes the parameter being estimated.

In this paper, our object of interest is on providing 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 novelty of this paper is to explicitly find the influence function of the “standard” GMM estimators under misspecification as a set of augmented moment conditions which include (i) an original moment restriction and (ii) a Jacobian of the moment function, (iii) with recentering. This result, while implicit in the form of the asymptotic variance of the misspecified-GMM in the literature, further suggests a new class of estimators based on the augmented moments. We show in the following sections that the proposed class of estimators include the “standard” GMM as a special case, and we can consider more efficient estimator within the class that has the smallest asymptotic variance that are consistent for the *same* pseudo-true value θ_W .

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

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

where

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

Based on (5), we can derive an asymptotic distribution of the GMM estimators under misspecification,

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

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

There exist a few papers in the literature, which consider over-identified GMM models as a just-identified GMM estimator with an augmented parameter vectors. One can generally interpret the overidentified GMM models as a just-identified GMM (Newey and McFadden (1994)), and Imbens (1997) finds the connection with the misspecification-robust variance from the just-identified GMM estimator with an augmented parameter vector. 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.

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

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

M-estimator based on the sample analogs of augmented moment conditions

$$\Lambda \tilde{\psi}_n(\tilde{\theta}(\Lambda)) = 0 \quad (8)$$

where $\tilde{\psi}_n(\theta) = \frac{1}{n} \sum_{i=1}^n \tilde{\psi}(X_i, \theta)$. 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 $\tilde{\theta}(\Lambda)$.

- Assumption 1.** 1. The observations $\{X_i\}_{i=1}^n$ are independent and identically distributed (i.i.d.).
2. For $W > 0$, $\theta_W = \arg \min_{\theta} g(\theta)'Wg(\theta)$ is unique, and θ_W is in the interior of the compact parameter space $\Theta \subset \mathcal{R}^p$, 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 , and $E[\|g(X, \theta_W)\|^2] < \infty$, $E[\|G(X, \theta_W)\|^2] < \infty$.
5. $\Lambda E[\tilde{\psi}(X_i, \theta)] \neq 0$ for $\theta \neq \theta_W$. $\Lambda \Gamma(\theta_W)$ exists and is nonsingular.

Assumption 1.3 ensures uniform convergence of the $\tilde{\psi}_n(\theta)$ and the population objective function $E[\tilde{\psi}_n(\theta)]$ to show consistency. Assumption 1.4 is for the asymptotic normality of the normalized sum of $\tilde{\psi}_n(\theta_W)$. The first part of Assumption 1.5 is an identification condition. Necessary and sufficient condition for (local) identification is that $\Lambda \Gamma(\theta_W) = \Lambda_1 G(\theta_W) + \Lambda_2 F(\theta_W)$ has a full column rank. 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 $\Lambda_2 F(\theta_W)$. More primitive conditions on the identification assumption can be considered. For example, when the moment functions are linear in θ , this reduces to the condition that $\Lambda \Gamma(\theta_W) = \Lambda_1 G$ has a 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,

$$\sqrt{n}(\tilde{\theta}(\Lambda) - \theta_W) \xrightarrow{d} N(0, V_M(\Lambda, \theta_W)). \quad (9)$$

where

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

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

Theorem 1 shows that $\tilde{\theta}(\Lambda)$ is consistent for the same GMM pseudo-true value θ_W , and the augmented-M estimator is a sufficiently general class of estimator 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), $\tilde{\theta}(\Lambda)$ has the same influence function representation with the original misspecified-GMM estimator $\hat{\theta}_{GMM}(W)$ 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 estimator, through the following corollaries, as we can consider “efficient” choice of Λ to minimize the asymptotic variance $V_M(\Lambda, \theta_W)$. By using the standard theory of optimal estimating equations (Godambe (1960)) or semi-parametric efficiency of GMM, the optimal choice of Λ takes the following form;⁵

$$\begin{aligned}\Lambda_W^* &= \left(\frac{\partial}{\partial \theta'} E[\tilde{\psi}(X_i, \theta)] \right)' \big|_{\theta=\theta_W} \text{Var}[\tilde{\psi}(X_i, \theta_W)]^{-1} \\ &= \Gamma(\theta_W)' \Sigma(\theta_W)^{-1}.\end{aligned}\tag{10}$$

With this 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 $\tilde{\theta}(\Lambda)$ is constructed with Λ_W^* (or any consistent estimator of Λ_W^*), we label it as the *misspecification-efficient* (ME) estimator, $\hat{\theta}_{ME}(W) = \tilde{\theta}(\Lambda_W^*)$.

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

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

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 $\Sigma(\theta_W) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$.

⁵When $\Sigma(\theta_W)$ 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 $\text{vec}(G(X_i, \theta))$ has an overlap with the $g(X_i, \theta)$, 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.

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

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

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

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

Thus, using the extra moment conditions from Jacobian does not worsen the asymptotic efficiency of estimators even under correct specification.⁶ The ME estimator provides no efficiency gain when $F_G(\theta_W) = F(\theta_W) - \Sigma_{21} \Sigma_{11}^{-1} G(\theta_W) = 0$. In the linear IV model, where $F(\theta_W) = 0$ and G is constant, 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 $G(X_i, \theta)$.

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

⁶Under correct specification, $E[g(X_i, \theta_0)] = 0$, straightforward calculation shows that the asymptotic variance of the “standard” GMM, $V(W, \theta_0)$ in (7) reduces to the classical one $V(W, \theta_0) = (G(\theta_0)' W G(\theta_0))^{-1} G(\theta_0)' W \Sigma_{11}(\theta_0) W G(\theta_0) (G(\theta_0)' W G(\theta_0))^{-1}$, where $\Sigma_{11}(\theta) = E[g(X_i, \theta) g(X_i, \theta)'] - E[g(X_i, \theta)] E[g(X_i, \theta)]'$ is the $m \times m$ upper-left submatrix of $\Sigma(\theta)$. Then, the “standard” optimal-weight matrix is $W_0 = \Sigma_{11}(\theta_0)^{-1}$ and the asymptotic variance of $\hat{\theta}_{GMM}(\Sigma_{11}(\theta_0)^{-1})$ achieves smallest asymptotic variance in the class of (correctly specified) GMM estimator in the sense that $V(W, \theta_0) \geq V(\Sigma_{11}(\theta_0)^{-1}, \theta_0) = (G(\theta_0)' \Sigma_{11}(\theta_0)^{-1} G(\theta_0))^{-1}$ for any non-singular matrix $W > 0$.

⁷To 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, $\Sigma_{12} = E[Z_i Z_i' X_i \varepsilon_i] - E[g(X_i, \theta)] E[g(X_i, \theta)]' = E[\sigma_{\varepsilon X} Z_i Z_i'] - \sigma_{\varepsilon Z} E[X_i Z_i']$. Under correct specification ($\sigma_{\varepsilon Z} = 0$), $\Sigma_{12} = E[\sigma_{\varepsilon X} Z_i Z_i']$ is generally nonzero whenever there is endogeneity, $\sigma_{\varepsilon X} \neq 0$. Under misspecification ($\sigma_{\varepsilon Z} \neq 0$), Σ_{12} is generally nonzero unless there exists specific structure of the higher moments of Z_i so that $E[\sigma_{\varepsilon X}(Z_i) Z_i Z_i'] = \sigma_{\varepsilon Z} \pi' E[Z_i Z_i']$.

Λ_W^* and the asymptotic variance of ME estimator.

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

$$\Lambda_W^* = [G' \Sigma_{11,2}^{-1} \quad -G' \Sigma_{11,2}^{-1} \Sigma_{12} \Sigma_{22}^{-1}] \quad (15)$$

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

$$V_{ME}(W) = (G' \Sigma_{11,2}^{-1} G)^{-1}, \quad (16)$$

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 $W^* = \Sigma_{11,2}^{-1}$ 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 $g(X_i, \theta) - \Sigma_{12} \Sigma_{22}^{-1} \text{vec}(G(X_i)')$, 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 GMM estimator with the conventional efficient weighting matrix W_0 when $\Sigma_{12} \neq 0$ as we discussed.

Corollary 2 provides new insight into the conventional (correct-specification) efficient GMM variance formula. Corollary 2 shows that the conventional efficient GMM variance formula $(G' W G)^{-1}$ 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 conventional formula for efficient GMM is too “small”, even in large samples, for the standard GMM estimator, as it will be valid for the ME estimator. Researchers have found that the standard errors of the efficient GMM are often severely downward biased.⁸ Our results show that reporting the conventional efficient GMM variance formula in the linear GMM will still be useful as an efficiency frontier - showing researchers what could be gained by imposing moments from the Jacobian $G(X_i, \theta)$; but not to be used for an inference directly with the efficient GMM itself under misspecification. When there is no efficiency gain $\Sigma_{12} = 0$, $W^* = W_0 = \Sigma_{11}^{-1}$, using the misspecification-efficient weighting matrix reduces to the conventional case.

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

⁸Many researchers thus routinely used the finite-sample correction of Windmeijer (2005). Hwang et al. (2022) further shows, 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.

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 $r_i(\theta) = g_i(\theta) - \widehat{\Sigma}_{12}\widehat{\Sigma}_{22}^{-1}vec(G_i(\theta)')$ and the preliminary GMM estimator $\widehat{\theta} = \widehat{\theta}_{GMM}(W)$.⁹

Furthermore, Corollary 2 shows that under the linear model, $V_{ME} = (G'\Sigma_{11,2}^{-1}G)^{-1}$ is same for all different pseudo-true values θ_W , due to the invariance property of $\Sigma_{11,2}$. This implies that the 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 or pseudo-true value.

In general nonlinear settings, above arguments fail to hold as the Jacobian $G(X_i, \theta)$, and $\Sigma_{11,2}(\theta)$ changes 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 methods (will be discussed in the next Section 4). This bootstrap method can be viewed as a misspecification-robust and efficient version of the Hall and Horowitz (1996) recentering method.

The joint covariance matrix of the moment and Jacobian vectors, $\Sigma(\theta) = E[\psi_i(X_i, \theta)\psi_i(X_i, \theta)'] - E[\psi_i(X_i, \theta)]E[\psi_i(X_i, \theta)]'$ plays a central role in this paper, and the similar construction of the orthogonalized moments has been used in the weak-identification literature, though from a different perspective than ours. Kleibergen (2005) constructs a recentered Jacobian that is asymptotically uncorrelated with the sample moments, by projecting the Jacobian on the moment vector, and it was used to construct weak-identification robust test-statistics. Kleibergen and Zhan (2025) extend this construction to accommodate misspecification, proposing a double robust Lagrange multiplier test for the pseudo-true value of the continuously updating estimator (CUE), with $\Sigma_{11}(\theta)^{-1}$, which is the efficient choice 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 are $n \times 1, n \times p$, 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 $m \times m$ matrix $W > 0$, the standard GMM estimator is

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

⁹Under 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 $\Sigma_{11,2}$.

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, ME has a closed-form as follows

$$\begin{aligned} \hat{\theta}_{ME}(W) = & \left(\frac{1}{n} X'Z \Sigma_{11,2}^{-1} Z'X \right)^{-1} \left(X'Z \Sigma_{11,2}^{-1} \left(\frac{1}{n} Z'Y - E[g(X_i, \theta_W)] \right) \right. \\ & \left. - X'Z \Sigma_{11,2}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \left(\frac{1}{n} \text{vec}(Z'X) - E[\text{vec}(Z'X)] \right) \right) \quad (17) \end{aligned}$$

where $\Sigma_{11,2} = (\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21})$, $\Sigma(\theta_W) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$, and $\Sigma(\theta)$, i.e., $E\left[\begin{pmatrix} Z_i Y_i - Z_i X_i' \theta \\ \text{vec}(X_i Z_i') \end{pmatrix} \begin{pmatrix} Z_i Y_i - Z_i X_i' \theta \\ \text{vec}(X_i Z_i') \end{pmatrix}'\right] - E\left[\begin{pmatrix} Z_i Y_i - Z_i X_i' \theta \\ \text{vec}(X_i Z_i') \end{pmatrix}\right] E\left[\begin{pmatrix} Z_i Y_i - Z_i X_i' \theta \\ \text{vec}(X_i Z_i') \end{pmatrix}'\right]$. Sample analogs of $\Sigma_n(\hat{\theta}_W)^{-1}$ can also be used here. In the linear IV example, the ME estimator, (17) can be considered as efficient linear combinations of the recentered GMM/TSLS estimator and the Jacobian moment conditions.

Assuming that the data-generating process is $Y_i = \theta_0 X_i + Z_{2i} + \varepsilon_i$, $X_i = Z_{1i} + 2Z_{2i} + v_i$, $Z_i = (Z_{1i}, Z_{2i})' \stackrel{i.i.d.}{\sim} N(0, I_2)$, $(\varepsilon_i, v_i) \sim N(0, [1, 0.5; 0.5, 1])$ with $\theta_0 = 0$. When we consider the weight matrix $W = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ 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 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 $V_{ME} = \frac{969}{6116} \approx 0.1584$ (do 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 72% and 88% for $\rho = -0.5$ and -0.9 , respectively.¹⁰

Remark 1 (*s*-step and the iterated GMM estimators under misspecification). *In this paper, we only consider the one-step GMM estimator in the usual sense with the fixed W . It is well-known that the GMM pseudo-true value changes with each step of the iteration under misspecification, and the asymptotic distribution of the conventional *s*-step GMM depends on that of the previous-step estimator at which the efficient weight matrix is evaluated, making the asymptotic analysis complicated (see, Hall and Inoue (2003), Hwang, Kang and Lee (2022), and Hansen and Lee (2021) for an asymptotic distribution of two-step and the iterated GMM estimator under misspecification). However, using the “standard” efficient weighting-matrix W_0 is ambiguous under misspecification*

¹⁰Although 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 the count data models with endogenous regressor (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.

and it is not clear why we want to stick with this choice in each iteration, which we show in this paper is suboptimal. The choice of $W^* = \Sigma_{11,2}^{-1}(\theta)$ deserves more attention, at least in the linear GMM setup. The FOC condition for the efficient GMM estimator based on $W^* = \Sigma_{11,2}^{-1}(\theta_W)$ with the one-step GMM and the initial weighting matrix W is $G' \Sigma_{11,2}^{-1}(\theta_W) E[g(X_i, \theta_W^*)] = 0$. In general, the FOC condition $G' \Sigma_{11,2}^{-1}(\theta_0) E[g(X_i, \theta_0)] = 0$ holds for the iterated GMM pseudo-true values θ_0 with the $W = \Sigma_{11,2}^{-1}(\theta)$.

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 Hall and Horowitz (1996) recentered bootstrap GMM estimator.

The ME-GMM estimator, $\hat{\theta}_{ME-GMM}(\Delta)$, with the fixed weighting matrix $(mp+m) \times (mp+m)$ matrix $\Delta > 0$ solves

$$\hat{\theta}_{ME-GMM}(\Delta) = \arg \min_{\theta} \left(n^{-1} \sum_{i=1}^n \tilde{\psi}(X_i, \theta) \right)' \Delta \left(n^{-1} \sum_{i=1}^n \tilde{\psi}(X_i, \theta) \right) \quad (18)$$

where,

$$\tilde{\psi}(X_i, \theta) = \begin{pmatrix} g(X_i, \theta) - E[g(X_i, \theta_W)] \\ \text{vec}(G(X_i, \theta)') - E[\text{vec}(G(X_i, \theta_W)')] \end{pmatrix}.$$

We can easily show that the ME-GMM with the efficient weighting matrix $\Delta = E[\tilde{\psi}(X_i, \theta_W) \tilde{\psi}(X_i, \theta_W)']^{-1} = \Sigma(\theta_W)^{-1}$ is asymptotically equivalent to the ME estimator, $\hat{\theta}_{ME}(W)$. The consistency and the asymptotic normality of $\hat{\theta}_{ME-GMM}(W)$ immediately follow by the same Assumptions (Assumption 1.2-1.4).

We first note that Lee and Liao (2018) consider the same GMM problems under correct specification ($E[g(X_i, \theta_W)] = 0$) and singular Jacobian ($E[G(X_i, \theta_W)] = 0$). However, the ME estimator, $\hat{\theta}_{ME}(W)$, and the ME-GMM estimator in (18) are not fully feasible without estimation of the population moments $E[g(X_i, \theta_W)]$ and $E[\text{vec}(G(X_i, \theta_W)')]$, which was used for recentering. If we use the sample counterparts of the re-centered moments, e.g., replacing the sample analogs of $E[g(X_i, \theta_W)]$ and $E[\text{vec}(Z'X)]$, then the estimator degenerates to the original estimator $\hat{\theta}_W$, so it does not correctly account for variations in the limiting distributions. We consider the sample-split estimators in Section 5 as potential alternatives.

Here, we consider the bootstrap GMM estimator which mimics (18) with the efficient weighting matrix $\Delta = \Sigma(\theta_W)^{-1}$. Let $\{X_i^*\}_{i=1}^n$ as the bootstrap samples, independently sampled from the original sample $\{X_i\}_{i=1}^n$ with replacement. With re-centering, the bootstrap ME-GMM estimator solves

$$\hat{\theta}_{ME}^*(W) = \arg \min_{\theta} \left(n^{-1} \sum_{i=1}^n \tilde{\psi}^*(X_i^*, \theta) \right)' \Delta^* \left(n^{-1} \sum_{i=1}^n \tilde{\psi}^*(X_i^*, \theta) \right) \quad (19)$$

where,

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

,

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

$$\Delta^* = \left(E^*[\psi(X_i^*, \hat{\theta}_W)\psi(X_i^*, \hat{\theta}_W)'] - E^*[\psi(X_i^*, \hat{\theta}_W)]E^*[\psi(X_i^*, \hat{\theta}_W)]' \right)^{-1}. \quad (20)$$

The recentering by $g_n(\hat{\theta}_W)$, $\text{vec}(G_n(\hat{\theta}_W))$ ensures $E^*[g(X_i^*, \hat{\theta}_W) - g_n(\hat{\theta}_W)] = 0$, and the bootstrap moments $\tilde{\psi}^*(X_i^*, \theta)$ has mean zero at $\hat{\theta}_W$ conditional on the data. Δ^* is symmetric positive definite weighting matrix that is a bootstrap sample version of the efficient weighting matrix $\Delta = \Sigma(\theta_W)^{-1}$, where $\Sigma(\theta) = E[\psi_i(X_i, \theta)\psi_i(X_i, \theta)'] - E[\psi_i(X_i, \theta)]E[\psi_i(X_i, \theta)]'$.

The proposed bootstrap estimator utilizes the re-centering of the Jacobian estimator to ensure that the bootstrap expected Jacobian matrix evaluated at the estimator is zero, which is essential to correctly mimic the asymptotic distribution under the misspecification. This “double re-centering” idea has been also considered in weak-identification literature to accurately reflect the behavior of the original Jacobian (e.g., Dovonon and Goncalves (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. Again, this is because the bootstrap sample mean of the moment function evaluated at the estimator, $E^*[g(X_i^*, \theta_0)] = n^{-1}Z'\hat{e}$, is not necessarily equal to zero under correct specification. To achieve refinement, Hall and Horowitz (1996) proposed a recentered bootstrap GMM estimator;

$$\hat{\theta}_{HH}^* = \arg \min_{\theta} \left(n^{-1} \sum_{i=1}^n g(X_i^*, \theta) - g_n(\hat{\theta}_W) \right)' W^* \left(n^{-1} \sum_{i=1}^n g(X_i^*, \theta) - g_n(\hat{\theta}_W) \right) \quad (21)$$

where W_0^* is the bootstrap version of efficient weighting matrix under correct specification, $W_0 = \Sigma_{11}(\theta)^{-1}$, and $\hat{\theta}_W = \hat{\theta}_{GMM}(W)$ is the GMM estimator. $\hat{\theta}_{HH}^*$ can be considered as a special case of the proposed bootstrap ME-GMM estimator $\hat{\theta}_{ME}^*$ 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 Hall-Horowitz recentered bootstrap always creates a correctly-specified bootstrap world, regardless of whether the actual model is misspecified or not. 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 original and Jacobian moments.

Example: Linear IV (continued). Based on the FOC from (18) and using the linearity of the moment condition, ME-GMM has a closed-form

$$\hat{\theta}_{ME-GMM}(\Delta) = \left(\frac{1}{n} X'Z\Delta_{11}Z'X \right)^{-1} \left(X'Z\Delta_{11} \left(\frac{1}{n} Z'Y - E[g(X_i, \theta_W)] \right) + X'Z\Delta_{12} \left(\frac{1}{n} \text{vec}(Z'X) - E[\text{vec}(Z'X)] \right) \right) \quad (22)$$

where $\Delta = \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix}$. When $\Delta_{11} = (Z'Z)^{-1}$ and $\Delta_{12} = 0$, ME-GMM estimator is related to a recentered TSLS estimator,

$$\hat{\theta}_{ME-GMM}(\Delta) = \left(\frac{1}{n} X'Z(Z'Z)^{-1}Z'X \right)^{-1} X'Z(Z'Z)^{-1} \left(\frac{1}{n} Z'Y - E[g(X_i, \theta_W)] \right). \quad (23)$$

The choice of $\Delta_{11} = (Z'Z)^{-1}$, $\Delta_{12} = 0$ for ME-GMM is not efficient under misspecification, although its choice was motivated by the efficiency of GMM under correct specification and homoskedasticity.

With the bootstrap samples $\{Y_i^*, X_i^*, Z_i^*\}_{i=1}^n$, independently sampled from the original sample $\{Y_i, X_i, Z_i\}_{i=1}^n$, Hall and Horowitz (1996) version of the recentered bootstrap TSLS estimator is

$$\hat{\theta}_{HH}^* = (X^{*'}Z^*(Z^{*'}Z^*)^{-1}Z^{*'}X^*)^{-1}X^{*'}Z^*(Z^{*'}Z^*)^{-1}(Z^{*'}Y^* - Z'\hat{e})$$

where $\hat{e} = Y - X'\hat{\theta}_{TSLS}$. This is essentially related to a version of recentered bootstrap estimator of (23). Our proposed bootstrap GMM estimator in this setup is

$$\hat{\theta}_{ME}^*(W) = (X^{*'}Z^*\hat{\Sigma}_{11,2}^{-1}Z^{*'}X^*)^{-1} \left(X^{*'}Z^*\hat{\Sigma}_{11,2}^{-1}(Z^{*'}Y^* - Z'\hat{e}) - X^{*'}Z^*\hat{\Sigma}_{11,2}^{-1}\hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}(\text{vec}(Z^{*'}X^*) - \text{vec}(Z'X)) \right) \quad (24)$$

where $\hat{e} = Y - X'\hat{\theta}_W$, $\hat{\Sigma}_{11,2} = (\hat{\Sigma}_{11} - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21})$, $\hat{\Sigma}(\hat{\theta}_W) = \begin{bmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{bmatrix}$, $\hat{\Sigma}(\theta) = \frac{1}{n} \sum_{i=1}^n \psi(X_i, \theta)\psi(X_i, \theta)'$ - $\sum_{i=1}^n \psi(X_i, \theta) \sum_{i=1}^n \psi(X_i, \theta)'$, and $\psi(X_i, \theta) = ((Z_iY_i - Z_iX_i'\theta)', \text{vec}(X_iZ_i'))'$. In the linear IV setup, $\hat{\theta}_{ME}^*$ is not just a linear combination of the $\hat{\theta}_{HH}^*$ and the recentered Jacobian moments, but an efficient linear combination with $\hat{\Sigma}_{11,2}$.

We show below that the asymptotic distribution of the bootstrap estimator $\hat{\theta}_{ME}^*(W)$ is identical to that of the original ME estimator $\hat{\theta}_{ME}(W)$. We consider the following assumptions for the bootstrap validity. Let 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. $\hat{\theta}_{ME}^*(W) \xrightarrow{p^*} \theta_W$.

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. $\sqrt{n}(g_n^*(\hat{\theta}_W) - g_n(\hat{\theta}_W), \text{vec}(G_n^*(\hat{\theta}_W) - G_n(\hat{\theta}_W)))' \xrightarrow{d^*} N(0, \Sigma(\theta_W))$, where $\hat{\theta}_W = \hat{\theta}_{GMM}(W)$, and $\Sigma(\theta_W)$ is the same covariance matrix as in Theorem 1.

These conditions are standard assumptions under i.i.d. sampling, e.g., Giné and Zinn (1990), and Hahn (1996). Assumption 2.1 on the bootstrap consistency is a high-level assumptions, 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 i.i.d. setup.

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

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

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

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

Theorem 2 shows that the asymptotic distribution of the bootstrap estimator $\sqrt{n}(\hat{\theta}_{ME}^*(W) - \hat{\theta}_{GMM}(W))$ is equal to those of $\sqrt{n}(\hat{\theta}_{ME}(W) - \theta_W)$. Theorem 2 allows us to approximate the sampling variation of the oracle-ME estimator $\hat{\theta}_{ME}(W)$, and the theorem justifies the use of the bootstrap to obtain the minimax bounds that will be established in Section 6 (Theorem 3). Although it is necessary to have stronger conditions such as uniformly square integrable condition to guarantee convergence in moments from the convergence in distribution results, we can still use the trimmed bootstrap estimator of variance as the consistent estimator of the asymptotic variance $(\Gamma(\theta_W)' \Sigma(\theta_W)^{-1} \Gamma(\theta_W))^{-1}$.

4.1 Double-Recentered Hall and Horowitz (1996) Bootstrap Estimator

We note, however, that Theorem 2 does not guarantee the validity of the percentile bootstrap CI based on $\hat{\theta}_{ME}^*(W)$ because it is centered at $\hat{\theta}_{GMM}(W)$, which has a different asymptotic distribution. Using the double-recentering idea from the previous section to approximate the sampling variation of the oracle ME estimator, we propose a double-recentered (DR) 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) argues that recentering in the standard nonparametric GMM bootstrap can be detrimental and is not even needed if we use the analytic misspecification-robust variance estimator. In

our paper, however, we clarify this argument that additional considerations of re-centered Jacobian can achieve asymptotic validity under misspecification. This is also the case for the standard GMM bootstrap estimator without recentering, while Hall and Horowitz (1996) do not achieve asymptotic validity because it *only* recenters the original moment functions.¹¹

We also note that 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 as in Imbens (1997) corresponds to the standard GMM bootstrap estimator without recentering.

We first observe that the GMM-estimator satisfies the estimating equations (8) for the augmented moment conditions with the particular choice of $\Lambda = A(W, \theta_W) = [G(\theta_W)'W \quad g(\theta_W)'W \otimes I_p]$. We then define the misspecification-robust recentered bootstrap estimator $\hat{\theta}_{DR}^*$ as the solution to the bootstrap analog of the augmented estimating equation:

$$A_n(W, \hat{\theta}_W) \tilde{\psi}_n^*(\hat{\theta}_{DR}^*) = 0, \quad (25)$$

where

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

and

$$A_n(W, \hat{\theta}_W) = [G_n(\hat{\theta}_W)'W_n, \quad g_n(\hat{\theta}_W)'W \otimes I_p]$$

is the sample analog of the matrix $A(W, \theta_W) = [G(\theta_W)'W, \quad g(\theta_W)'W \otimes I_p]$. While the HH bootstrap only recenters the original moments g_n^* at $g_n(\hat{\theta}_W)$, the DR bootstrap recenters the full augmented vector ψ_n at its sample counterpart ψ_n^* , including Jacobian estimation variability, appropriately weighted by $\Lambda_n = A_n(W, \hat{\theta}_W)$ to mimic the distribution of $\hat{\theta}_W$.

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

1. draw $\{X_i^*\}_{i=1}^n$ with replacement;
2. with $A_n(W, \hat{\theta}_W)$, solve (25) for $\hat{\theta}_{DR}^*$ via Newton–Raphson initialized at $\hat{\theta}_W$; for the linear model, $\hat{\theta}_{DR}^*$ has a closed-form solution (see below linear IV example);
3. Confidence intervals are constructed as standard percentile or percentile- t intervals based on the $t^* = \frac{\hat{\theta}_{DR}^* - \hat{\theta}_W}{\text{s.e.}(\hat{\theta}_{DR}^*)}$ from the B bootstrap draws.

Example: Linear IV (continued). Since the sample and population Jacobian does not depend

¹¹Since, the main focus of the paper is on the efficiency results for θ_W and minimax bounds, we only considered the GMM estimator with fixed W . Thus, we do not need to consider variations from the weight matrix, which was typically considered for the construction of the Hessian and misspecification-robust variance constructions 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 $\psi(\cdot)$.

on θ , $\hat{\theta}_{DR}^*$ admits a closed-form solution by direct algebra from (25):

$$\begin{aligned}\hat{\theta}_{DR}^* = & (X'Z(Z'Z)^{-1}Z^{*'}X^*)^{-1} \left(X'Z(Z'Z)^{-1}(Z^{*'}Y^* - Z'\hat{e}) \right. \\ & \left. + (\hat{e}'Z(Z'Z)^{-1} \otimes I_p) \text{vec}(Z^{*'}X^* - Z'X) \right).\end{aligned}$$

So, DR bootstrap estimator $\hat{\theta}_{DR}^*$ is approximately equal to $\hat{\theta}_{HH}^* + \tilde{\Lambda}_2 \cdot \text{vec}(Z^{*'}X^* - Z'X)$ with some weights $\tilde{\Lambda}_2$. The last term is a correction term for the Jacobian variability. Under correct specification, $g_n(\hat{\theta}_{DR}) = n^{-1}Z'\hat{e} = O_p(n^{-1/2})$, so the last terms are finite-sample corrections.

Corollary 3. *Suppose that Assumption 1.1-1.5, and Assumption 2 hold. Then,*

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

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

We recommend to consider $\hat{\theta}_{DR}^*$ and the percentile confidence interval for the valid inference for θ_W . In addition, separately report the $\hat{V}_{ME}(W)$ and $\sup_{W \in \mathcal{W}} \hat{V}_{ME}(W)$ based on the general double bootstrap ME estimator $\hat{\theta}_{ME}^*$ or using analytical ME estimates with the standard GMM point estimate $\hat{\theta}_{GMM}(W)$ as evidence of the attainable efficiency frontier 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 Split-Sample ME Estimator

The bootstrap provides a consistent estimate of the misspecification-efficient variance, but the oracle ME point estimate that actually achieves that variance is computable only if “degree of misspecification” ($\gamma_1(W) = E[g(X_i, \theta_W)]$) and the “identification strength” ($\gamma_2(W) = E[G(X_i, \theta_W)]$) parameters are known, that was used for recentering in the ME estimator. While these components can be consistently estimated from the full sample, the population moments and Jacobian cannot be replaced by its sample analog, which suffers from a degeneracy in our setup by collapsing ME estimator back to the standard GMM estimator that invalidates the ME variance reduction.¹²

We consider a sample-splitting ME estimator which uses one-half of a sample to estimate $\gamma(W) = (\gamma_1(W)', \text{vec}(\gamma_2(W))')'$. Then estimated parameters $\hat{\gamma}(W)$ are then used to construct the final ME estimates using the other half of the samples. This construction parallels the split-sample IV estimator of Angrist and Krueger (1995), who uses the first-half of the sample to estimate the population-first stage $E[Z'Z]^{-1}E[Z'X]$ and use them in the IV estimates in the other half sample.

Specifically, we consider the following step to construct the sample-split estimator for $\hat{\theta}_{ME}(W)$;

¹²In Appendix B, we consider the valid inference methods for θ_W without assuming $\gamma(W) = (E[g(X_i, \theta_W)]', \text{vec}(E[G(X_i, \theta_W)]'))'$ is known using the worst-case bounds approach similar to Conley et al. (2012).

1. Randomly divide the samples into $K = 2$ independent subsample roughly equal size $n/2$, and let $\{Y_1, X_1\}, \{Y_2, X_2\}$ as data matrices for each subsample.
2. Use all observations with $\{Y_2, X_2\}$ to estimate the $\gamma(W) = (\gamma_1(W)', \text{vec}(\gamma_2(W))')'$ with the GMM estimator $\hat{\theta}_W$ and let these leave-fold-out estimators as $\hat{\gamma}_{-1}(W)$.
3. Solve the single-split ME estimator $\hat{\theta}^{SS}(W)$ using observations $\{X_1, Y_1\}$ with the $\hat{\gamma}_{-1}(W)$, and calculate the standard error using “efficient” misspecification-robust variance formula $\hat{V}_{ME}(W)$.
4. Repeat the above procedure, $s = 1, \dots, S$, and set $\hat{\theta}^{RSS}(W) = \frac{1}{S} \sum_{s=1}^S \hat{\theta}_s^{SS}(W)$, $SD(\hat{\theta}^{RSS}(W)) = \frac{1}{S-1} \sum_{s=1}^S (\hat{\theta}_s^{SS}(W) - \hat{\theta}^{RSS}(W))^2$. We also consider the variance estimators considered in Chernozhukov et al. (2018) based on median, i.e., $\text{median}\{\hat{V}_{ME,s}(W) + ((\hat{\theta}_s^{SS}(W) - \hat{\theta}_{median}^{SS}(W))(\hat{\theta}_s^{SS}(W) - \hat{\theta}_{median}^{SS}(W)))'\}_{s=1}^S$.
5. In the linear model, when $W^* = \Sigma_{11,2}^{-1}$, then estimation of $\gamma_1(W)$ is not needed as the recentering part will be eliminated by the FOC.

Since $n_1 = n/2$ in the split-sample ME estimator, the asymptotic variance is at least twice the asymptotic variance of the oracle-ME. Repeated single sample-splitting procedure can be done to reduce the variance from the single sample. Furthermore, when we construct ME estimator based on $W = \Sigma_{11,2}^{-1}$, we don't need recentering part $\gamma_1(W)$ due to the FOC with $W^* = \Sigma_{11,2}^{-1}$. This helps reduce computational costs and estimation noises of the sample-splitting estimate. We do not consider the general symmetric sample-splitting estimator (i.e., swap the role of each subsample and then average) or cross-fitting estimator because the symmetry of averaging again has similar degeneracy problems.

6 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 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 matrix $W \in \mathcal{W}$ 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 need 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, GMM estimator is semiparametrically 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.

Our uniform minimax bounds will be not only useful for a researcher who chooses W based on target a specific economic parameter θ_W , but also useful for a researcher who is completely agnostic to the choice of $W \in \mathcal{W}$. Many applied researchers may just pick one W (e.g., TSLS); however, an applied researcher who runs GMM already considers choosing W over a class of different weighting matrices $\mathcal{W}$ routinely; one-step $W = I$ or $(Z'Z)^{-1}$, two-step, and iterated until convergence. Under misspecification each iteration changes the pseudo-true value. So when results differ across one-step, two-step, and iterated GMM, that is direct evidence of sensitivity to W .

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 measure or the information criteria (e.g., Rivers and Vuong (2002), Marmer and Otsu (2012)). We can also consider the optimal weights $W^* = \arg \min_W V(W, \theta_W)$, that minimizes the asymptotic variance of misspecified-GMM in (7) or $W^* = \arg \min_W V_M(\Lambda^*, \theta_W)$ 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 of 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 put some restrictions to have particular causal effects interpretations.

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

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

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 $\pi = E[Z'X]$. While the weights always sum to 1, individual weights w_j are not constrained to be positive. A weight w_j can be negative if π_j (the strength of j -th instruments Z_j) and $\sum_{k=1}^m W_{kj} \pi_k$ (the weighted strength of full instrument vector Z) have opposite signs. When $W = E[Z'Z]^{-1}$, θ_W is the TSLS estimand and Angrist and Imbens (1995), Kolesár (2013) characterizes θ_W as a weighted average of LATE under monotonicity conditions (see

also Andrews (2017)). However, the TSLS weights are not guaranteed to be positive.¹³

In general, researchers can restrict the class of weighting matrix $\mathcal{W}^+$ with positive weights w_j as follows;

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

The class $\mathcal{W}^+$ consists of all matrices W that preserve the “sign” of the instrument strength, i.e., such that $\text{sign}((W\pi)_j) = \text{sign}(\pi_j)$. Note that any diagonal positive definite matrix (e.g., identity, or $\text{diag}((E[ZZ'])^{-1})$) is always in $\mathcal{W}^+$. The TSLS weighting matrix $W = (E[ZZ'])^{-1}$ is in this class only under strict conditions if we have covariates (e.g., Blandhol et al. (2025), and Słoczyński (2024)). Restricting to the class $\mathcal{W}^+$ allows us to interpret θ_W as a weighted average of the single-instrument estimands. Under the treatment effect heterogeneity, this condition guarantees that θ_W lies strictly between the minimum and maximum of the single-instrument LATE estimands θ_j . Instead of considering the “right” causal target θ_{W*} , regardless of which convex combination that researchers care about with multiple instruments, we provide here uniform (over $W \in \mathcal{W}^+$) asymptotic minimax bounds of the class of causal LATE estimands.

6.1 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 let $\mathcal{F}^C$ as follows;

$$\begin{aligned} \mathcal{F}^C = \{ & (F, \theta) \in \mathcal{F} \times \Theta : (i) \ \theta \text{ is in an open set } \Theta \subset \mathcal{R}^p \text{ such that } g(X, \beta) \text{ and } \partial g(X, \beta) / \partial \beta' \\ & \text{are continuous for } (X, \beta) \in \mathcal{X} \times \Theta; (ii) \ E_F g(X, \theta) = 0; \\ & (iii) \ E_F [g(X, \theta) g(X, \theta)'] \text{ is positive-definite; } (iv) \ \text{rank } E_F [\partial g(X, \theta) / \partial \theta'] = p\}, \end{aligned} \quad (26)$$

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 $\hat{\theta}_{GMM}(\Sigma_{11}(\theta_0)^{-1})$ 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 said to be 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-

¹³For the same DGP we consider in the linear IV example in Section 3, the pseudo-true value can be decomposed into $\theta_W = w_1 \theta_1 + w_2 \theta_2$, where $w_1 = \frac{1+2\rho}{5+4\rho}$, $w_2 = \frac{2(\rho+2)}{5+4\rho}$, $\theta_1 = 0, \theta_2 = 1/2$. For the TSLS ($\rho = 0$), the weights are $w_1 = \frac{1}{5}$ and $w_2 = \frac{4}{5}$, which are positive. Notably, both weights are positive for $\rho > -1/2$. However, for $\rho < -1/2$, the weight w_1 becomes negative.

true values as a minimizer of 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 ;¹⁴

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

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;

$$\begin{aligned} \mathcal{F}^W = & \{(F, \theta) \in \mathcal{F} \times \Theta : (i) \ \theta \text{ is in an open set } \Theta \subset \mathcal{R}^p \text{ such that } \psi(X, \beta) \text{ and } \partial\psi(X, \beta)/\partial\beta' \\ & \text{are continuous for } (X, \beta) \in \mathcal{X} \times \Theta; (ii) \ E_F \tilde{\psi}(X, \theta) = 0, \text{ where} \\ & \tilde{\psi}(X, \theta) = \psi(X, \theta) - E_{F_0}[\psi(x, \theta_W)], \theta_W = \arg \min_{\theta} E_{F_0} g(X_i, \theta)' W E_{F_0} g(X_i, \theta), \quad (27) \\ & \text{assuming it exists and unique; (iii) } E_F[\tilde{\psi}(X, \theta)\tilde{\psi}(X, \theta)'] \text{ is positive-definite;} \\ & (iv) \ \text{rank} E_F[\partial\psi(X, \theta)/\partial\theta'] = p\}. \end{aligned}$$

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 neighborhood, the recentering is fixed.

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 Condition 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 Condition in $\mathcal{F}^W$ is satisfied for all $(F, \theta) \in \Gamma^W$. Let $\mathcal{G}^\Gamma = \{(W, (F, \theta)) : W \in \mathcal{W}, (F, \theta) \in \Gamma^W\}$. Let $\theta_{1,W}$ be the first component of θ_W and let $T_n(X_1, \dots, X_n)$ be any (measurable) estimator for $\theta_{1,W}$. Then for any loss function*

¹⁴One can consider other pseudo-distance measure such as Kullback-Leibler (KL) divergence. The KL divergence from F to F_0 is $d(F, F_0) = \int \log(dF_0/dF) dF_0$ if F_0 is absolutely continuous with respect to F , and $d(F, F_0) = \infty$, otherwise. Then, the pseudo distance from $\mathcal{F}^C$ to F_0 is defined by $d(\mathcal{F}^C, F_0) = \inf_{F \in \mathcal{F}^C} d(F, F_0)$. Vuong (1989), Kitamura (2000), Shi (2015) use this generic choice for model selection tests in moment equality/inequality models.

$l(\cdot) \in \mathcal{L}$, we have

$$\liminf_{n \rightarrow \infty} \sup_{(W, (F, \theta)) \in \mathcal{G}^\Gamma} E_F \{l[\sqrt{n}(T_n - \theta_{1,W})]\} \geq \int_{-\infty}^{\infty} l(\sigma u) d\Phi(u),$$

where $\Phi(u)$ is a 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 maximises risk over the class $\mathcal{W}$. This bound can be achieved by the worst-case variance of the ME estimator $\hat{\theta}_{ME}(W)$, $W \in \mathcal{W}$, 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 cost.

One can straightforwardly show that the misspecified-GMM achieves different semiparametric efficiency bounds based on $V_{GMM}(W, \theta_W)$ by replacing moment condition (ii) in $\mathcal{F}^C$ with the (local) first order condition $E_F[\partial g(X, \theta)/\partial \theta]' W E_F[g(X, \theta)] = 0$ (Mukhin (2019)) or using the augmented parameter space approach as in Imbens (1997) with similar regularity (smoothness) conditions. However, as shown in Corollary 1, the standard misspecified-GMM Estimator, which uses the linear combination $\Lambda = [G(\theta_W)' W \quad g(\theta_W)' W \otimes I_p]$ dictated by the FOC of the GMM objective, is inefficient relative to the bound we consider here, because it ignores the potential information contained in the variation of the Jacobian. By using the optimal linear combinations of the augmented moments $\Lambda^* = \Gamma(\theta_W)' \Sigma(\theta_W)^{-1}$, the moment vector $\tilde{\psi}(X, \theta)$ provides an over-identified system for θ , exploiting more information from the degree of misspecification and the identification strength.

7 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

$$Y_i = X_i \theta + Z_i' \gamma + \varepsilon_i, \quad X_i = Z_i' \Pi + v_i, \quad (28)$$

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 $\gamma = (0, \delta)'$ 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 $\Pi = (\pi, 2\pi)'$ 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 $W = I, \hat{\Sigma}_{11,2}^{-1}$, and we report the moderately strong instrument case here, and the weak instrument design results, together with an additional results under the conventional optimal weighting $W = \hat{\Sigma}_{11}^{-1}$ are reported in Appendix C. 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 and $S = 100$ sample-split replications. SD is normalized, relative to the SD of standard GMM with the correctly-specified benchmark ($n = 200, \delta = 0$). We do not report the bias here as it is nearly zero for all estimators targeting the same θ_W .

Tables 1 and 2 report results for the one-step ($W = I$), and the two-step GMM ($W = \hat{\Sigma}_{11,2}^{-1}$), respectively. 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 $V_{ME}(W)$; (iv) Sample-splitting ME estimator, the median repeated sample-split estimates with the median standard deviation. We also considered the following bootstrap estimators; (v) HH (Hall and Horowitz, 1996) percentile bootstrap CI; and (vi) the DR (double-recentered) percentile bootstrap CI proposed in Section 4.

As the degree of misspecification (δ) increases, we observe several consistent patterns across DGPs; 1) the SD of GMM increases, and 2) the coverage of GMM with conventional SE undercovers, while the misspecification-robust SE restores coverage by correctly capturing the SD of GMM; 3) the oracle ME dominates in SD and CI length across all designs, while retaining coverage (coverage 94-95% across all $\{n, \delta\}$ in Tables 1 and 2), consistent with Corollary 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 $0.43 \rightarrow 1.38$, SD of ME $0.40 \rightarrow 0.79$, when $\delta = 0 \rightarrow 2$, $n = 1000$ in Table 1); the oracle ME estimator take into account increases in variance through optimal Jacobian recentering rather than letting the asymptotic variance inflate. The efficiency gain of the oracle ME over GMM are 25.8%, 54.3%, 67.4% for $W = I$, and 23.8%, 50.1%, 63.8% for $W = \hat{\Sigma}_{11,2}^{-1}$, when $\delta = 0.5, 1, 2$, respectively. We also observe that the efficient weighting choice $W = \hat{\Sigma}_{11,2}^{-1}$ reduces the SD of GMM.

Furthermore, among feasible procedures, coverage of the sample-splitting ME and DR is very close to the nominal level in most cases, while the HH bootstrap exhibits the coverage distortions. While the sample-splitting ME provides similar coverage/lengths, the DR CI lengths shows meaningfully smaller CI compare to the CI with analytic GMM misspecification-robust SE (10-17% smaller in Table 2 for $\delta \neq 0$). The misspecification-robust SE adds a Jacobian uncertainty term that inflates SE with the moment violations, and this SE can be large under mild or severe misspecification, which leads to over-coverage and larger CI lengths in Table 2.

The simulation results confirms the efficiency of the oracle ME estimator, and overall suggest that the DR bootstrap and sample-split procedures provides valid procedures, and reasonable alternatives to the standard GMM with analytic misspecification-robust standard errors.

Table 1: Monte Carlo results under $W = I$

Estimator	$\delta = 0$			$\delta = 0.5$			$\delta = 1$			$\delta = 2$		
	SD	Cov	Len	SD	Cov	Len	SD	Cov	Len	SD	Cov	Len
$n = 200$												
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	0.908
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
Oracle ME	0.93	0.944	0.367	0.87	0.949	0.349	1.08	0.945	0.422	1.78	0.942	0.695
RSS ME median	1.04	0.974	0.466	1.04	0.976	0.471	1.67	0.963	0.710	3.18	0.970	1.378
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	0.412	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	0.305	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
Oracle ME	0.58	0.945	0.232	0.56	0.945	0.221	0.69	0.944	0.267	1.09	0.952	0.440
RSS ME median	0.63	0.961	0.265	0.66	0.955	0.273	1.04	0.948	0.413	1.95	0.956	0.804
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	0.417
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
Oracle ME	0.40	0.958	0.164	0.39	0.948	0.156	0.47	0.950	0.189	0.79	0.951	0.311
RSS ME median	0.44	0.958	0.181	0.46	0.955	0.187	0.72	0.945	0.284	1.38	0.945	0.556
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

Notes. We consider (i) standard GMM estimator with the conventional SE; (ii) GMM with the misspecification-robust SE; (iii) the oracle ME estimator with efficient SE, $V_{ME}(W)$; (iv) Repeated sample-splitting ME estimator (RSS ME), the median repeated sample-split estimates with ± 1.96 SD; (v) Hall and Horowitz (1996) (HH) percentile bootstrap; and (vi) the Double-Recentered percentile bootstrap (DR). For GMM, Oracle ME, and RSS 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 $\delta = 0, W = I$), so values < 1 (> 1) indicate higher (lower) precision than this benchmark. “Cov” is empirical 95% coverage; “Len” is average CI length. Based on 2,000 Monte Carlo replications, $B = 1,000$ bootstrap draws, $S = 100$ sample-split replications.

8 Empirical Examples

We illustrate our proposed methods with two main empirical applications: returns to schooling example of Card (1995) and Angrist and Krueger (1991). We also consider the dynamic panel data model of income and democracy in Acemoglu et al. (2008) in Appendix D.

8.1 Card (1995)

We replicate the well-known college-proximity IV analysis by Card (1995) who uses the National Longitudinal Survey of Young Men for 1976. The sample size is $n = 3010$. The outcome is log wage, 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 3 compares several estimators across four specifications, and three weighting choices:

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

Estimator	$\delta = 0$			$\delta = 0.5$			$\delta = 1$			$\delta = 2$		
	SD	Cov	Len	SD	Cov	Len	SD	Cov	Len	SD	Cov	Len
$n = 200$												
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	0.947	0.395	0.94	0.976	0.447	1.48	0.985	0.700	2.95	0.976	1.340
Oracle ME	0.93	0.944	0.367	0.87	0.950	0.349	1.09	0.942	0.422	1.83	0.936	0.695
RSS ME median	1.04	0.971	0.458	0.97	0.982	0.480	1.52	0.985	0.737	3.01	0.987	1.441
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	0.956	0.249	0.61	0.979	0.288	0.94	0.984	0.459	1.84	0.980	0.873
Oracle ME	0.58	0.945	0.232	0.56	0.945	0.221	0.68	0.943	0.267	1.13	0.946	0.440
RSS ME median	0.63	0.958	0.264	0.61	0.975	0.280	0.95	0.977	0.439	1.88	0.978	0.858
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	0.250	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	0.954	0.176	0.42	0.983	0.207	0.66	0.987	0.328	1.31	0.980	0.624
Oracle ME	0.40	0.958	0.164	0.39	0.947	0.156	0.48	0.951	0.189	0.81	0.943	0.311
RSS ME median	0.44	0.957	0.180	0.43	0.971	0.195	0.67	0.973	0.305	1.32	0.970	0.600
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

$W = I$, $W = (Z'Z)^{-1}$ (2SLS), and $W = \hat{\Sigma}_{11,2}^{-1}$ (efficient).¹⁵ 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) treat experience and its square as endogenous, adding age and age squared to the instrument sets. These specifications correspond to Table 3 in Card (1995). 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 2SLS estimand to admit a weakly causal estimand.

Table 3 shows that GMM point estimates are very similar across specifications and W (0.164-0.175 in four columns with three W). The conventional and misspecification-robust SEs are very similar in this example (within 1.5%). The Hansen J test does not reject in any specifications (p-values ≥ 0.35), however, misspecification can still exist even when J test reject. The J -test is not informative to decide whether ME bound is informative, and the moment-Jacobian covariance Σ_{12} is more relevant in this direction. Although, the ME efficiency bound is essentially similar across specifications (0.0346-0.0366), it is 15-20% smaller than the conventional/misspecification-robust SE. The ME-GMM bootstrap SD tracks the bound closely, and the DR percentile CI is comparable

¹⁵In both of applications we consider here, we also considered conventional optimal weight $W = \hat{\Sigma}_{11}^{-1}$ for each specification. Results are almost identical and thus we do not report in the paper.

to the Wald CI with misspecification-robust SEs across specifications.

Similar to Mogstad and Torgovitsky (2024), who consider similar specifications in Card applications, we find that a Ramsey (1969) RESET test rejects the null of rich covariates in specifications (1) and (2). However, when treating experience as endogenous, we do not reject the RESET test ($p\text{-value} = 0.88$) in Columns in (3) and (4) at 1% significance level, which indicates that the first-stage nonlinearity is captured by the dummy interactions terms. Since the point estimates are similar across all specification, the same caveat emphasized by Mogstad and Torgovitsky (2024) still applies: two estimates can be very similar even when one corresponds to a non-negatively weighted causal estimand and the other does not.

8.2 Angrist and Krueger (1991)

In this section, we consider Angrist and Krueger (1991) quarter-of-birth instruments example. We use the US Census 1980 extract of men born 1930-1939 ($n = 329,509$). The outcome is the log wage, and 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 4, Column (1) uses the baseline QOB dummies ($m = 3$); (2) uses QOB $\times$ YOB interactions ($m = 30$), (3) adds QOB $\times$ State interactions ($m = 180$) which may exhibits many and weak instruments 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 4 parallel Table 3 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 significantly smaller than the conventional/misspecification-robust SE when $W = I$. When we use 2SLS ($(Z'Z)^{-1}$) and efficient weights ($\hat{\Sigma}_{11,2}^{-1}$), the ME efficiency bounds are almost similar to the conventional SE, but still 10-20% smaller than the misspecification-robust SE as predicted by our theory.

Adding interaction terms barely changes the GMM point estimates. However, GMM point estimate under $W = I$ is systematically higher (0.091–0.113) than under 2SLS or efficient weighting (0.081–0.099). Furthermore, DR bootstrap CI is close to the Wald CIs under 2SLS and efficient weighting, but it is noticeably wider when $W = I$. In all specifications, we reject a Ramsey (1969) RESET test. ME-GMM inference still applies to the pseudo-true value θ_W , but the rich covariates condition of Blandhol et al. (2025) fails, so causal interpretations of all of 2SLS/GMM estimates in Table 4 may be fragile in this application.

9 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

Table 3: Returns to schooling: Card (1995)

	exp/exp^2 as exogenous		exp/exp^2 as endogenous	
	(1)	(2)	(3)	(4)
<i>Panel A: $W = I$ (identity)</i>				
GMM point	0.1763	0.1754	0.1674	0.1689
Conv. SE	(0.0448)	(0.0415)	(0.0430)	(0.0425)
Misspec-robust SE	(0.0445)	(0.0413)	(0.0427)	(0.0423)
ME efficiency bound	(0.0366)	(0.0348)	(0.0347)	(0.0346)
RSS ME median	0.1831	0.1801	0.1743	0.1829
	(0.0648)	(0.0622)	(0.0637)	(0.0636)
ME-GMM Boot SD	(0.0375)	(0.0356)	(0.0372)	(0.0363)
DR Boot 95% CI	[0.096,0.289]	[0.103,0.283]	[0.094,0.294]	[0.097,0.290]
<i>Panel B: $W = (Z'Z/n)^{-1}$ (2SLS)</i>				
GMM point	0.1611	0.1642	0.1597	0.1644
Conv. SE	(0.0405)	(0.0384)	(0.0408)	(0.0411)
Misspec-robust SE	(0.0413)	(0.0390)	(0.0416)	(0.0415)
ME efficiency bound	(0.0366)	(0.0348)	(0.0347)	(0.0346)
RSS ME median	0.1748	0.1744	0.1654	0.1725
	(0.0674)	(0.0621)	(0.0629)	(0.0638)
ME-GMM Boot SD	(0.0382)	(0.0360)	(0.0379)	(0.0362)
DR Boot 95% CI	[0.090,0.259]	[0.097,0.257]	[0.090,0.279]	[0.091,0.286]
<i>Panel C: $W = \hat{\Sigma}_{11,2}^{-1}$ (efficient)</i>				
GMM point	0.1606	0.1640	0.1594	0.1643
Conv. SE	(0.0404)	(0.0384)	(0.0408)	(0.0411)
Misspec-robust SE	(0.0413)	(0.0390)	(0.0417)	(0.0416)
ME efficiency bound	(0.0366)	(0.0348)	(0.0347)	(0.0346)
RSS ME median	0.1653	0.1757	0.1719	0.1682
	(0.0645)	(0.0583)	(0.0635)	(0.0608)
ME-GMM Boot SD	(0.0391)	(0.0358)	(0.0373)	(0.0371)
DR Boot 95% CI	[0.089,0.268]	[0.093,0.262]	[0.089,0.274]	[0.096,0.280]
<i>Panel D: Diagnostics</i>				
First-stage F	13.5	15.6	8.65	8.88
J -test p -value	0.351	0.419	0.461	0.609
RESET F (p -value)	25.98 (0.000)	7.48 (0.001)	3.96 (0.019)	0.13 (0.883)
# excluded IVs	2	2	4	4
# exogenous controls	6	17	4	7

Notes. We report returns to education coefficients with standard errors or bootstrap SD (in parentheses). Column (1) consider baseline controls include indicators for Black, residence in an SMSA in 1976, residence in the South in 1976, and experience and experience squared. Column (2) additionally controls cubic and quartic 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: $B = 2,000$. RSS ME median estimates are the median across $S = 100$ replications, with standard errors for the median similarly computed as in Chernozhukov et al. (2018). DR bootstrap percentile 95% CI in Section 4.1 is reported in brackets.

Table 4: Returns to schooling: Angrist and Krueger (1991)

	(1)		(2)		(3)	
	basic	flexible	basic	flexible	basic	flexible
<i>Panel A: $W = I$ (identity)</i>						
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 efficiency bound	(0.0204)	(0.0207)	(0.0164)	(0.0166)	(0.0095)	(0.0096)
RSS ME median	0.1102	0.1099	0.0828	0.0804	0.0972	0.0890
	(0.0342)	(0.0366)	(0.0282)	(0.0337)	(0.0243)	(0.0264)
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]
<i>Panel B: $W = (Z'Z/n)^{-1}$ (2SLS)</i>						
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)	(0.0115)
ME efficiency bound	(0.0204)	(0.0207)	(0.0164)	(0.0166)	(0.0095)	(0.0096)
RSS ME median	0.0998	0.1012	0.0734	0.0787	0.0774	0.0740
	(0.0332)	(0.0343)	(0.0303)	(0.0293)	(0.0255)	(0.0202)
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]
<i>Panel C: $W = \hat{\Sigma}_{11,2}^{-1}$ (efficient)</i>						
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)	(0.0178)	(0.0179)	(0.0116)	(0.0117)
ME efficiency bound	(0.0204)	(0.0207)	(0.0164)	(0.0166)	(0.0095)	(0.0096)
RSS ME median	0.0998	0.0979	0.0781	0.0707	0.0795	0.0754
	(0.0322)	(0.0348)	(0.0288)	(0.0300)	(0.0267)	(0.0215)
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]
<i>Panel D: Diagnostics</i>						
First-stage F (ed)	30.5		4.75		2.43	
Hansen J p -value	0.310	0.348	0.821	0.838	0.774	0.808
RESET F (p -value)	251.9 (<.001)	72.0 (<.001)	248.3 (<.001)	70.5 (<.001)	68.4 (<.001)	66.1 (<.001)

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 $\times$ YOB interactions ($m = 30$), (3) adds QOB $\times$ 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: $B = 2,000$. RSS ME median estimates are the median across $S = 100$ replications, with standard errors for the median similarly computed as in Chernozhukov et al. (2018). DR bootstrap percentile 95% CI in Section 4.1 is reported in brackets.

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.

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, the relationship between the ME estimator 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 generally difficult to compare in a unified framework. We note however, that the continuously updating estimator (CUE) implicitly updates the weighting matrix as a function of θ , which may share connections with the augmented-moment approach. Third, extending the framework to accommodate weak identification, where the Jacobian $G(\theta_W)$ is near-singular, would be important given the connection to the literature (e.g., Lee and Liao (2018) and Kleibergen (2005)) that already exploits Jacobian information. Finally, developing practical guidance on the choice of the weighting matrix class $\mathcal{W}$ would provide applied researchers with concrete recommendations for implementation.

References

- Abadir, K.M., and Magnus, J.R. (2005). *Matrix Algebra*. Cambridge University Press.
- Acemoglu, D., Johnson, S., Robinson, J. A., and Yared, P. (2008). Income and Democracy. *American Economic Review*, 98(3), 808-842.
- 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. (2017). On the Structure of IV Estimands. *Journal of Econometrics*, 211 (1), 294-307.
- Andrews, I., Barnhard, H., and Carlson, J. (2024). True and pseudo-true parameters. *Working Paper*.
- 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., Chen, J., and Techhio, 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. (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.
- 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(1), 63-69.
- Cervellati, M., Jung, F., Sunde, U., and Vischer, T. (2014). Income and democracy: Comment. *American Economic Review*, 104(2), 707-719.
- Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3), 305-334.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21, 1-68.
- 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 Goncalves, S. (2017). Bootstrapping the GMM overidentification test under first-order underidentification. *Journal of Econometrics*, 201, 43-71.
- Dovonon, P., 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(1), 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.
- Kitamura, Y. (2000). Comparing misspecified dynamic econometric models using nonparametric likelihood, *Working Paper*, University of Pennsylvania.
- 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, S. (2014). Asymptotic refinements of a misspecification-robust bootstrap for generalized method of moments estimators *Journal of Econometrics*, 178(3), 398-413.
- Lee, J.H, and Liao, Z. (2018). On standard Inference for GMM with Local Identification Failure of Known Forms. *Econometric Theory*, 34, 790-814.
- 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.
- Shi, X. (2015). Model selection tests for moment inequality models, *Journal of Econometrics*, 187(1), 1-17.

- 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.
- Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica*, 57(2), 307-333.
- 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.

Appendix A: 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, $F(X_i, \theta) = \frac{\partial}{\partial \theta'} \text{vec}(G(X_i, \theta)')$, $F_n(\theta) = n^{-1} \sum_{i=1}^n F(X_i, \theta)$ and $F(\theta) = E[F_n(\theta)]$. Using the first-order Taylor expansion from the FOC of the GMM estimator,

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

we have

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

where

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

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

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

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, $\text{vec}(AB) = (B' \otimes I) \text{vec}(A)$.

Under Assumption 1, which is the standard regularity conditions for the GMM estimators, we have $\hat{\theta}_{gmm}(W) \xrightarrow{p} \theta_W$, $G_n(\theta_W) \xrightarrow{p} G(\theta_W)$, $H_n(\tilde{\theta}) \xrightarrow{p} H(\theta_W)$, where $H(\theta) = G(\theta)' W G(\theta) + (g(\theta)' W \otimes$

$I_p)F(\theta) = A(W, \theta)\Gamma(\theta)$. Thus,

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

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

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

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

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

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

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. $\square$

Proof of Corollary 1

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

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

We now show that $V_{AM}(\Lambda, \theta_W) - (\Gamma(\theta_W)' \Sigma(\theta_W)^{-1} \Gamma(\theta_W))^{-1}$ is positive semi-definite. 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 G is positive semidefinite, then HGH' 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

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

This implies,

$$V_{AM}(\Lambda, \theta_W) = (\Lambda\Gamma(\theta_W))^{-1}\Lambda\Sigma(\theta_W)\Lambda'((\Lambda\Gamma(\theta_W))')^{-1} - (\Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Gamma(\theta_W))^{-1} \geq 0.$$

For the last part of Corollary, 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

$$\begin{aligned} \Gamma(\theta_W)'\Sigma(\theta_W)^{-1}\Gamma(\theta_W) &= G(\theta_W)'\Sigma_{11,2}^{-1}G(\theta_W) - F(\theta_W)'\Sigma_{22,1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}G(\theta_W) \\ &\quad - G(\theta_W)'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22,1}^{-1}F(\theta_W)' + F(\theta_W)'\Sigma_{22,1}^{-1}F(\theta_W) \end{aligned} \quad (29)$$

By the following Woodbury matrix identity for nonsingular matrix 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 $\Sigma_{11,2}$. Thus, (29) 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. $\square$

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 (29) 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

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

where $C = (\delta' \otimes I_m) K_{p,m}$ and $K_{p,m}$ is the $mp \times mp$ commutation matrix satisfies $\text{vec}(G(X_i)) = K_{p,m} \text{vec}(G(X_i)')$ (Abadir and Magnus (2005)). $L(\theta) = \begin{pmatrix} I_m & C \\ 0 & I_{mp} \end{pmatrix}$ is block lower-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

$$\begin{aligned} \Sigma_{11}(\theta) &= \Sigma_{11}(\theta_0) + C\Sigma_{21}(\theta_0) + \Sigma_{12}(\theta_0)C' + C\Sigma_{22}C', \\ \Sigma_{12}(\theta) &= \Sigma_{12}(\theta_0) + C\Sigma_{22}, \\ \Sigma_{22}(\theta) &= \Sigma_{22}, \\ \Sigma_{12}(\theta)\Sigma_{22}^{-1}\Sigma_{21}(\theta) &= \Sigma_{12}(\theta_0)\Sigma_{22}^{-1}\Sigma_{21}(\theta_0) + \Sigma_{12}(\theta_0)C' + C\Sigma_{21}(\theta_0) + C\Sigma_{22}C'. \end{aligned}$$

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, $\Sigma_{11,2}(\theta_W)$ is therefore independent of W . This completes the proof. $\square$

Proof of Theorem 2

It is straightforward to show that ME-GMM estimator is same as the augmented M-estimator $\tilde{\theta}(\Lambda)$ with $\Lambda = \Gamma(\theta_W)'\Delta$. Then, the first results in Theorem 2 follows by Theorem 1 with Assumption 1.(5) holds by the assumptions given in the proof.

The bootstrap ME-GMM estimator $\hat{\theta}_{ME}^*$ in (19) satisfies the first-order condition;

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

where

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

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

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

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

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

$$\sqrt{n}(\hat{\theta}_{ME}^* - \hat{\theta}_W) = (\Gamma_n^*(\hat{\theta}_{ME}^*)' \Delta^* \Gamma_n^*(\tilde{\theta}^*))^{-1} \Gamma_n^*(\hat{\theta}_{ME}^*)' \Delta^* \sqrt{n} \tilde{\psi}_n^*(\hat{\theta}_W) \quad (30)$$

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

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

Next, we consider $\sqrt{n} \tilde{\psi}_n^*(\hat{\theta}_W)$ in (30),

$$\sqrt{n} \tilde{\psi}_n^*(\hat{\theta}_W) = \sqrt{n} \begin{pmatrix} g_n^*(\hat{\theta}_W) - g_n(\hat{\theta}_W) \\ \text{vec}(G_n^*(\hat{\theta}_W)') - \text{vec}(G_n(\hat{\theta}_W)')' \end{pmatrix}. \quad (31)$$

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, by the triangular array CLT under Assumption 2 gives $\sqrt{n} \tilde{\psi}_n^*(\hat{\theta}_W) \xrightarrow{d^*} N(0, \Sigma(\theta_W))$. The conclusion then follows by the Slutsky theorem,

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

Proof of Corollary 3 The DR estimator $\hat{\theta}_{DR}^*$ in (25) satisfies the following;

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

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

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

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

Proof of Theorem 3

By construction, for any finite n , the following holds

$$\begin{aligned} \sup_{(W, (F, \theta)) \in \mathcal{G}^\Gamma} E_F \{l[\sqrt{n}(T_n - \theta_{1,W})]\} &= \sup_{W \in \mathcal{W}} \sup_{(F, \theta) \in \mathcal{F}^W} E_F \{l[\sqrt{n}(T_n - \theta_{1,W})]\} \\ &\geq \sup_{(F, \theta) \in \mathcal{F}^{\widetilde{W}}} E_F \{l[\sqrt{n}(T_n - \theta_{1,\widetilde{W}})]\} \end{aligned} \quad (32)$$

for any arbitrary element $\widetilde{W} \in \mathcal{W}$.

Since the inequality (32) holds for all n , we have that

$$\liminf_{n \rightarrow \infty} \sup_{(W, (F, \theta)) \in \mathcal{G}^\Gamma} E_F \{l[\sqrt{n}(T_n - \theta_{1,W})]\} \geq \liminf_{n \rightarrow \infty} \sup_{(F, \theta) \in \mathcal{F}^{\widetilde{W}}} E_F \{l[\sqrt{n}(T_n - \theta_{1,\widetilde{W}})]\} \quad (33)$$

Next, for the specific $\widetilde{W}$, we can similarly follow the approach in Chamberlain (1987), i.e., we can construct the multinomial distribution which is arbitrarily close to (F_0, θ_W) that satisfies the conditions in $\mathcal{F}^W$ and apply the local asymptotic minimax bounds for the multinomial distribution (Lemma 2 in Chamberlain (1987)). To be more specific, we can construct the multinomial distributions F with a finite support and probabilities $p = (p_1, \dots, p_r)$ on the support points, which satisfies the moment condition $\sum_{j=1, \dots, r} p_j \phi(X_j, \theta) = E_{F_0}[\psi(X_i, \theta_W)]$ and then compute the local asymptotic minimax bound on that parametric submodel.

By Theorem 2 of Chamberlain (1987), with the moment function $\tilde{\psi}(\cdot)$ and the pairs (F, θ) that satisfy the regularity conditions the space of probability distributions $\mathcal{F}^{\widetilde{W}}$, the bounds in (33) is bounded below by the following;

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

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

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

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

$$\liminf_{n \rightarrow \infty} \sup_{(W, (F, \theta)) \in \mathcal{G}^\Gamma} E_F \{l[\sqrt{n}(T_n - \theta_{1,W})]\} \geq \sup_W \int_{-\infty}^{\infty} l(\sigma_W u) d\Phi(u) = \int_{-\infty}^{\infty} l(\sup_W \sigma_W u) d\Phi(u)$$

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. □

Appendix B: Worst-Case Variance Bounds over Misspecification and Identification Strength

In this section, we consider the inference problem about θ_W without assuming $\gamma(W) = (\gamma_1(W)', \text{vec}(\gamma_2(W)'))'$ is known. We assume only that the support of γ is known and consider confidence interval for θ_W by taking a union of valid confidence intervals specific to each γ . This approach is closely related to Conley, Hansen, and Rossi (2012), where they construct the union of confidence intervals over the nuisance parameter (“violation of the exclusion restrictions”) which is related to γ_1 we consider here. In our framework, the nuisance parameter γ not only include the degree of misspecification (γ_1), but also contains the identification strength (γ_2). Interestingly, we can consider the trade-off relationship in the construction of the parameter spaces from the F.O.C of the overidentified GMM conditions; $\gamma_2'W\gamma_1 = 0$, if γ is equal to true value $\gamma(W)$.

Define the parameter space

$$\bar{\Gamma}_W = \{\gamma = (\gamma_1', \text{vec}(\gamma_2)')' \in \mathbb{R}^{m+mp} : 0 \leq \|\gamma_1\| \leq \bar{\gamma}_1, \underline{\gamma}_2 \leq \|\gamma_2\| \leq \bar{\gamma}_2, \underline{\gamma}_2 > 0, \gamma_2'W\gamma_1 = 0\}. \quad (34)$$

$\bar{\gamma}_1$ denotes the worst-case degree of misspecification and $\|\gamma_1\| = 0$ corresponds to the correct-specification. $\underline{\gamma}_2 > 0$ rules out weak identification. We do not consider data-dependent choice of $\bar{\gamma}_1, \underline{\gamma}_2$ in this paper, but practically we can consider $\bar{\gamma}_1$ as J-statistics (which estimate $\|\gamma_1\|^2$ with W), and some pre-test identification strength measure for $\underline{\gamma}_2$ such as effective F-statistics in the linear IV setup. The FOC condition requires γ_1 to lie in the null space of $\gamma_2'W$, and thus the direction of γ_1 is restricted. As we discussed earlier, when $W = \Sigma_{11,2}^{-1}$ in the linear model, the recentering part $\gamma_1(W)$ will be eliminated by the FOC $G'\Sigma_{11,2}^{-1}\gamma_1(W) = 0$ from Corollary 2. Since $\gamma_1(W)$ is not needed for $\hat{\theta}_{ME}(W)$, the parameter space $\bar{\Gamma}_W$ only requires restrictions on γ_2 . This reduces not only computational costs, but also the conservativeness of the CI considered here.

For any $\gamma \in \bar{\Gamma}_W$, we can define the ME-estimator defined in (18) as follows;

$$\hat{\theta}_{ME}(W; \gamma) = \arg \min_{\theta} \left(\psi_n(\theta) - \gamma \right)' \hat{\Sigma}(\hat{\theta}_W)^{-1} \left(\psi_n(\theta) - \gamma \right) \quad (35)$$

where $\psi_n(\theta) = \frac{1}{n} \sum_{i=1}^n \psi(X_i, \theta)$, $\psi(X_i, \theta) = (g(X_i, \theta)', \text{vec}(G(X_i, \theta))')'$, and $\hat{\Sigma}(\hat{\theta}_W)$ is a consistent estimator of $\Sigma(\theta_W)$ with the preliminary GMM estimator $\hat{\theta}_W$. Although, we can consider the pseudo-true value $\theta_W(\gamma)$ as a minimizer of the population objective function of (35), but for simplicity, we assume that for all $\gamma \in \bar{\Gamma}$, there exists unique value of $\theta = \theta_W(\gamma)$ such that $E[\psi(X_i, \theta_W(\gamma))] = \gamma$. Then, it will follow under the conventional regularity conditions that

$$\sqrt{n}(\hat{\theta}_{ME}(W; \gamma) - \theta_W(\gamma)) \xrightarrow{d} N(0, V_{ME}(W; \gamma)) \quad (36)$$

where $V_{ME}(W; \gamma) = (\Gamma(\theta_W(\gamma))' \Sigma(\theta_W)^{-1} \Gamma(\theta_W(\gamma)))^{-1} \Gamma(\theta_W(\gamma))' \Sigma(\theta_W)^{-1} \Sigma_\gamma \Sigma(\theta_W)^{-1} \Gamma(\theta_W(\gamma)) (\Gamma(\theta_W(\gamma))' \Sigma(\theta_W)^{-1} \Gamma(\theta_W(\gamma)))^{-1}$, and $\Sigma_\gamma = E[\psi_i(X_i, \theta_W(\gamma)) \psi_i(X_i, \theta_W(\gamma))'] - \gamma \gamma'$. Note that at the true values $\gamma = \gamma(W)$, we have $\theta_W(\gamma(W)) = \theta_W$ since $E[\psi(X_i, \theta_W(\gamma))] = \gamma(W)$, and thus $\Sigma_\gamma = \Sigma(\theta_W)$, $V_{ME}(W, \gamma) = V_{ME}(W)$ when evaluated at the $\gamma(W)$. Theorem 3 can be extended to the worst-case minimax bounds $\sup_{\gamma \in \bar{\Gamma}} V_{ME}(W; \gamma)$ uniformly over $\gamma \in \bar{\Gamma}$. In the linear model, $V_{ME}(W) = (\gamma_2' \Sigma_{11,2}^{-1} \gamma_2)^{-1}$ and it depends only on γ_2 , and the bounds for $V_{ME}(W)$ can be easily calculated with the $\|\gamma_2\|^2$ and the Rayleigh quotient for $\Sigma_{11,2}^{-1}$, which reduces the dimension significantly.

For simplicity, we suppose that θ is a scalar ($p = 1$), in the following construction. We can construct the union of confidence interval as follows

$$CI_{1-\alpha} = \bigcup_{\gamma \in \bar{\Gamma}_W} CI_{1-\alpha}(\gamma)$$

where $CI_{1-\alpha}(\gamma) = \hat{\theta}_{ME}(W; \gamma) \pm z_{1-\alpha/2} \sqrt{\hat{V}_{ME}(W; \gamma)/n}$ and $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution. Reporting $CI_{1-\alpha}(\gamma)$ as a function of (γ_1, γ_2) allow researchers to do transparent sensitivity analysis for assumptions about misspecification and identification strength. Since for each $\gamma \in \bar{\Gamma}_W$, $CI_{1-\alpha}(\gamma)$ covers $\theta_W(\gamma)$ with probability $1 - \alpha$ asymptotically, and $\gamma(W) \in \bar{\Gamma}$, we have a valid coverage for θ_W ,

$$\lim_{n \rightarrow \infty} P(\theta_W \in CI_{1-\alpha}) \geq 1 - \alpha.$$

Conley et al. (2012) considers similar type of CI based on TSLS estimates $\hat{\theta}(\gamma)$ and the conventional TSLS standard errors, with γ denotes the direct effect of instruments. The key differences with our approach is that we use the misspecification-efficient estimator $\hat{\theta}_{ME}$ with the misspecification-efficient variance $V_{ME}(W)$ which was shown to be more efficient than the standard misspecification-robust GMM variances. We use the augmented moment system and exploiting additional information from the Jacobian with the potential efficiency gains from the moment and Jacobian correlations.

Appendix C: Additional Simulations

In this Appendix, we report the additional simulation results for the same simulation design in Section 7 with the conventional optimal weighting $W = \hat{\Sigma}_{11}^{-1}$ and the moderately weak instrument setup.

Table 5 reports the simulation results with the correct-specification optimal weighting matrix $W = \hat{\Sigma}_{11}^{-1}$. Comparing Tables 2 (with the efficient ME weight $W = \hat{\Sigma}_{11,2}^{-1}$) and 5, we found that SD of the GMM estimator and the lengths of CI with misspecification-robust standard errors under $W = \hat{\Sigma}_{11,2}^{-1}$ is smaller than those of $W = \hat{\Sigma}_{11}^{-1}$, especially when the degree of misspecification is large ($\delta \in \{1, 2\}$), but they are nearly identical when $\delta \in \{0, 0.5\}$. We note that the population

pseudo-true values are meaningfully different only when the misspecification is large, i.e., $\delta = 2$. Even when $W = \hat{\Sigma}_{11}^{-1}$ is used, the oracle ME bound is identical, the RSS-ME median CI lengths, and DR CI lengths are very similar with these two weights across all (n, δ) .

Table 5: Monte Carlo results under $W = \hat{\Sigma}_{11}^{-1}$

Estimator	$\delta = 0$			$\delta = 0.5$			$\delta = 1$			$\delta = 2$		
	SD	Cov	Len	SD	Cov	Len	SD	Cov	Len	SD	Cov	Len
$n = 200$												
GMM + Conv. SE	1.00	0.942	0.384	0.98	0.929	0.347	1.57	0.867	0.471	2.99	0.870	0.902
GMM + Robust SE	1.00	0.947	0.394	0.98	0.974	0.457	1.57	0.983	0.740	2.99	0.982	1.461
Oracle ME	0.93	0.944	0.367	0.87	0.949	0.349	1.08	0.945	0.422	1.78	0.943	0.695
RSS ME median	1.04	0.968	0.455	1.01	0.975	0.477	1.60	0.978	0.725	3.07	0.981	1.401
HH Bootstrap	1.00	0.942	0.393	0.90	0.926	0.357	1.24	0.872	0.483	2.39	0.875	0.925
DR Bootstrap	1.02	0.943	0.400	1.02	0.951	0.404	1.55	0.941	0.604	3.02	0.947	1.181
$n = 500$												
GMM + Conv. SE	0.62	0.955	0.245	0.63	0.919	0.223	0.99	0.875	0.303	1.85	0.881	0.585
GMM + Robust SE	0.62	0.957	0.249	0.63	0.978	0.296	0.99	0.984	0.486	1.85	0.988	0.952
Oracle ME	0.58	0.945	0.232	0.56	0.945	0.221	0.69	0.945	0.267	1.09	0.952	0.440
RSS ME median	0.63	0.962	0.263	0.64	0.967	0.277	1.00	0.967	0.429	1.86	0.965	0.827
HH Bootstrap	0.62	0.954	0.247	0.56	0.918	0.224	0.77	0.874	0.305	1.49	0.880	0.591
DR Bootstrap	0.63	0.955	0.249	0.64	0.957	0.254	0.96	0.939	0.382	1.87	0.946	0.742
$n = 1000$												
GMM + Conv. SE	0.43	0.953	0.174	0.44	0.922	0.158	0.69	0.877	0.216	1.34	0.874	0.415
GMM + Robust SE	0.43	0.955	0.175	0.44	0.983	0.211	0.69	0.986	0.347	1.34	0.983	0.679
Oracle ME	0.40	0.958	0.164	0.39	0.948	0.156	0.47	0.950	0.189	0.79	0.951	0.311
RSS ME median	0.44	0.957	0.180	0.45	0.964	0.191	0.70	0.961	0.296	1.35	0.956	0.568
HH Bootstrap	0.44	0.953	0.175	0.40	0.921	0.159	0.54	0.881	0.218	1.04	0.873	0.418
DR Bootstrap	0.44	0.955	0.176	0.45	0.953	0.180	0.67	0.942	0.270	1.31	0.939	0.522

Notes. See Table 1 for further notes. SD is normalized relative to the $W = I$, $\delta = 0$, $n = 200$ benchmark.

Tables 6 and 7 report simulation results under the moderately weak instrument setup. In particular, we set the first-stage coefficient so that the scaled concentration parameter (μ^2/m) is 10 for $n = 200$. We note that our theory does not cover the weak instruments setup. We report the median bootstrap standard deviation as a robust measure in this setup.

The results are qualitatively similar to the strong instrument cases; conventional SE/HH bootstrap undercover as δ increases, the misspecification-robust SE Wald CI/Sample-splitting methods, and DR percentile CI coverage are close to nominal coverage. The oracle ME has smaller SD/lengths, while the coverage for weak instruments slightly undercover $n = 200$, although the coverage improves when n increases. Using the efficient $W = \hat{\Sigma}_{11,2}$ performs slightly worse (larger SD, lower coverage) in small samples $n = 200, 500$ because of the estimation noise for $\hat{\Sigma}_{11,2}$ under weak instruments, but the gap decreases when $n = 1000$.

Table 6: Monte Carlo results under weak instrument setup ($W = I$)

Estimator	$\delta = 0$			$\delta = 0.5$			$\delta = 1$			$\delta = 2$		
	SD	Cov	Len	SD	Cov	Len	SD	Cov	Len	SD	Cov	Len
$n = 200$												
GMM + Conv. SE	1.00	0.946	0.847	1.51	0.898	1.055	2.92	0.859	2.036	6.28	0.844	4.173
GMM + Robust SE	1.00	0.950	0.873	1.51	0.949	1.275	2.92	0.917	2.485	6.28	0.918	5.135
Oracle ME	0.89	0.920	0.781	0.93	0.936	0.846	1.34	0.927	1.169	2.47	0.931	2.061
RSS ME median	1.08	0.989	1.393	1.53	0.980	1.854	2.99	0.965	3.398	6.24	0.958	6.950
HH Bootstrap	0.95	0.951	0.914	1.20	0.914	1.140	2.36	0.895	2.200	4.81	0.898	4.524
DR Bootstrap	1.03	0.950	0.989	1.55	0.950	1.466	3.15	0.949	2.930	6.50	0.949	6.087
$n = 500$												
GMM + Conv. SE	0.59	0.958	0.541	0.94	0.891	0.690	1.74	0.893	1.354	3.63	0.892	2.865
GMM + Robust SE	0.59	0.961	0.548	0.94	0.949	0.821	1.74	0.942	1.623	3.63	0.945	3.377
Oracle ME	0.54	0.938	0.494	0.57	0.945	0.535	0.78	0.948	0.739	1.40	0.936	1.303
RSS ME median	0.62	0.981	0.650	0.97	0.966	0.940	1.75	0.956	1.799	3.63	0.956	3.728
HH Bootstrap	0.58	0.958	0.560	0.74	0.896	0.713	1.46	0.904	1.411	3.09	0.910	2.972
DR Bootstrap	0.59	0.960	0.576	0.89	0.942	0.867	1.79	0.951	1.738	3.74	0.957	3.602
$n = 1000$												
GMM + Conv. SE	0.41	0.951	0.388	0.62	0.898	0.490	1.24	0.899	0.966	2.54	0.902	2.041
GMM + Robust SE	0.41	0.952	0.390	0.62	0.948	0.581	1.24	0.936	1.157	2.54	0.952	2.404
Oracle ME	0.36	0.946	0.349	0.39	0.956	0.378	0.55	0.948	0.523	0.96	0.941	0.922
RSS ME median	0.42	0.965	0.424	0.63	0.959	0.616	1.24	0.948	1.207	2.54	0.959	2.498
HH Bootstrap	0.41	0.950	0.395	0.51	0.898	0.496	1.01	0.894	0.983	2.15	0.911	2.079
DR Bootstrap	0.41	0.952	0.401	0.61	0.952	0.596	1.24	0.942	1.198	2.57	0.955	2.495

Notes. We consider (i) standard GMM estimator with the conventional SE; (ii) GMM with the misspecification-robust SE; (iii) the oracle ME estimator with efficient SE, $V_{ME}(W)$; (iv) Repeated sample-splitting ME estimator (RSS ME), the median repeated sample-split estimates with ± 1.96 SD; (v) Hall and Horowitz (1996) (HH) percentile bootstrap; and (vi) the Double-Recentered percentile bootstrap (DR). For GMM, Oracle ME, and RSS ME, “SD” is the MC standard deviation of the point estimate across MC draws; for HH and DR bootstrap, “SD” is the median bootstrap standard deviation across MC draws. “SD” is reported relative to the correctly-specified benchmark SD (GMM with $\delta = 0, W = I$), so values < 1 (> 1) indicate higher (lower) precision than this benchmark. “Cov” is empirical 95% coverage; “Len” is average CI length. Based on 2,000 Monte Carlo replications, $B = 1,000$ bootstrap draws, $S = 100$ sample-split replications.

Appendix D: Dynamic Panel Data Model

We consider the following dynamic panel regression of Acemoglu, Johnson, Robinson, and Yared (2008, AJRY hereafter),

$$y_{it} = \alpha y_{it-1} + \gamma x_{it-1} + \mu_t + \delta_i + u_{it}, \quad (37)$$

where y_{it} is a measure of democracy, x_{it} is log income per capita, $i = 1, \dots, N$ denotes countries and $t = 1, \dots, T$ denotes the time periods. AJRY data set includes $N = 127$ countries over 1960-2000 at both 5-year and 10-year frequencies. Following Hansen and Lee (2021, HL hereafter), we also consider augmented specification with the additional interaction terms $x_{it-1} \times c_i$, where c_i is a country-specific dummy for “historically strong institutions” allowing income effect vary across groups as in Cervellati, Jung, Sunde, and Vischer (2014, CJSV hereafter). Same as CJSV, and HL, we consider three different measure of institutions quality: (1) the level of constraints on the

Table 7: Monte Carlo results under weak instrument setup ($W = \hat{\Sigma}_{11,2}^{-1}$)

Estimator	$\delta = 0$			$\delta = 0.5$			$\delta = 1$			$\delta = 2$		
	SD	Cov	Len	SD	Cov	Len	SD	Cov	Len	SD	Cov	Len
$n = 200$												
GMM + Conv. SE	1.00	0.943	0.841	1.47	0.895	1.026	2.91	0.852	1.879	6.23	0.818	3.686
GMM + Robust SE	1.00	0.951	0.883	1.47	0.959	1.360	2.91	0.953	2.671	6.23	0.940	5.559
Oracle ME	0.89	0.920	0.781	0.94	0.941	0.846	1.37	0.919	1.169	2.59	0.913	2.061
RSS ME median	1.09	0.987	1.382	1.46	0.982	1.885	2.88	0.973	3.593	6.02	0.974	7.333
HH Bootstrap	0.94	0.950	0.910	1.16	0.919	1.113	2.16	0.890	2.032	4.20	0.865	3.970
DR Bootstrap	1.04	0.953	0.990	1.54	0.948	1.457	3.14	0.942	2.938	6.78	0.930	6.260
$n = 500$												
GMM + Conv. SE	0.59	0.955	0.540	0.92	0.897	0.675	1.77	0.872	1.250	3.88	0.841	2.535
GMM + Robust SE	0.59	0.961	0.553	0.92	0.965	0.889	1.77	0.959	1.792	3.88	0.949	3.718
Oracle ME	0.54	0.938	0.494	0.57	0.946	0.535	0.80	0.936	0.739	1.50	0.915	1.303
RSS ME median	0.61	0.981	0.652	0.94	0.978	0.970	1.76	0.971	1.911	3.81	0.969	3.990
HH Bootstrap	0.57	0.958	0.559	0.72	0.895	0.695	1.34	0.882	1.301	2.70	0.846	2.615
DR Bootstrap	0.59	0.962	0.574	0.89	0.941	0.861	1.83	0.950	1.771	3.93	0.942	3.804
$n = 1000$												
GMM + Conv. SE	0.41	0.950	0.387	0.61	0.900	0.479	1.28	0.861	0.889	2.72	0.844	1.814
GMM + Robust SE	0.41	0.954	0.391	0.61	0.966	0.631	1.28	0.961	1.284	2.72	0.958	2.682
Oracle ME	0.36	0.946	0.349	0.39	0.954	0.378	0.56	0.939	0.523	1.03	0.925	0.922
RSS ME median	0.42	0.962	0.424	0.62	0.970	0.644	1.28	0.963	1.291	2.72	0.959	2.745
HH Bootstrap	0.41	0.952	0.394	0.50	0.897	0.487	0.93	0.863	0.904	1.90	0.842	1.837
DR Bootstrap	0.41	0.951	0.401	0.61	0.950	0.598	1.26	0.944	1.224	2.72	0.947	2.635

executive in 1900; (2) whether the country became independent before 1900; (3) whether the colony was subject to the rule of a late colonial power.

We consider the first-difference GMM estimator of Arellano and Bond (1991) across specifications and weighting matrices. When using $W = W_{AB}$ (2SLS-type) in the second column in Tables 8 and 9, we exactly replicate Hansen and Lee (2021), Table III (columns 1 and 3) and IV (Columns 1, 3, and 5) with the conventional SE (HL’s label is “Arellano-Bond SE”), and misspecification-robust SE. We also consider $W = I$ and the conventional two-step efficient weight $W = \hat{\Sigma}_{11}^{-1}$. Same as HL, we report the cluster-robust version of SEs and consider the cluster (or pairs) bootstrap for the bootstrap SD and percentile CI. Due to the small number of clusters and rank issues, we only report the i.i.d version of the ME efficiency bounds $V_{ME} = (G'\Sigma_{11,2}^{-1}G)^{-1}$, so they may not be directly comparable to cluster-robust SE. We note that our current theory does not cover the cluster-robust inference, and we leave it for a future research.

Tables 8 and 9 show some interesting findings. First, point estimates are W -sensitive, but mostly settle at the two-step in many cases. For AJRY 5-year, $\hat{\beta}_{Inc}$ moves from -0.445 ($W = I$) to -0.129 ($W = AB$) to -0.012 ($W = \hat{\Sigma}_{11}^{-1}$), and $\hat{\alpha}_{Dem}$ from -0.10 to 0.49 to 0.53 ; the same patterns for the AJRY 10-year and CJSV panels. This is consistent with HL’s Figure 1, although not explicitly mentioned, where one-step GMM with $W = I$ is a poor choice in terms of point

estimates sensitivity, whereas one-step $W = AB$ is much closer to the iterated/two-step GMM estimators.

Second, the gap between the conventional/misspecification-robust SE, and the ME efficiency bounds can be significantly different across W . J -test rejects in almost all specifications, and the misspecification-robust SE exceeds the conventional SE by 10–25% across AJRY and CJSV under $W = AB$. However, under $W = I$, the gap is much bigger (e.g., AJRY 5-year Inc_{t-1} has $\text{misspec}/\text{conv} = 0.392/0.210 = 1.87$), and under the two-step efficient weight the two SEs are very similar. The ME efficiency bound is much smaller than misspec-robust SE when $W \in \{I, AB\}$, but is similar or below when $W = \hat{\Sigma}_{11}^{-1}$ and the ME-GMM bootstrap SD tracks the bound to within roughly 20% at $W \in \{I, AB\}$.

We also found there is no strong evidence of either an income effect or institutional heterogeneity in the income-democracy relationship. HL found that there is no strong evidence of the income heterogeneity based on the iterated GMM with the misspecification-robust SE. When, $W = AB$ or $\hat{\Sigma}_{11}^{-1}$, the DR percentile CIs tell a similar story: four out of six confidence intervals for the interaction coefficient excludes zero, but three only marginally significant, except $[0.17, 0.046]$ (No Late Colonial, $W = \hat{\Sigma}_{11}^{-1}$).

Table 8: Dynamic panel: AJRY income–democracy, 5-year and 10-year

Coefficient	Quantity	Weight matrix		
		I (one-step)	AB (one-step)	$\widehat{\Sigma}_{11}^{-1}$ (two-step)
<i>AJRY 5-year</i> ($N = 127, m = 55$)				
α (Dem_{t-1})	GMM point	−0.099	0.489	0.528
	Conv. SE	(0.111)	(0.085)	(0.047)
	Misspec. SE	(0.211)	(0.095)	(0.037)
	ME efficiency bound	(0.048)	(0.048)	(0.048)
	ME-GMM Boot SD	(0.060)	(0.048)	(0.048)
	DR 95% CI	[−0.46,0.38]	[0.28,0.69]	[0.43,0.64]
γ (Income_{t-1})	GMM point	−0.445	−0.129	−0.012
	Conv. SE	(0.210)	(0.076)	(0.047)
	Misspec. SE	(0.392)	(0.088)	(0.037)
	ME efficiency bound	(0.091)	(0.091)	(0.091)
	ME-GMM Boot SD	(0.111)	(0.086)	(0.095)
	DR 95% CI	[−1.34,0.33]	[−0.35,0.05]	[−0.12,0.10]
J p -value		0.002	0.006	0.042
<i>AJRY 10-year</i> ($N = 118, m = 15$)				
α (Dem_{t-1})	GMM point	−0.164	0.226	0.274
	Conv. SE	(0.196)	(0.123)	(0.111)
	Misspec. SE	(0.428)	(0.125)	(0.099)
	ME efficiency bound	(0.087)	(0.087)	(0.087)
	ME-GMM Boot SD	(0.258)	(0.129)	(0.135)
	DR 95% CI	[−1.16,0.84]	[−0.03,0.53]	[0.01,0.54]
γ (Income_{t-1})	GMM point	−0.832	−0.318	−0.274
	Conv. SE	(0.425)	(0.180)	(0.169)
	Misspec. SE	(0.773)	(0.183)	(0.179)
	ME efficiency bound	(0.166)	(0.166)	(0.166)
	ME-GMM Boot SD	(0.605)	(0.262)	(0.280)
	DR 95% CI	[−3.27,1.14]	[−0.85,0.04]	[−0.85,0.08]
J p -value		0.004	0.020	0.078

Notes. We report point estimates of α (Democracy_{it-1}) and γ (Income_{it-1}) in the dynamic panel model of AJRY (Acemoglu, Johnson, Robinson, and Yared, 2008) (equation (37)). We consider first-differenced dynamic panel GMM estimator with standard errors and bootstrap SD (in parentheses). We report the cluster-robust version of SEs and consider the cluster (or pairs) bootstrap for the bootstrap SD and percentile CI with $B = 1,000$ bootstrap draws. “Conv. SE” is the conventional (Arellano–Bond SE in HL). “Misspec. SE” is the misspecification-robust GMM SEs. “ME efficiency bound” is the oracle ME standard error $\widehat{V}_{ME}(W)^{1/2}/\sqrt{n}$ with $V_{ME}(W) = (G'\Sigma_{11,2}^{-1}G)^{-1}$ as in Corollary 2). “ME-GMM Boot SD” is the SD of the bootstrap ME-GMM estimator, and we use MAD-based robust scale $1.4826 \times \text{median}|b - \text{med}(b)|$. “DR 95% CI” is the double-recentered percentile interval and is reported in bracket.

Table 9: Dynamic panel: CJSV 5-year income–democracy with institutional interactions

Coefficient		Weight matrix		
		I (one-step)	AB (one-step)	$\hat{\Sigma}_{11}^{-1}$ (two-step)
<i>Constraints</i> ($N = 79, m = 56$)				
γ (Inc_{t-1})	GMM point	−0.564	−0.417	−0.400
	Conv. SE	(0.316)	(0.194)	(0.093)
	Misspec. SE	(0.897)	(0.221)	(0.117)
	ME efficiency bound	(0.091)	(0.091)	(0.091)
	ME-GMM Boot SD	(0.131)	(0.124)	(0.127)
	DR 95% CI	[−3.22,1.14]	[−1.07,0.00]	[−0.80,−0.03]
β ($\text{Inc}_{t-1} \times c_i$)	GMM point	0.453	0.345	0.294
	Conv. SE	(0.196)	(0.162)	(0.066)
	Misspec. SE	(0.353)	(0.168)	(0.060)
	ME efficiency bound	(0.151)	(0.151)	(0.151)
	ME-GMM Boot SD	(0.260)	(0.226)	(0.242)
	DR 95% CI	[−0.37,1.63]	[0.03,0.82]	[−0.08,0.49]
J p -value		0.026	0.032	0.030
<i>Independence</i> ($N = 99, m = 56$)				
γ (Inc_{t-1})	GMM point	−0.309	−0.270	−0.085
	Conv. SE	(0.206)	(0.113)	(0.067)
	Misspec. SE	(0.447)	(0.134)	(0.070)
	ME efficiency bound	(0.082)	(0.082)	(0.082)
	ME-GMM Boot SD	(0.090)	(0.096)	(0.113)
	DR 95% CI	[−1.45,0.67]	[−0.58,−0.02]	[−0.28,0.11]
β ($\text{Inc}_{t-1} \times c_i$)	GMM point	0.257	0.224	0.173
	Conv. SE	(0.147)	(0.121)	(0.054)
	Misspec. SE	(0.145)	(0.125)	(0.047)
	ME efficiency bound	(0.136)	(0.136)	(0.136)
	ME-GMM Boot SD	(0.211)	(0.201)	(0.156)
	DR 95% CI	[−0.12,0.73]	[−0.06,0.55]	[0.05,0.33]
J p -value		0.017	0.029	0.037
<i>No Late Colonial</i> ($N = 100, m = 56$)				
γ (Inc_{t-1})	GMM point	−0.346	−0.303	−0.151
	Conv. SE	(0.200)	(0.110)	(0.071)
	Misspec. SE	(0.417)	(0.121)	(0.056)
	ME efficiency bound	(0.081)	(0.081)	(0.081)
	ME-GMM Boot SD	(0.092)	(0.090)	(0.085)
	DR 95% CI	[−1.43,0.39]	[−0.62,−0.07]	[−0.32,0.03]
β ($\text{Inc}_{t-1} \times c_i$)	GMM point	0.367	0.318	0.299
	Conv. SE	(0.168)	(0.122)	(0.061)
	Misspec. SE	(0.176)	(0.130)	(0.055)
	ME efficiency bound	(0.129)	(0.129)	(0.129)
	ME-GMM Boot SD	(0.197)	(0.205)	(0.154)
	DR 95% CI	[0.02,0.97]	[0.04,0.63]	[0.17,0.46]
J p -value		0.031	0.090	0.148

Notes. Same model as in Table 8 (equation (37)), augmented with the additional interaction terms $x_{it-1} \times c_i$, where $c_i \in \{\textit{Constraints}, \textit{Independence}, \textit{NoLateColonial}\}$ is a country-specific dummy for “historically strong institutions” as in CJSV (Cervellati, Jung, Sunde, and Vischer, 2014).