Supplement to "Efficient GMM and Weighting Matrix under Misspecification"

Byunghoon Kang*

October 15, 2026

Abstract

This supplement contains additional theoretical results, estimators, simulations, and empirical results to accompany the main paper, "Efficient GMM and Weighting Matrix under Misspecification." Appendix A provides additional details on the linear IV example. Appendix B considers worst-case variance bounds over misspecification and identification strength. Appendix C considers the split-sample ME estimator. Appendix D reports additional simulations. Appendix E contains the dynamic panel data application of Arellano and Bond (1991). Appendix F provides the decomposition of the ME efficiency gain.

A Additional Details on the Linear IV Example

This appendix provides the closed-form expressions of the ME-GMM estimator, the bootstrap ME-GMM estimator, and the double-recentered (DR) bootstrap estimator in the linear IV example of Sections 3.3 and 4 of the main paper.

Based on the FOC from (19) in the main paper and using the linearity of the moment condition, the ME-GMM has a closed-form

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

$$\widehat{\theta}_{ME-GMM}(\Delta) = (\frac{1}{n}X'Z\Delta_{11}Z'X)^{-1}\Big(X'Z\Delta_{11}(\frac{1}{n}Z'Y - E[g(X_i,\theta_W)]) + X'Z\Delta_{12}(\frac{1}{n}vec(X'Z) - E[vec(X_iZ_i')])\Big) \quad (A.1)$$

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,

$$\widehat{\theta}_{ME-GMM}(\Delta) = (\frac{1}{n}X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}(\frac{1}{n}Z'Y - E[g(X_i,\theta_W)]). \quad (A.2)$$

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 $\{Y_i, X_i, Z_i\}_{i=1}^n$, the Hall and Horowitz (1996) version of the recentered bootstrap TSLS estimator is

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

where $\widehat{e} = Y - X\widehat{\theta}_{TSLS}$. This is essentially related to a version of recentered bootstrap estimator of (A.2).

Our proposed bootstrap ME-GMM estimator in (20) of the main paper has the closed form

$$\widehat{\theta}^*_{ME}(W) = (X^{*\prime}Z^*\widehat{\Sigma}^{-1}_{11,2}Z^{*\prime}X^*)^{-1}\Big(X^{*\prime}Z^*\widehat{\Sigma}^{-1}_{11,2}(Z^{*\prime}Y^* - Z'\widehat{e}) - X^{*\prime}Z^*\widehat{\Sigma}^{-1}_{11,2}\widehat{\Sigma}_{12}\widehat{\Sigma}^{-1}_{22}(vec(X^{*\prime}Z^*) - vec(X'Z))\Big) \quad (A.3)$$

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

with $\widehat{\Sigma}_{11,2}$.

Since the sample and population Jacobians do not depend on θ, $\widehat{\theta}^*_{DR}$ with TSLS weights $W = (Z'Z)^{-1}$ admits a closed-form solution by direct algebra from (23) of the main paper:

$$\begin{aligned}\widehat{\theta}^*_{DR} &= \big(X'Z(Z'Z)^{-1}Z^{*\prime}X^*\big)^{-1}\Big(X'Z(Z'Z)^{-1}(Z^{*\prime}Y^* - Z'\hat{e}) \\ &+ (\hat{e}'Z(Z'Z)^{-1}\otimes I_p)\mathrm{vec}(X^{*\prime}Z^* - X'Z)\Big).\end{aligned}$$

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

B Worst-Case Variance Bounds over Misspecification and Identification Strength

In this section, we consider the inference on θ_W without assuming that the recentering vector $\gamma(W) = (\gamma_1(W)', vec(\gamma_2(W)')')'$ is known. We assume only that the support of γ is known and consider a 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 includes 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 FOC of the overidentified GMM conditions; $\gamma_2'W\gamma_1 = 0$, if γ is equal to the true value $\gamma(W)$.

Define the parameter space

$$\begin{aligned}\overline{\Gamma}_W = \big\{\gamma = (\gamma_1', vec(\gamma_2')')' \in \mathbb{R}^{m+mp} : 0 \le ||\gamma_1|| \le \overline{\gamma}_1,\ \ \underline{\gamma}_2 \le ||\gamma_2|| \le \overline{\gamma}_2, \\ \underline{\gamma}_2 > 0,\ \ \gamma_2'W\gamma_1 = 0\big\}.\end{aligned} \tag{B.1}$$

$\overline{\gamma}_1$ 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 $\overline{\gamma}_1, \underline{\gamma}_2$ in this paper, but practically we can consider $\overline{\gamma}_1$ 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. 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 $\widehat{\theta}_{ME}(W)$, the parameter space $\overline{\Gamma}_W$ only requires restrictions on γ_2. This reduces not only computational costs, but also the conservativeness of the CI considered here.

For any $\gamma \in \overline{\Gamma}_W$, we can define the ME-estimator in (19) of the main paper as follows;

$$\widehat{\theta}_{ME}(W;\gamma) = \arg\min_{\theta} \left(\psi_n(\theta) - \gamma\right)' \widehat{\Sigma}(\widehat{\theta}_W)^{-1} \left(\psi_n(\theta) - \gamma\right) \tag{B.2}$$

where $\psi_n(\theta) = \frac{1}{n}\sum_{i=1}^n \psi(X_i,\theta), \psi(X_i,\theta) = (g(X_i,\theta)', vec(G(X_i,\theta)')')'$, and $\widehat{\Sigma}(\widehat{\theta}_W)$ is a consistent estimator of $\Sigma(\theta_W)$ with the preliminary GMM estimator $\widehat{\theta}_W$. Although we can consider the pseudo-true value $\theta_W(\gamma)$ as a minimizer of the population objective function of (B.2), for simplicity, we assume that for all $\gamma \in \overline{\Gamma}_W$, there exists a 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

$$\sqrt{n}(\widehat{\theta}_{ME}(W;\gamma) - \theta_W(\gamma)) \xrightarrow{d} N(0, V_{ME}(W;\gamma)) \tag{B.3}$$

where

$$\begin{aligned} V_{ME}(W;\gamma) = \big(\Gamma(\theta_W(\gamma))'\Sigma(\theta_W)^{-1}\Gamma(\theta_W(\gamma))\big)^{-1}\Gamma(\theta_W(\gamma))'\Sigma(\theta_W)^{-1}\Sigma_\gamma \\ \times \Sigma(\theta_W)^{-1}\Gamma(\theta_W(\gamma))\big(\Gamma(\theta_W(\gamma))'\Sigma(\theta_W)^{-1}\Gamma(\theta_W(\gamma))\big)^{-1}, \end{aligned}$$

and $\Sigma_\gamma = E[\psi(X_i,\theta_W(\gamma))\psi(X_i,\theta_W(\gamma))'] - \gamma\gamma'$. Note that at the true values $\gamma = \gamma(W)$, we have $\theta_W(\gamma(W)) = \theta_W$ since $E[\psi(X_i,\theta_W)] = \gamma(W)$ and $\theta_W(\gamma)$ is unique, 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 $\sup_{\gamma\in\overline{\Gamma}_W} V_{ME}(W;\gamma)$ uniformly over $\gamma \in \overline{\Gamma}_W$. 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 intervals as follows

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

where $CI_{1-\alpha}(\gamma) = \widehat{\theta}_{ME}(W;\gamma) \pm z_{1-\alpha/2}\sqrt{\widehat{V}_{ME}(W;\gamma)/n}$ 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) allows researchers to do transparent sensitivity analysis for assumptions about misspecification and identification strength. Since for each $\gamma \in \overline{\Gamma}_W$, $CI_{1-\alpha}(\gamma)$ covers $\theta_W(\gamma)$ with probability $1-\alpha$ asymptotically, and $\gamma(W) \in \overline{\Gamma}_W$, we have valid coverage for θ_W,

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

Conley et al. (2012) consider a similar type of CI based on TSLS estimates $\widehat{\theta}(\gamma)$ and the conventional TSLS standard errors, with γ denoting the direct effect of instruments. The key difference from our approach is that we use the misspecification-efficient estimator $\widehat{\theta}_{ME}$ 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 exploit additional information from the Jacobian with the potential efficiency gains from the moment and Jacobian correlations.

C 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 the "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, which were 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 their sample analogs, which suffers from a degeneracy in our setup by collapsing ME estimator back to the standard GMM estimator that invalidates the ME variance reduction.

We consider a sample-splitting ME estimator which uses one-half of a sample to estimate $\gamma(W) = (\gamma_1(W)', vec(\gamma_2(W)')')'$. Then estimated parameters $\widehat{\gamma}(W)$ are 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 use the first-half of the sample to estimate the population-first stage $E[Z'Z]^{-1}E[Z'X]$ and use it in the IV estimates in the other half sample.

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

1. Randomly divide the samples into $K = 2$ independent subsamples of roughly equal size $n/2$, and denote $\{Y_1, X_1\}, \{Y_2, X_2\}$ as data matrices for each subsample.

2. Use all observations with $\{Y_2, X_2\}$ to estimate the $\gamma(W) = (\gamma_1(W)', vec(\gamma_2(W)')')'$ with the GMM estimator $\widehat{\theta}_W$ and let these leave-one-fold-out estimators as $\widehat{\gamma}_{-1}(W)$.

3. Solve the single-split ME estimator $\widehat{\theta}^{SS}(W)$ using observations $\{X_1, Y_1\}$ with the $\widehat{\gamma}_{-1}(W)$, and calculate the standard error using "efficient" misspecification-robust variance formula $\widehat{V}_{ME}(W)$.

4. Repeat the above procedure, $s = 1, ..., S$, and set $\widehat{\theta}^{RSS}(W) = \frac{1}{S}\sum_{s=1}^{S}\widehat{\theta}_s^{SS}(W)$, $SD(\widehat{\theta}^{RSS}(W))^2 = \frac{1}{S-1}\sum_{s=1}^{S}(\widehat{\theta}_s^{SS}(W) - \widehat{\theta}^{RSS}(W))^2$. We also use the variance estimators considered in Chernozhukov et al. (2018) based on median, i.e., $median\{\widehat{V}_{ME,s}(W) + (\widehat{\theta}_s^{SS}(W) - \widehat{\theta}_{median}^{SS}(W))(\widehat{\theta}_s^{SS}(W) - \widehat{\theta}_{median}^{SS}(W))'\}_{s=1}^{S}$.

Because the sample size is $n/2$ 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 $W = \Sigma_{11,2}^{-1}$ in the linear model, we don't need the recentering part $\gamma_1(W)$ due to the FOC with $W^* = \Sigma_{11,2}^{-1}$. This helps reduce computational costs and estimation noise 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.

Table C.1 reports the Monte Carlo results of the repeated sample-splitting ME estimator for the simulation design in Section 6 under $W = I$ and $W = \widehat{\Sigma}_{11,2}^{-1}$, which complement Table 1 in the main paper. We report the median of the $S = 100$

repeated sample-split estimates (RSS ME median) with the Wald CI based on the median variance estimator described above. The RSS ME median CI has coverage close to or above the nominal level across all designs (0.945-0.987), and its length is larger than those of the Wald CI with the misspecification-robust SE and the DR percentile CI in Table 1, as expected since each split uses only half of the sample. The SD of the RSS ME median is close to that of the standard GMM estimator.

Table C.1: Monte Carlo results for the repeated sample-splitting ME estimator

	$\delta = 0$			$\delta = 0.5$			$\delta = 1$			$\delta = 2$		
RSS ME median	SD	Cov	Len	SD	Cov	Len	SD	Cov	Len	SD	Cov	Len
Panel A: $W = I$												
$n = 200$	1.04	0.974	0.466	1.04	0.976	0.471	1.67	0.963	0.710	3.18	0.970	1.378
$n = 500$	0.63	0.961	0.265	0.66	0.955	0.273	1.04	0.948	0.413	1.95	0.956	0.804
$n = 1000$	0.44	0.958	0.181	0.46	0.955	0.187	0.72	0.945	0.284	1.38	0.945	0.556
Panel B: $W = \widehat{\Sigma}_{11,2}^{-1}$												
$n = 200$	1.04	0.971	0.458	0.97	0.982	0.480	1.52	0.985	0.737	3.01	0.987	1.440
$n = 500$	0.63	0.958	0.264	0.61	0.975	0.280	0.95	0.977	0.439	1.88	0.978	0.858
$n = 1000$	0.44	0.957	0.180	0.43	0.971	0.195	0.67	0.973	0.304	1.32	0.970	0.600

Notes. Repeated sample-splitting ME estimator (RSS ME): the median of the repeated sample-split estimates with ± 1.96 SD. "SD" is the MC standard deviation of the point estimate across MC draws, reported relative to the correctly-specified benchmark SD as in Table 1 (GMM with $\delta = 0, W = I$, $n = 200$). "Cov" is empirical 95% coverage; "Len" is median CI length. Based on 2,000 Monte Carlo replications and $S = 100$ sample-split replications. See Table 1 for the results of the other estimators under the same designs.

Tables C.2 and C.3 report the RSS ME median estimates for the empirical applications in Section 7, with the same specifications and weighting matrices as in Tables 2 and 3. The RSS ME median estimates are close to the GMM point estimates, while their standard errors are larger than the misspecification-robust SE.

Table C.2: Repeated Sample-Splitting (RSS) ME: Card (1995)

	exp/exp^2 as exogenous		exp/exp^2 as endogenous	
	(1)	(2)	(3)	(4)
$W = I$ (identity)	0.1831	0.1899	0.1792	0.1786
	(0.0648)	(0.0619)	(0.0646)	(0.0645)
$W = (Z'Z)^{-1}$ (TSLS)	0.1748	0.1740	0.1693	0.1726
	(0.0674)	(0.0580)	(0.0613)	(0.0592)
$W = \widehat{\Sigma}_{11,2}^{-1}$ (efficient)	0.1619	0.1692	0.1642	0.1654
	(0.0650)	(0.0602)	(0.0632)	(0.0603)

Notes. We report the RSS ME median estimates of the returns to education with standard errors (in parentheses), for the specifications in Table 2; see the notes in Table 2 for the specifications and the other estimators. 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). $n = 3010$.

Table C.3: Repeated Sample-Splitting (RSS) ME: Angrist and Krueger (1991)

	(1) QOB (m=3)		(2) + QOB × YOB (m=30)		(3) + QOB × State (m=180)	
	basic	flexible	basic	flexible	basic	flexible
$W = I$ (identity)	0.1102	0.1099	0.0828	0.0804	0.0972	0.0890
	(0.0342)	(0.0366)	(0.0282)	(0.0337)	(0.0243)	(0.0264)
$W = (Z'Z)^{-1}$ (TSLS)	0.0998	0.1012	0.0734	0.0787	0.0774	0.0740
	(0.0332)	(0.0343)	(0.0303)	(0.0293)	(0.0255)	(0.0202)
$W = \widehat{\Sigma}_{11,2}^{-1}$ (efficient)	0.0998	0.0979	0.0781	0.0707	0.0795	0.0754
	(0.0322)	(0.0348)	(0.0288)	(0.0300)	(0.0267)	(0.0215)

Notes. We report the RSS ME median estimates of the returns to education with standard errors (in parentheses), for the specifications in Table 3; see the notes in Table 3 for the specifications and the other estimators. 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). $n = 329,509$.

D Additional Simulations

In this appendix, we report the additional simulation results for the same simulation design in Section 6 with the conventional optimal weighting $W = \widehat{\Sigma}_{11}^{-1}$ and the moderately weak instrument setup.

Table D.1 reports the simulation results with the correct-specification optimal weighting matrix $W = \widehat{\Sigma}_{11}^{-1}$. The GMM estimators under the two weights are nearly identical when $\delta \in \{0, 0.5\}$ and the population pseudo-true values are meaningfully different only when the misspecification is large, i.e., $\delta = 2$. Comparing Table 1 (with the efficient ME weight $W = \widehat{\Sigma}_{11,2}^{-1}$) and Table D.1, we find that SD of the GMM estimator under $W = \widehat{\Sigma}_{11,2}^{-1}$ is slightly smaller than that of $W = \widehat{\Sigma}_{11}^{-1}$, especially

when the degree of misspecification is large ($\delta \in \{1, 2\}$). The Wald CI with the misspecification-robust SE under-covers under $W = \widehat{\Sigma}_{11}^{-1}$ (0.90-0.92 for $\delta = 1, 2$). The oracle ME bound is identical across the two weights, and the RSS-ME median CI lengths, and DR CI lengths are very similar across all (n, δ), with DR CI coverage close to the nominal level.

Table D.1: Monte Carlo results under $W = \widehat{\Sigma}_{11}^{-1}$

	$\delta = 0$			$\delta = 0.5$			$\delta = 1$			$\delta = 2$		
Estimator	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	0.942	0.386	0.98	0.938	0.367	1.57	0.903	0.527	2.99	0.913	1.019
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	0.968	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	0.394	0.90	0.926	0.357	1.24	0.872	0.483	2.39	0.875	0.924
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	0.955	0.246	0.63	0.935	0.234	0.99	0.911	0.338	1.85	0.920	0.654
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	0.962	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	0.953	0.174	0.44	0.935	0.167	0.69	0.916	0.241	1.34	0.905	0.465
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	0.217	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	0.521

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 $\delta = 0, W = I$, $n = 200$), so values < 1 (> 1) indicate higher (lower) precision than this benchmark. "Cov" is empirical 95% coverage; "Len" is median CI length. Based on 2,000 Monte Carlo replications, $B = 1{,}000$ bootstrap draws, $S = 100$ sample-split replications.

Tables D.2 and D.3 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, while the misspecification-robust SE Wald CI/sample-splitting methods, and DR percentile CI have coverage close to the nominal level (the robust-SE Wald CI slightly undercovers). The oracle ME has smaller SD/lengths, while the coverage for weak instruments is slightly below the nominal level at $n = 200$, although the coverage improves when n increases.

Table D.2: Monte Carlo results under weak instrument setup ($W = I$)

	$\delta = 0$			$\delta = 0.5$			$\delta = 1$			$\delta = 2$		
Estimator	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	6.29	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	6.29	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	1.21	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	0.489	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

Table D.3: Monte Carlo results under weak instrument setup ($W = \widehat{\Sigma}_{11,2}^{-1}$)

	$\delta = 0$			$\delta = 0.5$			$\delta = 1$			$\delta = 2$		
Estimator	SD	Cov	Len	SD	Cov	Len	SD	Cov	Len	SD	Cov	Len
$n = 200$												
GMM + Conv. SE	1.00	0.943	0.841	1.48	0.895	1.026	2.91	0.852	1.879	6.23	0.818	3.686
GMM + Robust SE	1.00	0.946	0.862	1.48	0.941	1.234	2.91	0.923	2.458	6.23	0.916	5.220
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	1.09	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	2.031	4.20	0.865	3.970
DR Bootstrap	1.04	0.953	0.990	1.54	0.948	1.456	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	0.957	0.546	0.92	0.940	0.800	1.77	0.941	1.658	3.88	0.932	3.564
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	0.61	0.981	0.651	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	0.952	0.390	0.61	0.943	0.571	1.28	0.945	1.181	2.72	0.945	2.556
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	0.962	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

E Dynamic Panel Data Model

In this Appendix, we consider the following dynamic panel regression of Arellano and Bond (1991),

$$y_{it} = \alpha_1 y_{it-1} + \alpha_2 y_{it-2} + x_{it}'\beta + \lambda_t + \eta_i + u_{it}, \tag{E.1}$$

where y_{it} is the log of employment for company i at time t, $x_{it} = (w_{it}, w_{it-1}, k_{it}, ys_{it}, ys_{it-1})$ is a vector of strictly exogenous regressors, w_{it} is the log of the real product wage, k_{it} is the log of gross capital, and ys_{it} is the log of industry output. The data set includes an unbalanced panel of 140 UK firms over the years 1976-1984. We consider the same model as Windmeijer (2005, Table 2), which is the same as Arellano and Bond (1991, Table 4(b)).

In Table E.1, we consider the first-difference GMM estimator of Arellano and Bond (1991) across different weighting matrices and the instrument sets Z_i. Besides $W = I$, we consider the one-step weights $W = (\sum_i Z_i' H Z_i)^{-1}$ (TSLS-type), where H is a matrix with 2 on the diagonal, and -1 on the first off-diagonals. We also consider the conventional two-step efficient weight $W = (\sum_i Z_i' \hat{u}_i \hat{u}_i' Z_i)^{-1}$ with the one-step residuals $\hat{u}_i$. The ME efficient weight $W = \widehat{\Sigma}_{11,2}^{-1}$ is computed from the firm-level moments and Jacobians. Panel A uses the full instrument sets using all available lagged levels, and Panel B uses the collapsed GMM-style instruments due to the concern of many instruments (Roodman (2009)). When using one-step and two-step with the full instrument set in the second and third columns in Table E.1, we replicate Windmeijer (2005) Table 2 and Arellano and Bond (1991) Table 4(b) for the point estimates/conventional SE; we add the misspecification-robust SE, and the ME efficiency bounds, and the DR percentile bootstrap CI.

The results from Panel A should be read with caution. First of all, with the same concern as in Roodman (2009), the full instrument set does not reject (J-test p-value 0.220), while the collapsed instrument set rejects the J-test at the 5% (p-value 0.040). Furthermore, the ME bound in Panel A is much smaller than the conventional/misspecification-robust SEs, which can be due to the small-sample problems of the estimated weighting matrix with many instruments, including $\widehat{\Sigma}_{11,2}^{-1}$, rather than an efficiency gain. In Panel B with the collapsed instruments, the estimates of the AR(1) coefficient α_1 range from 0.65 to 0.85, and the misspecification-robust SE (0.31-1.60) exceeds the conventional SE (0.26-0.40) with the ME bound 0.194, in

which the ME-GMM bootstrap SD tracks it closely (0.165-0.173). Under $W = I$, the gap between the ME bound and the conventional/robust SEs is much bigger, and the DR bootstrap CIs are wider. Under the efficient weights (two-step, ME-efficient), the ME efficiency gain is much larger than in the cross-sectional examples. This is partly because the first-differencing makes the error terms and the lagged dependent variable correlated, so that the first term of the decomposition in Remark 1 is large.

Table E.1: Dynamic panel: Arellano and Bond (1991) employment equation

Coefficient		Weight matrix W: Identity I	One-step $(\sum_i Z_i'HZ_i)^{-1}$	Two-step $(\sum_i Z_i'\hat{u}_i\hat{u}_i'Z_i)^{-1}$	ME-efficient $\widehat{\Sigma}_{11,2}^{-1}$
Panel A. Full instrument set $m = 27$, $F = 3.2$, $J = 30.11$ $(p = 0.220)$					
α_1	GMM point	0.572	0.535	0.474	0.442
	Conv. SE	(0.311)	(0.166)	(0.086)	(0.284)
	Misspec. SE	(0.412)	(0.338)	(0.101)	(0.486)
	ME bound	(0.027)	(0.027)	(0.027)	(0.027)
	ME-GMM Boot SD	(0.023)	(0.024)	(0.024)	(0.024)
	DR 95% CI	$[-1.09, 1.95]$	$[-0.41, 1.26]$	$[0.18, 0.71]$	$[-0.49, 1.67]$
α_2	GMM point	0.103	−0.075	−0.053	−0.055
	Conv. SE	(0.224)	(0.068)	(0.027)	(0.100)
	Misspec. SE	(0.359)	(0.104)	(0.029)	(0.143)
	ME bound	(0.008)	(0.008)	(0.008)	(0.008)
	ME-GMM Boot SD	(0.008)	(0.008)	(0.008)	(0.008)
	DR 95% CI	$[-0.95, 0.97]$	$[-0.32, 0.23]$	$[-0.12, 0.02]$	$[-0.41, 0.20]$
Panel B. Collapsed instrument set $m = 7$, $F = 5.3$, $J = 11.63$ $(p = 0.040)$					
α_1	GMM point	0.705	0.823	0.854	0.650
	Conv. SE	(0.400)	(0.293)	(0.270)	(0.264)
	Misspec. SE	(1.601)	(0.773)	(0.306)	(0.593)
	ME bound	(0.194)	(0.194)	(0.194)	(0.194)
	ME-GMM Boot SD	(0.165)	(0.165)	(0.167)	(0.173)
	DR 95% CI	$[-6.25, 5.32]$	$[-1.84, 4.23]$	$[-0.24, 2.96]$	$[-1.75, 3.04]$
α_2	GMM point	0.449	−0.145	−0.170	−0.088
	Conv. SE	(0.312)	(0.069)	(0.065)	(0.067)
	Misspec. SE	(1.760)	(0.163)	(0.065)	(0.134)
	ME bound	(0.046)	(0.046)	(0.046)	(0.046)
	ME-GMM Boot SD	(0.042)	(0.041)	(0.041)	(0.042)
	DR 95% CI	$[-5.04, 3.63]$	$[-0.90, 0.41]$	$[-0.53, 0.05]$	$[-0.64, 0.36]$

Notes: We report point estimates of α_1, α_2, which are the coefficients of y_{it-1} and y_{it-2} in the dynamic panel model of Windmeijer (2005), Arellano and Bond (1991) (equation (E.1)). We use the first-differenced dynamic panel GMM estimator. "Conv. SE" is the conventional standard errors, and "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. "DR 95% CI" is the double-recentered percentile interval and is reported in brackets. We use $B = 2,000$ bootstrap draws. The dataset includes $N = 140$ firms, 611 first-differenced observations in 1979-1984. m is the number of excluded instruments. F is the first-stage F for $\Delta y_{i,t-1}$ and J is the Sargan-Hansen J statistic for the two-step estimator.

F The Decomposition of the ME Efficiency Gain

This appendix provides a detailed derivation of the decomposition in Remark 1 and the closed-form formula in the linear homoskedastic IV model. The closed form is useful for understanding the ratio of the conventional SE, the misspecification-robust SE and the ME efficiency bound.

We first derive the general form of the misspecification-robust GMM variance $V_{GMM}(W, \theta_W)$ which is helpful for the calculations. Consider the linear moment $g(X_i, \theta) = g(X_i, 0) + G(X_i)\theta$ with $G = E[G(X_i)]$ and $F(\theta) = 0$. Fix $W > 0$, and let $r_i = g(X_i, \theta_W) - g(\theta_W) - \Sigma_{12}\Sigma_{22}^{-1}(vec(G(X_i)') - E[vec(G(X_i)')])$. Recall that $A(W, \theta_W) = [G'W,\ g(\theta_W)'W \otimes I_p]$, and we have the FOC, $G'Wg(\theta_W) = 0$. We note that

$$A(W, \theta_W)(\psi(X_i, \theta_W) - E[\psi(X_i, \theta_W)]) = G'Wr_i + C_W(vec(G(X_i)') - E[vec(G(X_i)')]),$$

where $C_W = G'W\Sigma_{12}\Sigma_{22}^{-1} + (g(\theta_W)'W \otimes I_p)$ and $Var(r_i) = \Sigma_{11,2}$ and $Cov(r_i, vec(G(X_i)')) = 0$ by construction. Therefore, we have that

$$V_{GMM}(W, \theta_W) = (G'WG)^{-1}(G'W\Sigma_{11,2}WG + C_W\Sigma_{22}C_W')(G'WG)^{-1},$$

and $V_{ME} = (G'\Sigma_{11,2}^{-1}G)^{-1}$ as in Corollary 2.

Now we consider the linear IV model

$$Y_i = X_i\theta_0 + Z_i'\eta + \varepsilon_i, \qquad X_i = Z_i'\pi + v_i,$$

with a scalar endogenous regressor X_i, excluded instruments $Z_i \in \mathbb{R}^m$, the first-stage coefficient π, the direct effect of instruments η, and $E[\varepsilon_i \mid Z_i] = E[v_i \mid Z_i] = 0$. The moment is $g(X_i, \theta) = Z_i(Y_i - X_i\theta)$, and the Jacobian is $G(X_i) = -Z_iX_i$, $G = E[G(X_i)] = -Q_{ZZ}\pi$ where $Q_{ZZ} = E[Z_iZ_i'] > 0$. The pseudo-true value θ_W is defined by the first-order condition $G'Wg(\theta_W) = 0$, and we can deduce that $\theta_W = \theta_0 + (\pi'Q_{ZZ}\eta)/(\pi'Q_{ZZ}\pi)$ when $W \propto Q_{ZZ}^{-1}$ (TSLS). We define the residual evaluated at θ_W,

$$\varepsilon_{W,i} \equiv Y_i - X_i\theta_W = e_i + a_i$$

where $e_i = \varepsilon_i - v_i(\theta_W - \theta_0), a_i = Z_i'(\eta - \pi(\theta_W - \theta_0))$.

We further assume the following conditions for the derivation: 1) (conditional

homoskedasticity) $E[(\varepsilon_i, v_i)'(\varepsilon_i, v_i) \mid Z_i]$ does not depend on Z_i; 2) (local first-stage and the misspecification), $\pi = c_\pi/\sqrt{n}$ and $\eta = c_\eta/\sqrt{n}$. Under these assumptions, we have $E[e_i^2 \mid Z_i] = \sigma_e^2$, $E[e_i v_i \mid Z_i] = \sigma_{ev}$ and $E[v_i^2 \mid Z_i] = \sigma_v^2$ that are constants, and let $\rho = \sigma_{ev}/(\sigma_e \sigma_v)$. We impose the local misspecification for the moment as $g(\theta_W) = O(n^{-1/2})$, but not in the parameter as $\theta_W - \theta_0 = O(1)$. We also note that a_i and $g(\theta_W)$ are $O(n^{-1/2})$, and thus $\rho = \lim corr(\varepsilon_{W,i}, X_i)$.

We can then calculate the terms that appear in the variance formula of the V_{ME}, V_{GMM}, up to the order $O(n^{-1})$,

$$\Sigma_{11} = \sigma_e^2 Q_{ZZ}, \quad \Sigma_{22} = \sigma_v^2 Q_{ZZ}, \quad \Sigma_{12} = -\sigma_{ev} Q_{ZZ}, \quad \Sigma_{12}\Sigma_{22}^{-1} = -\sigma_{ev}/\sigma_v^2 I_m,$$

$$\Sigma_{11,2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = (\sigma_e^2 - \frac{\sigma_{ev}^2}{\sigma_v^2}) Q_{ZZ} = \sigma_e^2(1-\rho^2) Q_{ZZ},$$

and thus we have

$$V_{ME} = \sigma_e^2(1-\rho^2)(\pi' Q_{ZZ} \pi)^{-1}.$$

Finally, using the formula for V_{GMM}

$$V_{GMM}(W, \theta_W) = (G'WG)^{-1}(G'W\Sigma_{11,2}WG + C_W \Sigma_{22} C_W')(G'WG)^{-1}$$

with $C_W = G'W\Sigma_{12}\Sigma_{22}^{-1} + (g(\theta_W)'W \otimes I_p)$, where we can further have $G'W\Sigma_{11,2}WG = \sigma_e^2(1-\rho^2)(\pi' Q_{ZZ}\pi)$, and $C_W \Sigma_{22} C_W' = \sigma_{ev}^2/\sigma_v^2(\pi' Q_{ZZ}\pi) + \sigma_v^2 g(\theta_W)' Q_{ZZ}^{-1} g(\theta_W)$ by using the FOC, $G'W g(\theta_W) = 0$. Thus, we have

$$\begin{aligned} V_{GMM}(W, \theta_W) &= (\pi' Q_{ZZ}\pi)^{-1}(\sigma_e^2(1-\rho^2)(\pi' Q_{ZZ}\pi) + \sigma_{ev}^2/\sigma_v^2(\pi' Q_{ZZ}\pi) \\ &\qquad + \sigma_v^2 g(\theta_W)' Q_{ZZ}^{-1} g(\theta_W))(\pi' Q_{ZZ}\pi)^{-1} \\ &= (\pi' Q_{ZZ}\pi)^{-1}(\sigma_e^2(1-\rho^2) + \frac{\sigma_{ev}^2}{\sigma_v^2}) \\ &\qquad + (\pi' Q_{ZZ}\pi)^{-1} \frac{n g(\theta_W)' Q_{ZZ}^{-1} g(\theta_W)}{n(\pi' Q_{ZZ}\pi)/\sigma_v^2}. \end{aligned}$$

In sum, we have the decomposition as in (18) in the main paper,

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

where $J_n = ng(\theta_W)'\Sigma_{11}^{-1}g(\theta_W) = O(1)$, and $\mu_n^2 = n\pi' Q_{ZZ}\pi/\sigma_v^2 = O(1)$. The ME efficiency gap is approximately measured by the correlation between the moments and the Jacobian ($\frac{\rho^2}{1-\rho^2}$), and the Hansen J statistic scaled by the strength of identification $\frac{J_n}{\mu_n^2(1-\rho^2)}$.

Furthermore, we can also compare the conventional/misspecification-robust SE with the ME bound. Note that the conventional variance $V_0 = (G'\Sigma_{11}^{-1}G)^{-1} = \sigma_e^2(\pi' Q_{ZZ}\pi)^{-1}$. Therefore, to leading order, we have

$$V_{ME} \approx V_0(1-\rho^2), \quad V_{GMM} \approx V_0\Big(1+\frac{J_n}{\mu_n^2}\Big),$$

Thus,

$$\frac{SE_{ME}}{SE_{conv}} \approx \sqrt{1-\rho^2}, \qquad \frac{SE_{robust}}{SE_{conv}} \approx \sqrt{1+\frac{J_n}{\mu_n^2}}, \qquad \frac{SE_{robust}}{SE_{ME}} \approx \sqrt{\frac{1+J_n/\mu_n^2}{1-\rho^2}}. \quad \text{(F.1)}$$

The misspecification-robust SE inflates the conventional SE by the scaled J statistic J_n/μ_n^2. In practice, J_n can be replaced by the usual J statistic, and the concentration parameter μ_n^2 can be approximated by mF with the number of excluded instruments m and the first-stage F statistic. It is also worth noting that the above derivation is useful to compare the ratio of the variance formulas, but under the weak first-stage, all three variances are proportional to $(\pi' Q_{ZZ}\pi)^{-1} = O(n)$.

Supplemental Appendix References

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

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

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

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

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.

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

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

Roodman, D. (2009). A note on the theme of too many instruments, *Oxford Bulletin of Economics and Statistics*, 71(1), 135-158.

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