Supplementary Material

Supplement to "Unbiased instrumental variables estimation under known first-stage sign"

(*Quantitative Economics*, Vol. 8, No. 2, July 2017, 479–503)

Isaiah Andrews
Department of Economics, MIT

Timothy B. Armstrong
Department of Economics, Yale University

This appendix contains proofs and additional results for the main paper. Appendix A gives proofs for results stated in the main text. Appendix B derives asymptotic results for models with nonnormal errors and an unknown reduced-form error variance. Appendix C relates our results to those of Hirano and Porter (2015). Appendix D derives a lower bound on the risk of unbiased estimators in overidentified models, discusses cases in which the bound in attained, and proves that there is no uniformly minimum risk unbiased estimator in such models. Appendix F gives additional simulation results for the just-identified case, while Appendix G details our simulation design for the overidentified case.

Appendix A: Proofs

This appendix contains proofs of the results in the main text. The notation is the same as in the main text.

A.1 *Single-instrument case*

This section proves the results from Section 2, which treats the single-instrument case ($k = 1$). We prove Lemma 2.1 and Theorems 2.1, 2.2, and 2.3.

We first prove Lemma 2.1, which shows unbiasedness of $\hat{\tau}$ for $1/\pi$. As discussed in the main text, this result is known in the literature (see, e.g., pp. 181–182 of Voinov and Nikulin (1993)). We give a constructive proof based on elementary calculus (Voinov and Nikulin provide a derivation based on the bilateral Laplace transform).

Proof of Lemma 2.1. Since $\xi_2/\sigma_2 \sim N(\pi/\sigma_2, 1)$, we have

$$E_{\pi,\beta}\hat{\tau}(\xi_2, \sigma_2^2) = \frac{1}{\sigma_2}\int \frac{1-\Phi(x)}{\phi(x)}\phi(x-\pi/\sigma_2)\,dx$$

Isaiah Andrews: iandrews@mit.edu
Timothy B. Armstrong: timothy.armstrong@yale.edu

Copyright © 2017 The Authors. Quantitative Economics. The Econometric Society. Licensed under the Creative Commons Attribution-NonCommercial License 4.0. Available at http://www.qeconomics.org.
DOI: 10.3982/QE700

$$
\begin{aligned}
&= \frac{1}{\sigma_2} \int \big(1-\Phi(x)\big) \exp\big((\pi/\sigma_2)x - (\pi/\sigma_2)^2/2\big)\, dx \\
&= \frac{1}{\sigma_2} \exp\big(-(\pi/\sigma_2)^2/2\big) \Big\{ \big[\big(1-\Phi(x)\big)(\sigma_2/\pi)\exp\big((\pi/\sigma_2)x\big)\big]_{x=-\infty}^{\infty} \\
&\quad + \int (\sigma_2/\pi)\exp\big((\pi/\sigma_2)x\big)\phi(x)\, dx \Big\},
\end{aligned}
$$

using integration by parts to obtain the last equality. Since the first term in brackets in the last line is zero, this is equal to

$$
\frac{1}{\sigma_2} \int (\sigma_2/\pi)\exp\big((\pi/\sigma_2)x - (\pi/\sigma_2)^2/2\big)\phi(x)\, dx = \frac{1}{\pi}\int \phi(x-\pi/\sigma_2)\, dx = \frac{1}{\pi}. \qquad \square
$$

We note that $\hat{\tau}$ has an infinite $1+\varepsilon$ moment for $\varepsilon > 0$.

Lemma A.1. *The expectation of* $\hat{\tau}(\xi_2, \sigma_2^2)^{1+\varepsilon}$ *is infinite for all* π *and* $\varepsilon > 0$.

Proof. By calculations similar to those in the proof of Lemma 2.1,

$$
E_{\pi,\beta}\hat{\tau}\big(\xi_2, \sigma_2^2\big)^{1+\varepsilon} = \frac{1}{\sigma_2^{1+\varepsilon}} \int \frac{\big(1-\Phi(x)\big)^{1+\varepsilon}}{\phi(x)^{\varepsilon}} \exp\big((\pi/\sigma_2)x - (\pi/\sigma_2)^2/2\big)\, dx.
$$

For $x < 0$, $1-\Phi(x) \geq 1/2$, so the integrand is bounded from below by a constant times $\exp(\varepsilon x^2/2 + (\pi/\sigma_2)x)$, which is bounded away from zero as $x \to -\infty$. $\square$

Proof of Theorem 2.1. To establish unbiasedness, note that since ξ_2 and $\xi_1 - \frac{\sigma_{12}}{\sigma_2^2}\xi_2$ are jointly normal with zero covariance, they are independent. Thus,

$$
E_{\pi,\beta}\hat{\beta}_U(\xi, \Sigma) = (E_{\pi,\beta}\hat{\tau})\left[E_{\pi,\beta}\left(\xi_1 - \frac{\sigma_{12}}{\sigma_2^2}\xi_2\right)\right] + \frac{\sigma_{12}}{\sigma_2^2} = \frac{1}{\pi}\left(\pi\beta - \frac{\sigma_{12}}{\sigma_{22}}\pi\right) + \frac{\sigma_{12}}{\sigma_{22}} = \beta
$$

since $E_{\pi,\beta}\hat{\tau} = 1/\pi$ by Lemma 2.1.

To establish uniqueness, consider any unbiased estimator $\hat{\beta}(\xi, \Sigma)$. By unbiasedness

$$
E_{\pi,\beta}\big[\hat{\beta}(\xi, \Sigma) - \hat{\beta}_U(\xi, \Sigma)\big] = 0 \quad \forall \beta \in B, \pi \in \Pi.
$$

The parameter space contains an open set by assumption, so by Theorem 4.3.1 of Lehmann and Romano (2005), the family of distributions of ξ under $(\pi, \beta) \in \Theta$ is complete. Thus $\hat{\beta}(\xi, \Sigma) - \hat{\beta}_U(\xi, \Sigma) = 0$ almost surely for all $(\pi, \beta) \in \Theta$ by the definition of completeness. $\square$

Proof of Theorem 2.2. If $E_{\pi,\beta}|\hat{\beta}_U(\xi, \Sigma)|^{1+\varepsilon}$ were finite, then $E_{\pi,\beta}|\hat{\beta}_U(\xi, \Sigma) - \sigma_{12}/\sigma_2^2|^{1+\varepsilon}$ would be finite as well by Minkowski's inequality. But

$$
E_{\pi,\beta}\big|\hat{\beta}_U(\xi, \Sigma) - \sigma_{12}/\sigma_2^2\big|^{1+\varepsilon} = E_{\pi,\beta}\big|\hat{\tau}\big(\xi_2, \sigma_2^2\big)\big|^{1+\varepsilon} E_{\pi,\beta}\left|\xi_1 - \frac{\sigma_{12}}{\sigma_{22}^2}\xi_2\right|^{1+\varepsilon},
$$

and the second term is nonzero since Σ is positive definite. Thus, the $1+\varepsilon$ absolute moment is infinite by Lemma A.1. The claim that any unbiased estimator has infinite $1+\varepsilon$ moment follows from the Rao–Blackwell theorem: since $\hat{\beta}_U(\xi,\Sigma)=E[\tilde{\beta}(\xi,\Sigma)|\xi]$ for any unbiased estimator $\tilde{\beta}$ by the uniqueness of the nonrandomized unbiased estimator based on ξ, Jensen's inequality implies that the $1+\varepsilon$ moment of $|\tilde{\beta}|$ is bounded from below by the (infinite) $1+\varepsilon$ moment of $|\hat{\beta}_U|$. $\square$

We now consider the behavior of $\hat{\beta}_U$ relative to the usual 2SLS estimator (which, in the single-instrument case considered here, is given by $\hat{\beta}_{2\text{SLS}}=\xi_1/\xi_2$) as $\pi\to\infty$.

Proof of Theorem 2.3. Note that

$$\hat{\beta}_U-\hat{\beta}_{2\text{SLS}}=\left(\hat{\tau}\big(\xi_2,\sigma_2^2\big)-\frac{1}{\xi_2}\right)\left(\xi_1-\frac{\sigma_{12}}{\sigma_2^2}\xi_2\right)=\big(\xi_2\hat{\tau}\big(\xi_2,\sigma_2^2\big)-1\big)\left(\frac{\xi_1}{\xi_2}-\frac{\sigma_{12}}{\sigma_2^2}\right).$$

As $\pi\to\infty$, $\xi_1/\xi_2=\hat{\beta}_{2\text{SLS}}=\mathcal{O}_P(1)$, so it suffices to show that $\pi(\xi_2\hat{\tau}(\xi_2,\sigma_2^2)-1)=o_P(1)$ as $\pi\to\infty$. Note that, by Section 2.3.4 of Small (2010),

$$\pi\big|\xi_2\hat{\tau}\big(\xi_2,\sigma_2^2\big)-1\big|=\pi\left|\frac{\xi_2}{\sigma_2}\frac{1-\Phi(\xi_2/\sigma_2)}{\phi(\xi_2/\sigma_2)}-1\right|\le\pi\frac{\sigma_2^2}{\xi_2^2}=\frac{\pi}{\xi_2}\frac{\sigma_2^2}{\xi_2}.$$

This converges in probability to zero since $\pi/\xi_2\xrightarrow{p}1$ and $\frac{\sigma_2^2}{\xi_2}\xrightarrow{p}0$ as $\pi\to\infty$. $\square$

The following lemma regarding the mean absolute deviation of $\hat{\beta}_U$ will be useful in the next section.

Lemma A.2. *For a constant $K(\beta,\Sigma)$ depending only on Σ and β (but not on π),*

$$\pi E_{\pi,\beta}\big|\hat{\beta}_U(\xi,\Sigma)-\beta\big|\le K(\beta,\Sigma).$$

Proof. We have

$$\begin{aligned}\pi(\hat{\beta}_U-\beta)&=\pi\left[\hat{\tau}\cdot\left(\xi_1-\frac{\sigma_{12}}{\sigma_2^2}\xi_2\right)+\frac{\sigma_{12}}{\sigma_2^2}-\beta\right]=\pi\hat{\tau}\cdot\left(\xi_1-\frac{\sigma_{12}}{\sigma_2^2}\xi_2\right)+\pi\frac{\sigma_{12}}{\sigma_2^2}-\pi\beta\\&=\pi\hat{\tau}\cdot\left(\xi_1-\beta\pi-\frac{\sigma_{12}}{\sigma_2^2}(\xi_2-\pi)\right)+\pi\hat{\tau}\beta\pi-\pi\hat{\tau}\frac{\sigma_{12}}{\sigma_2^2}\pi+\pi\frac{\sigma_{12}}{\sigma_2^2}-\pi\beta\\&=\pi\hat{\tau}\cdot\left(\xi_1-\beta\pi-\frac{\sigma_{12}}{\sigma_2^2}(\xi_2-\pi)\right)+\pi(\pi\hat{\tau}-1)\left(\beta-\frac{\sigma_{12}}{\sigma_2^2}\right).\end{aligned}$$

Using this and the fact that ξ_2 and $\xi_1-\frac{\sigma_{12}}{\sigma_2^2}\xi_2$ are independent, it follows that

$$\pi E_{\pi,\beta}|\hat{\beta}_U-\beta|\le E_{\pi,\beta}\left|\xi_1-\beta\pi-\frac{\sigma_{12}}{\sigma_2^2}(\xi_2-\pi)\right|+\pi E_{\pi,\beta}|\pi\hat{\tau}-1|\left|\beta-\frac{\sigma_{12}}{\sigma_2^2}\right|,$$

where we have used the fact that $E_{\pi,\beta}\pi\hat{\tau}=1$. The only term in the above expression that depends on π is $\pi E_{\pi,\beta}|\pi\hat{\tau}-1|$. Note that this is bounded above by $\pi E_{\pi,\beta}\pi\hat{\tau}+\pi=2\pi$

(using unbiasedness and positivity of $\hat{\tau}$), so we can assume an arbitrary lower bound on π when bounding this term.

Letting $\tilde{\pi} = \pi/\sigma_2$, we have $\xi_2/\sigma_2 \sim N(\tilde{\pi}, 1)$, so that

$$\frac{\pi}{\sigma_2} E_{\pi,\beta}|\pi\hat{\tau} - 1| = \frac{\pi}{\sigma_2} E_{\pi,\beta}\left|\frac{\pi}{\sigma_2}\frac{1-\Phi(\xi_2/\sigma_2)}{\phi(\xi_2/\sigma_2)} - 1\right| = \tilde{\pi}\int\left|\tilde{\pi}\frac{1-\Phi(z)}{\phi(z)} - 1\right|\phi(z-\tilde{\pi})\,dz.$$

Let $\varepsilon > 0$ be a constant to be determined later in the proof. By (1.1) in Baricz (2008),

$$\begin{aligned}
&\tilde{\pi}^2\int_{z\geq\tilde{\pi}\varepsilon}\left|\frac{1-\Phi(z)}{\phi(z)} - \frac{1}{\tilde{\pi}}\right|\phi(z-\tilde{\pi})\,dz \\
&\quad\leq \tilde{\pi}^2\int_{z\geq\tilde{\pi}\varepsilon}\left|\frac{1}{z} - \frac{1}{\tilde{\pi}}\right|\phi(z-\tilde{\pi})\,dz + \tilde{\pi}^2\int_{z\geq\tilde{\pi}\varepsilon}\left|\frac{z}{z^2+1} - \frac{1}{\tilde{\pi}}\right|\phi(z-\tilde{\pi})\,dz.
\end{aligned}$$

The first term is

$$\tilde{\pi}^2\int_{z\geq\tilde{\pi}\varepsilon}\left|\frac{\tilde{\pi}-z}{\tilde{\pi}z}\right|\phi(z-\tilde{\pi})\,dz \leq \tilde{\pi}^2\int_{z\geq\tilde{\pi}\varepsilon}\left|\frac{\tilde{\pi}-z}{\tilde{\pi}^2\varepsilon}\right|\phi(z-\tilde{\pi})\,dz \leq \frac{1}{\varepsilon}\int|u|\phi(u)\,du.$$

The second term is

$$\begin{aligned}
\tilde{\pi}^2\int_{z\geq\tilde{\pi}\varepsilon}\left|\frac{1}{z+1/z} - \frac{1}{\tilde{\pi}}\right|\phi(z-\tilde{\pi})\,dz &= \tilde{\pi}^2\int_{z\geq\tilde{\pi}\varepsilon}\left|\frac{\tilde{\pi}-(z+1/z)}{\tilde{\pi}(z+1/z)}\right|\phi(z-\tilde{\pi})\,dz \\
&\leq \tilde{\pi}^2\int_{z\geq\tilde{\pi}\varepsilon}\frac{|\tilde{\pi}-z| + \frac{1}{\varepsilon\tilde{\pi}}}{\tilde{\pi}^2\varepsilon}\phi(z-\tilde{\pi})\,dz \\
&\leq \frac{1}{\varepsilon}\int\left(|u| + \frac{1}{\varepsilon\tilde{\pi}}\right)\phi(u)\,dz.
\end{aligned}$$

We also have

$$\begin{aligned}
&\tilde{\pi}^2\int_{z<\tilde{\pi}\varepsilon}\left|\frac{1-\Phi(z)}{\phi(z)} - \frac{1}{\tilde{\pi}}\right|\phi(z-\tilde{\pi})\,dz \\
&\quad\leq \tilde{\pi}^2\int_{z<\tilde{\pi}\varepsilon}\frac{1-\Phi(z)}{\phi(z)}\phi(z-\tilde{\pi})\,dz + \tilde{\pi}\int_{z<\tilde{\pi}\varepsilon}\phi(z-\tilde{\pi})\,dz.
\end{aligned}$$

The second term is equal to $\tilde{\pi}\Phi(\tilde{\pi}\varepsilon - \tilde{\pi})$, which is bounded uniformly over $\tilde{\pi}$ for $\varepsilon < 1$. The first term is

$$\begin{aligned}
&\tilde{\pi}^2\int_{z<\tilde{\pi}\varepsilon}\big(1-\Phi(z)\big)\exp\left(\tilde{\pi}z - \frac{1}{2}\tilde{\pi}^2\right)dz \\
&= \tilde{\pi}^2\int_{z<\tilde{\pi}\varepsilon}\int_{t\geq z}\phi(t)\exp\left(\tilde{\pi}z - \frac{1}{2}\tilde{\pi}^2\right)dt\,dz \\
&= \tilde{\pi}^2\int_{t\in\mathbb{R}}\int_{z\leq\min\{t,\tilde{\pi}\varepsilon\}}\phi(t)\exp\left(\tilde{\pi}z - \frac{1}{2}\tilde{\pi}^2\right)dz\,dt \\
&= \tilde{\pi}^2\exp\left(-\frac{1}{2}\tilde{\pi}^2\right)\int_{t\in\mathbb{R}}\phi(t)\left[\frac{1}{\tilde{\pi}}\exp(\tilde{\pi}z)\right]_{z=-\infty}^{\min\{t,\tilde{\pi}\varepsilon\}}dt
\end{aligned}$$

$$= \tilde{\pi} \exp\Big(-\frac{1}{2}\tilde{\pi}^2\Big) \int_{t\in\mathbb{R}} \phi(t) \exp\big(\tilde{\pi} \min\{t, \tilde{\pi}\varepsilon\}\big)\, dt$$
$$\le \tilde{\pi} \exp\Big(-\frac{1}{2}\tilde{\pi}^2 + \varepsilon\tilde{\pi}^2\Big).$$

For $\varepsilon < 1/2$, this is uniformly bounded over all $\tilde{\pi} > 0$. □

A.2 *Multiple instrument case*

This section proves Theorem 3.1 and extends this theorem to cover unbiased estimators that are efficient under strong-instrument asymptotics in the heteroskedastic case. In particular, we prove an extension of this theorem allowing for unbiased estimators that are asymptotically equivalent to a GMM estimator of the form $\hat{\beta}_{\text{GMM},W} = \frac{\xi_2'\hat{W}\xi_1}{\xi_2'\hat{W}\xi_2}$, where $\hat{W} = \hat{W}(\xi)$ is a data dependent weighting matrix. For Theorem 3.1, $\hat{W}$ is the deterministic matrix $Z'Z$. In models with non-homoskedastic errors, the two-step GMM estimator with weighting matrix

$$\hat{W} = \big(\Sigma_{11} - \hat{\beta}_{2\text{SLS}}(\Sigma_{12} + \Sigma_{21}) + \hat{\beta}_{2\text{SLS}}^2\Sigma_{22}\big)^{-1} \tag{10}$$

is asymptotically efficient under strong instruments. Here, $\hat{W}$ is an estimate of the inverse of the variance matrix of the moments $\xi_1 - \beta\xi_2$, which the GMM estimator sets close to zero. Let

$$\hat{w}^*_{\text{GMM},i}\big(\xi^{(b)}\big) = \frac{\xi_2^{(b)\prime}\hat{W}\big(\xi^{(b)}\big)e_ie_i'\xi_2^{(b)}}{\xi_2^{(b)\prime}\hat{W}\big(\xi^{(b)}\big)\xi_2^{(b)}}, \tag{11}$$

where

$$\hat{W}\big(\xi^{(b)}\big) = \big(\Sigma_{11} - \hat{\beta}\big(\xi^{(b)}\big)(\Sigma_{12} + \Sigma_{21}) + \hat{\beta}\big(\xi^{(b)}\big)^2\Sigma_{22}\big)^{-1}$$

for a preliminary estimator $\hat{\beta}(\xi^{(b)})$ of β based on $\xi^{(b)}$. The Rao–Blackwellized estimator formed by replacing $\hat{w}^*$ with $\hat{w}^*_{\text{GMM}}$ in the definition of $\hat{\beta}^*_{\text{RB}}$ gives an unbiased estimator that is asymptotically efficient under strong-instrument asymptotics with non-homoskedastic errors, as we now show by proving an extension of Theorem 3.1 that covers the weight matrix in (10) in addition to the matrix $Z'Z$ used in Theorem 3.1.

Consider the GMM estimator $\hat{\beta}_{\text{GMM},W} = \frac{\xi_2'\hat{W}\xi_1}{\xi_2'\hat{W}\xi_2}$, where $\hat{W} = \hat{W}(\xi)$ is a data dependent weighting matrix. For Theorem 3.1, $\hat{W}$ is the deterministic matrix $Z'Z$, while, in the extension discussed above, $\hat{W}$ is defined in (10). In both cases, $\hat{W} \xrightarrow{p} W^*$ for some positive definite matrix W^* under the strong-instrument asymptotics in the theorem. For this W^*, define the oracle weights

$$w_i^* = \pi_i\frac{\pi'W^*e_i}{\pi'W^*\pi} = \frac{\pi'W^*e_ie_i'\pi}{\pi'W^*\pi}$$

and the oracle estimator

$$\hat{\beta}^{\mathrm{o}}_{\mathrm{RB}} = \hat{\beta}_{\mathrm{RB}}\big(\xi, \Sigma; w^*\big) = \hat{\beta}_w\big(\xi, \Sigma; w^*\big) = \sum_{i=1}^{k} w_i^* \hat{\beta}_U\big(\xi(i), \Sigma(i)\big).$$

Define the estimated weights as in (11),

$$\hat{w}_i^* = \hat{w}_i^*\big(\xi^{(b)}\big) = \frac{\xi_2^{(b)\prime}\hat{W}\big(\xi^{(b)}\big)e_i e_i' \xi_2^{(b)}}{\xi_2^{(b)\prime}\hat{W}\big(\xi^{(b)}\big)\xi_2^{(b)}},$$

and the Rao–Blackwellized estimator based on the estimated weights

$$\hat{\beta}^*_{\mathrm{RB}} = \hat{\beta}_{\mathrm{RB}}\big(\xi, \Sigma; \hat{w}^*\big) = E\big[\hat{\beta}_w\big(\xi^{(a)}, 2\Sigma; \hat{w}^*\big)|\xi\big] = \sum_{i=1}^{k} E\big[\hat{w}_i^*\big(\xi_2^{(b)}\big)\hat{\beta}_U\big(\xi^{(a)}(i), 2\Sigma(i)\big)|\xi\big].$$

In the general case, we will assume that $\hat{w}_i^*(\xi^{(b)})$ is uniformly bounded (this holds for equivalence with 2SLS under the conditions of Theorem 3.1, since $\sup_{\|u\|_{Z'Z}=1} u'Z'Ze_ie_i'u$ is bounded, and one can likewise show that it holds for the two-step GMM provided Σ has full rank). Let us also define the oracle linear combination of 2SLS estimators,

$$\hat{\beta}^{\mathrm{o}}_{\mathrm{2SLS}} = \sum_{i=1}^{k} w_i^* \frac{\xi_{1,i}}{\xi_{2,i}}.$$

Lemma A.3. *Suppose that $\hat{w}$ is deterministic: $\hat{w}(\xi^{(b)}) = w$ for some constant vector w. Then $\hat{\beta}_{\mathrm{RB}}(\xi, \Sigma; w) = \hat{\beta}_w(\xi, \Sigma; w)$.*

Proof. We have

$$\hat{\beta}_{\mathrm{RB}}(\xi, \Sigma; w) = E\left[\sum_{i=1}^{k} w_i \hat{\beta}_U\big(\xi^{(a)}(i), 2\Sigma(i)\big)\Big|\xi\right] = \sum_{i=1}^{k} w_i E\big[\hat{\beta}_U\big(\xi^{(a)}(i), 2\Sigma(i)\big)|\xi\big].$$

Since $\xi^{(a)}(i) = \zeta(i) + \xi(i)$ (where $\zeta(i) = (\zeta_i, \zeta_{k+i})'$), $\xi^{(a)}(i)$ is independent of $\{\xi(j)\}_{j\neq i}$ conditional on $\xi(i)$. Thus, $E[\hat{\beta}_U(\xi^{(a)}(i), 2\Sigma(i))|\xi] = E[\hat{\beta}_U(\xi^{(a)}(i), 2\Sigma(i))|\xi(i)]$. Since $E[\hat{\beta}_U(\xi^{(a)}(i), 2\Sigma(i))|\xi(i)]$ is an unbiased estimator for β that is a deterministic function of $\xi(i)$, it must be equal to $\hat{\beta}_U(\xi(i), \Sigma(i))$, the unique nonrandom unbiased estimator based on $\xi(i)$ (where uniqueness follows by completeness since the parameter space $\{(\beta\pi_i, \pi_i)|\pi_i \in \mathbb{R}_+, \beta \in \mathbb{R}\}$ contains an open rectangle). Plugging this in to the above display gives the result. □

Lemma A.4. *Let $\|\pi\| \to \infty$ with $\|\pi\|/\min_i \pi_i = \mathcal{O}(1)$. Then $\|\pi\|(\hat{\beta}_{\mathrm{GMM},W} - \hat{\beta}^{\mathrm{o}}_{\mathrm{2SLS}}) \xrightarrow{p} 0$.*

Proof. Note that

$$\begin{aligned}\hat{\beta}_{\mathrm{GMM},W} - \hat{\beta}^{\mathrm{o}}_{\mathrm{2SLS}} &= \frac{\xi_2'\hat{W}\xi_1}{\xi_2'\hat{W}\xi_2} - \sum_{i=1}^{k} w_i^* \frac{\xi_{1,i}}{\xi_{2,i}} \\ &= \sum_{i=1}^{k}\left(\frac{\xi_2'\hat{W}e_ie_i'\xi_2}{\xi_2'\hat{W}\xi_2} - w_i^*\right)\frac{\xi_{1,i}}{\xi_{2,i}}\end{aligned}$$

$$= \sum_{i=1}^{k} \left(\frac{\xi_2' \hat{W} e_i e_i' \xi_2}{\xi_2' \hat{W} \xi_2} - \frac{\pi' W^* e_i e_i' \pi}{\pi' W^* \pi} \right) \frac{\xi_{1,i}}{\xi_{2,i}}$$

$$= \sum_{i=1}^{k} \left(\frac{\xi_2' \hat{W} e_i e_i' \xi_2}{\xi_2' \hat{W} \xi_2} - \frac{\pi' W^* e_i e_i' \pi}{\pi' W^* \pi} \right) \left(\frac{\xi_{1,i}}{\xi_{2,i}} - \beta \right),$$

where the last equality follows since $\sum_{i=1}^{k} \frac{\xi_2' \hat{W} e_i e_i' \xi_2}{\xi_2' \hat{W} \xi_2} = \sum_{i=1}^{k} \frac{\pi' W^* e_i e_i' \pi}{\pi' W^* \pi} = 1$ with probability 1. For each i, $\pi_i(\xi_{1,i}/\xi_{2,i} - \beta) = \mathcal{O}_P(1)$ and $\frac{\xi_2' \hat{W} e_i e_i' \xi_2}{\xi_2' \hat{W} \xi_2} - \frac{\pi' W^* e_i e_i' \pi}{\pi' W^* \pi} \xrightarrow{p} 0$ as the elements of π approach infinity. Combining this with the above display and the fact that $\|\pi\|/\min_i \pi_i = \mathcal{O}(1)$ gives the result. □

Lemma A.5. *Let $\|\pi\| \to \infty$ with $\|\pi\|/\min_i \pi_i = \mathcal{O}(1)$. Then $\|\pi\|(\hat{\beta}^{o}_{2SLS} - \hat{\beta}^{o}_{RB}) \xrightarrow{p} 0$.*

Proof. By Lemma A.3,

$$\|\pi\| \big(\hat{\beta}^{o}_{2SLS} - \hat{\beta}^{o}_{RB} \big) = \|\pi\| \sum_{i=1}^{k} w_i^* \left(\frac{\xi_{1,i}}{\xi_{2,i}} - \hat{\beta}_U \big(\xi(i), \Sigma(i) \big) \right).$$

By Theorem 2.3, $\pi_i(\frac{\xi_{1,i}}{\xi_{2,i}} - \hat{\beta}_U(\xi(i), \Sigma(i))) \xrightarrow{p} 0$. Combining this with the boundedness of $\|\pi\|/\min_i \pi_i$ gives the result. □

Lemma A.6. *Let $\|\pi\| \to \infty$ with $\|\pi\|/\min_i \pi_i = \mathcal{O}(1)$. Then $\|\pi\|(\hat{\beta}^{o}_{RB} - \hat{\beta}^{*}_{RB}) \xrightarrow{p} 0$.*

Proof. We have

$$\hat{\beta}^{o}_{RB} - \hat{\beta}^{*}_{RB} = \sum_{i=1}^{k} E\big[\big(w_i^* - \hat{w}_i^*(\xi^{(b)}) \big) \hat{\beta}_U \big(\xi^{(a)}(i), 2\Sigma(i) \big) | \xi \big]$$

$$= \sum_{i=1}^{k} E\big[\big(w_i^* - \hat{w}_i^*(\xi^{(b)}) \big) \big(\hat{\beta}_U \big(\xi^{(a)}(i), 2\Sigma(i) \big) - \beta \big) | \xi \big]$$

using the fact that $\sum_{i=1}^{k} w_i^* = \sum_{i=1}^{k} \hat{w}_i^*(\xi^{(b)}) = 1$ with probability 1. Thus,

$$E_{\beta,\pi} \big| \hat{\beta}^{o}_{RB} - \hat{\beta}^{*}_{RB} \big| \le \sum_{i=1}^{k} E_{\beta,\pi} \big| \big(w_i^* - \hat{w}_i^*(\xi^{(b)}) \big) \big(\hat{\beta}_U \big(\xi^{(a)}(i), 2\Sigma(i) \big) - \beta \big) \big|$$

$$= \sum_{i=1}^{k} E_{\beta,\pi} \big| w_i^* - \hat{w}_i^*(\xi^{(b)}) \big| E_{\beta,\pi} \big| \hat{\beta}_U \big(\xi^{(a)}(i), 2\Sigma(i) \big) - \beta \big|.$$

As $\|\pi\| \to \infty$, $\hat{w}_i^*(\xi^{(b)}) - w_i^* \xrightarrow{p} 0$, so since $\hat{w}_i^*(\xi^{(b)})$ is bounded, $E_{\beta,\pi}|w_i^* - \hat{w}_i^*(\xi^{(b)})| \to 0$. Thus, it suffices to show that $\pi_i E_{\beta,\pi} |\hat{\beta}_U(\xi^{(a)}(i), 2\Sigma(i)) - \beta| = \mathcal{O}(1)$ for each i. But this follows by Lemma A.2, which completes the proof. □

Appendix B: Nonnormal errors and unknown reduced-form variance

This appendix derives asymptotic results for the case with nonnormal errors and an estimated reduced-form covariance matrix. Appendix B.1 shows asymptotic unbiasedness in the weak-instrument case. Appendix B.2 shows asymptotic equivalence with 2SLS in the strong-instrument case (where, in the case with multiple instruments, the weights are chosen appropriately). The results are proved using some auxiliary lemmas, which are stated and proved in Appendix B.3.

Throughout this appendix, we consider a sequence of reduced-form estimators

$$\tilde{\xi} = \begin{pmatrix} (Z'Z)^{-1} Z'Y \\ (Z'Z)^{-1} Z'X \end{pmatrix},$$

which we assume satisfy a central limit theorem,

$$\sqrt{T}\left(\tilde{\xi} - \begin{pmatrix} \pi_T \beta \\ \pi_T \end{pmatrix}\right) \xrightarrow{d} N(0, \Sigma^*), \tag{12}$$

where π_T is a sequence of parameter values and Σ^* is a positive definite matrix. Following Staiger and Stock (1997), we distinguish between the case of weak instruments, in which π_T converges to 0 at a $\sqrt{T}$ rate, and the case of strong instruments, in which π_T converges to a vector in the interior of the positive orthant. Formally, the weak-instrument case is given by the condition that

$$\sqrt{T}\pi_T \to \pi^* \quad \text{where } \pi_i^* > 0 \text{ for all } i, \tag{13}$$

while the strong-instrument case is given by the condition that

$$\pi_T \to \pi^* \quad \text{where } \pi_i^* > 0 \text{ for all } i. \tag{14}$$

In both cases, we assume the availability of a consistent estimator $\tilde{\Sigma}$ for the asymptotic variance of the reduced-form estimators:

$$\tilde{\Sigma} \xrightarrow{p} \Sigma^*. \tag{15}$$

The estimator is then formed as

$$\begin{aligned} \hat{\beta}_{\mathrm{RB}}(\tilde{\xi}, \tilde{\Sigma}/T, \hat{w}) &= E_{\tilde{\Sigma}/T}\big[\hat{\beta}_w\big(\tilde{\xi}^{(a)}, 2\tilde{\Sigma}/T, \hat{w}\big(\tilde{\xi}^{(b)}\big)\big)|\tilde{\xi}\big] \\ &= \int \hat{\beta}_w\big(\tilde{\xi} + T^{-1/2}\tilde{\Sigma}^{1/2}\eta, 2\tilde{\Sigma}/T, \hat{w}\big(\tilde{\xi} - T^{-1/2}\tilde{\Sigma}^{1/2}\eta\big)\big)\, dP_{N(0,I_{2k})}(\eta), \end{aligned}$$

where $\tilde{\xi}^{(a)} = \tilde{\xi} + T^{-1/2}\tilde{\Sigma}^{1/2}\eta$ and $\tilde{\xi}^{(b)} = \tilde{\xi} - T^{-1/2}\tilde{\Sigma}^{1/2}\eta$ for $\eta \sim N(0, I_{2k})$ independent of $\tilde{\xi}$ and $\tilde{\Sigma}$, and we use the subscript in the expectation to denote the dependence of the conditional distribution of $\tilde{\xi}^{(a)}$ and $\tilde{\xi}^{(b)}$ on $\tilde{\Sigma}/T$. In the single-instrument case, $\hat{\beta}_{\mathrm{RB}}(\tilde{\xi}, \tilde{\Sigma}/T, \hat{w})$ reduces to $\hat{\beta}_U(\tilde{\xi}, \tilde{\Sigma}/T)$.

For the weights $\hat{w}$, we assume that $\hat{w}(\xi^{(b)})$ is bounded and continuous in $\xi^{(b)}$ with $\sum_{i=1}^k \hat{w}_i(\xi^{(b)}) = 1$ and $\hat{w}_i(a\xi^{(b)}) = \hat{w}_i(\xi^{(b)})$ for any scalar a, as holds for all the weights discussed above. Using the fact that $\hat{\beta}_U(\sqrt{a}x, a\Omega) = \hat{\beta}_U(x, \Omega)$ for any scalar a and any x and Ω, we have, under the above conditions on $\hat{w}$,

$$\begin{aligned}\hat{\beta}_{\mathrm{RB}}(\tilde{\xi}, \tilde{\Sigma}/T, \hat{w}) &= \int \hat{\beta}_w\big(\sqrt{T}\tilde{\xi} + \tilde{\Sigma}^{1/2}\eta, 2\tilde{\Sigma}, \hat{w}\big(\sqrt{T}\tilde{\xi} - \tilde{\Sigma}^{1/2}\eta\big)\big)\, dP_{N(0,I_{2k})}(\eta) \\ &= \hat{\beta}_{\mathrm{RB}}(\sqrt{T}\tilde{\xi}, \tilde{\Sigma}, \hat{w}).\end{aligned}$$

Thus, we can focus on the behavior of $\sqrt{T}\tilde{\xi}$ and $\tilde{\Sigma}$, which are asymptotically nondegenerate in the weak-instrument case.

B.1 *Weak-instrument case*

The following theorem shows that the estimator $\hat{\beta}_{\mathrm{RB}}$ converges in distribution to a random variable with mean β. Note that since convergence in distribution does not imply convergence of moments, this does not imply that the bias of $\hat{\beta}_{\mathrm{RB}}$ converges to zero. While it seems likely this stronger form of asymptotic unbiasedness could be achieved under further conditions by truncating $\hat{\beta}_{\mathrm{RB}}$ at a slowly increasing sequence of points, we leave this extension for future research.

Theorem B.1. *Let* (12), (13), *and* (15) *hold, and suppose that* $\hat{w}(\xi^{(b)})$ *is bounded and continuous in* $\xi^{(b)}$ *with* $\hat{w}_i(a\xi^{(b)}) = \hat{w}_i(\xi^{(b)})$ *for any scalar* a. *Then*

$$\hat{\beta}_{\mathrm{RB}}(\tilde{\xi}, \tilde{\Sigma}/T, \hat{w}) = \hat{\beta}_{\mathrm{RB}}(\sqrt{T}\tilde{\xi}, \tilde{\Sigma}, \hat{w}) \xrightarrow{d} \hat{\beta}_{\mathrm{RB}}\big(\xi^*, \Sigma^*, \hat{w}\big),$$

where $\xi^* \sim N((\pi^{*\prime}\beta, \pi^{*\prime})', \Sigma^*)$ *and* $E[\hat{\beta}_{\mathrm{RB}}(\xi^*, \Sigma^*, \hat{w})] = \beta$.

Proof. Since $\sqrt{T}\tilde{\xi} \xrightarrow{d} \xi^*$ and $\tilde{\Sigma} \xrightarrow{p} \Sigma^*$, the first display follows by the continuous mapping theorem as long as $\hat{\beta}_{\mathrm{RB}}(\xi^*, \Sigma^*, \hat{w})$ is continuous in ξ^* and Σ^*. Since

$$\hat{\beta}_{\mathrm{RB}}\big(\xi^*, \Sigma^*, \hat{w}\big) = \int \hat{\beta}_w\big(\xi^* + \Sigma^{*1/2}\eta, 2\Sigma^*, \hat{w}\big(\xi^* - \Sigma^{*1/2}\eta\big)\big)\, dP_{N(0,I_{2k})}(\eta) \tag{16}$$

and the integrand is continuous in ξ^* and Σ^*, it suffices to show uniform integrability over ξ^* and Σ^* in an arbitrarily small neighborhood of any point. The pth moment of the integrand in the above display is bounded by a constant times the sum over i of

$$\int\int \big|\hat{\beta}_U\big(\xi^*(i) + \Sigma^{*1/2}(i)z, 2\Sigma^*(i)\big)\big|^p \phi(z_1)\phi(z_2)\, dz_1\, dz_2 = R\big(\xi^*(i), \Sigma^*(i), 0, p\big),$$

where R is defined below in Appendix B.3. By Lemma B.1 below, this is equal to

$$\tilde{R}\left(\frac{\xi_2^*(i)}{\sqrt{\Sigma_{22}^*(i)}}, \frac{\xi_1^*(i)}{\sqrt{\Sigma_{22}^*(i)}}, \left(\frac{\Sigma_{11}^*(i)}{\Sigma_{22}^*(i)} - \frac{\Sigma_{12}^*(i)^2}{\Sigma_{22}^*(i)^2}\right)^{1/2}, -\frac{\Sigma_{12}^*(i)}{\Sigma_{22}^*(i)}, p\right),$$

which is bounded uniformly over a small enough neighborhood of any ξ^* and Σ^* with Σ^* positive definite by Lemma B.2 below as long as $p < 2$. Setting $1 < p < 2$, it follows that uniform integrability holds for (16) so that $\hat{\beta}_{\mathrm{RB}}(\xi^*, \Sigma^*, \hat{w})$ is continuous, thereby giving the result. $\square$

B.2 *Strong-instrument asymptotics*

Let $\tilde{W}(\tilde{\xi}^{(b)}, \tilde{\Sigma})$ and $\hat{W}$ be weighting matrices that converge in probability to some positive definite symmetric matrix W. Let

$$\hat{w}^*_{\mathrm{GMM},i}\big(\tilde{\xi}^{(b)}\big) = \frac{\tilde{\xi}_2^{(b)\prime}\tilde{W}\big(\tilde{\xi}^{(b)}, \tilde{\Sigma}\big)e_i e_i' \tilde{\xi}_2^{(b)}}{\tilde{\xi}_2^{(b)\prime}\tilde{W}\big(\tilde{\xi}^{(b)}, \tilde{\Sigma}\big)\tilde{\xi}_2^{(b)}},$$

where e_i is the ith standard basis vector in $\mathbb{R}^k$, and let

$$\hat{\beta}_{\mathrm{GMM},\hat{W}} = \frac{\tilde{\xi}_2'\hat{W}\tilde{\xi}_1}{\tilde{\xi}_2'\hat{W}\tilde{\xi}_1}.$$

The following theorem shows that $\hat{\beta}_{\mathrm{GMM},\hat{W}}$ and $\hat{\beta}_{\mathrm{RB}}(\sqrt{T}\tilde{\xi}, \tilde{\Sigma}, \hat{w}^*_{\mathrm{GMM}})$ are asymptotically equivalent in the strong-instrument case. For the case where $\tilde{W}(\tilde{\xi}^{(b)}, \tilde{\Sigma}) = \hat{W} = Z'Z/T$, this gives asymptotic equivalence to 2SLS.

Theorem B.2. *Let $\tilde{W}(\tilde{\xi}^{(b)}, \tilde{\Sigma})$ and $\hat{W}$ be weighting matrices that converge in probability to the same positive definite matrix W, such that $\hat{w}^*_{\mathrm{GMM},i}$ defined above is uniformly bounded over $\tilde{\xi}^{(b)}$. Then, under* (12), (14), *and* (15),

$$\sqrt{T}\big(\hat{\beta}_{\mathrm{RB}}\big(\sqrt{T}\tilde{\xi}, \tilde{\Sigma}, \hat{w}^*_{\mathrm{GMM}}\big) - \hat{\beta}_{\mathrm{GMM},\hat{W}}\big) \xrightarrow{p} 0.$$

Proof. As with the normal case, define the oracle linear combination of 2SLS estimators

$$\hat{\beta}^{\mathrm{o}}_{\mathrm{2SLS}} = \sum_{i=1}^k w_i^* \frac{\tilde{\xi}_{1,i}}{\tilde{\xi}_{2,i}},$$

where $w_i^* = \frac{\pi^{*\prime}We_ie_i'\pi^*}{\pi^{*\prime}W\pi^*}$. We have $\sqrt{T}(\hat{\beta}_{\mathrm{RB}}(\sqrt{T}\tilde{\xi}, \tilde{\Sigma}, \hat{w}^*_{\mathrm{GMM}}) - \hat{\beta}_{\mathrm{GMM},\hat{W}}) = I + II + III$, where $I \equiv \sqrt{T}(\hat{\beta}_{\mathrm{RB}}(\sqrt{T}\tilde{\xi}, \tilde{\Sigma}, \hat{w}^*_{\mathrm{GMM}}) - \hat{\beta}_{\mathrm{RB}}(\sqrt{T}\tilde{\xi}, \tilde{\Sigma}, w^*))$, $II \equiv \sqrt{T}(\hat{\beta}_{\mathrm{RB}}(\sqrt{T}\tilde{\xi}, \tilde{\Sigma}, w^*) - \hat{\beta}^{\mathrm{o}}_{\mathrm{2SLS}})$, and $III \equiv \sqrt{T}(\hat{\beta}^{\mathrm{o}}_{\mathrm{2SLS}} - \hat{\beta}_{\mathrm{GMM},\hat{W}})$.

For the first term, note that

$$\begin{aligned} I &= \sqrt{T}\sum_{i=1}^k E_{\tilde{\Sigma}}\big[\big(\hat{w}^*_{\mathrm{GMM},i}\big(\tilde{\xi}^{(b)}\big) - w_i^*\big)\hat{\beta}_U\big(\sqrt{T}\tilde{\xi}^{(a)}(i), \tilde{\Sigma}(i)\big) | \tilde{\xi}\big] \\ &= \sqrt{T}\sum_{i=1}^k E_{\tilde{\Sigma}}\big[\big(\hat{w}^*_{\mathrm{GMM},i}\big(\tilde{\xi}^{(b)}\big) - w_i^*\big)\big(\hat{\beta}_U\big(\sqrt{T}\tilde{\xi}^{(a)}(i), \tilde{\Sigma}(i)\big) - \beta\big) | \tilde{\xi}\big], \end{aligned}$$

where the last equality follows since $\sum_{i=1}^k \hat{w}^*_{\mathrm{GMM},i}(\tilde{\xi}^{(b)}) = \sum_{i=1}^k w_i^* = 1$ with probability 1. Thus, by Hölder's inequality,

$$|I| \le \sqrt{T}\sum_{i=1}^k \big(E_{\tilde{\Sigma}}\big[\big|\hat{w}^*_{\mathrm{GMM},i}\big(\tilde{\xi}^{(b)}\big) - w_i^*\big|^q | \tilde{\xi}\big]\big)^{1/q}\big(E_{\tilde{\Sigma}}\big[\big|\hat{\beta}_U\big(\sqrt{T}\tilde{\xi}^{(a)}(i), \tilde{\Sigma}(i)\big) - \beta\big|^p | \tilde{\xi}\big]\big)^{1/p}$$

for any p and q with $p, q > 1$ and $1/p + 1/q = 1$ such that these conditional expectations exist. Under (14), $\hat{w}^*_{\text{GMM},i}(\tilde{\xi}^{(b)}) \xrightarrow{p} w^*_i$ so, since $\hat{w}^*_{\text{GMM},i}(\tilde{\xi}^{(b)})$ is uniformly bounded, $E_{\tilde{\Sigma}}[|\hat{w}^*_{\text{GMM},i}(\tilde{\xi}^{(b)}) - w^*_i|^q|\tilde{\xi}]$ will converge to zero for any q. Thus, for this term, it suffices to bound

$$
\begin{aligned}
&\sqrt{T}\big(E_{\tilde{\Sigma}}\big[\big|\hat{\beta}_U\big(\sqrt{T}\tilde{\xi}^{(a)}(i), \tilde{\Sigma}(i)\big) - \beta\big|^p|\tilde{\xi}\big]\big)^{1/p} \\
&\quad = \sqrt{T}R\big(\sqrt{T}\tilde{\xi}(i), \tilde{\Sigma}(i), \beta, p\big)^{1/p} \\
&\quad = \sqrt{T}\tilde{R}\left(\frac{\sqrt{T}\tilde{\xi}_2(i)}{\sqrt{\tilde{\Sigma}_{22}(i)}}, \frac{\sqrt{T}\big(\tilde{\xi}_1(i) - \beta\tilde{\xi}_2(i)\big)}{\sqrt{\tilde{\Sigma}_{22}(i)}}, \left(\frac{\tilde{\Sigma}_{11}(i)}{\tilde{\Sigma}_{22}(i)} - \frac{\tilde{\Sigma}_{12}(i)^2}{\tilde{\Sigma}_{22}(i)^2}\right)^{1/2}, \beta - \frac{\tilde{\Sigma}_{12}(i)}{\tilde{\Sigma}_{22}(i)}, p\right)^{1/p}
\end{aligned}
$$

for R and $\tilde{R}$ as defined in Appendix B.3 below. By Lemma B.3 below, this is equal to $\sqrt{\tilde{\Sigma}_{22}(i)}/\tilde{\xi}_2(i)$ times a $\mathcal{O}_P(1)$ term for $p < 2$. Since $\sqrt{\tilde{\Sigma}_{22}(i)}/\tilde{\xi}_2(i) \xrightarrow{p} \sqrt{\Sigma^*_{22}(i)}/\pi^*_i$, it follows that the above display is also $\mathcal{O}_P(1)$. Thus, $I \xrightarrow{p} 0$.

For the second term, we have

$$
\begin{aligned}
II &= \sqrt{T}\sum_{i=1} w^*_i\left(\hat{\beta}_U\big(\sqrt{T}\tilde{\xi}(i), \tilde{\Sigma}(i)\big) - \frac{\tilde{\xi}_1(i)}{\tilde{\xi}_2(i)}\right) \\
&= \sqrt{T}\sum_{i=1} w^*_i\big(\sqrt{T}\tilde{\xi}_2(i)\hat{\tau}\big(\sqrt{T}\tilde{\xi}_2(i), \tilde{\Sigma}_{22}(i)\big) - 1\big)\left(\frac{\tilde{\xi}_1(i)}{\tilde{\xi}_2(i)} - \frac{\tilde{\Sigma}_{12}(i)}{\tilde{\Sigma}_{22}(i)}\right).
\end{aligned}
$$

For each i, $\frac{\tilde{\xi}_1(i)}{\tilde{\xi}_2(i)} - \frac{\tilde{\Sigma}_{12}(i)}{\tilde{\Sigma}_{22}(i)}$ converges in probability to a finite constant and, by Section 2.3.4 of Small (2010),

$$
\sqrt{T}\big|\sqrt{T}\tilde{\xi}_2(i)\hat{\tau}\big(\sqrt{T}\tilde{\xi}_2(i), \tilde{\Sigma}_{22}(i)\big) - 1\big| \le \sqrt{T}\frac{\tilde{\Sigma}_{22}(i)}{T\tilde{\xi}_2(i)^2} \xrightarrow{p} 0.
$$

The third term converges in probability to zero by standard arguments. We have

$$
III = \sqrt{T}\sum_{i=1}^{k}\left(w^*_i - \frac{\tilde{\xi}_2'\hat{W}e_ie_i'\tilde{\xi}_2}{\tilde{\xi}_2'\hat{W}\tilde{\xi}_2}\right)\frac{\tilde{\xi}_{1,i}}{\tilde{\xi}_{2,i}} = \sqrt{T}\sum_{i=1}^{k}\left(w^*_i - \frac{\tilde{\xi}_2'\hat{W}e_ie_i'\tilde{\xi}_2}{\tilde{\xi}_2'\hat{W}\tilde{\xi}_2}\right)\left(\frac{\tilde{\xi}_{1,i}}{\tilde{\xi}_{2,i}} - \beta\right),
$$

where the last equality follows since $\sum_{i=1}^k w^*_i = \sum_{i=1}^k \frac{\tilde{\xi}_2'\hat{W}e_ie_i'\tilde{\xi}_2}{\tilde{\xi}_2'\hat{W}\tilde{\xi}_2}$ with probability 1. The result then follows from Slutsky's theorem. □

B.3 *Auxiliary lemmas*

For $p \ge 1$, $x \in \mathbb{R}^2$, Ω a 2×2 matrix, and $b \in \mathbb{R}$, let

$$
R(x, \Omega, b, p) = \int\int\big|\hat{\beta}_U\big(x + \Omega^{1/2}z, 2\Omega\big) - b\big|^p\phi(z_1)\phi(z_2)\,dz_1\,dz_2
$$

and let

$$\begin{aligned}&\tilde{R}(t, c_1, c_2, c_3, p)\\&\quad = \int\int \left|\hat{\tau}(t+z_2, 2)(c_1 + c_2 z_1) + \left[\hat{\tau}(t+z_2, 2)t - 1\right]c_3\right|^p \phi(z_1)\phi(z_2)\, dz_1\, dz_2.\end{aligned}$$

LEMMA B.1. *For R and $\tilde{R}$ defined above,*

$$R(x, \Omega, b, p) = \tilde{R}\left(\frac{x_2}{\sqrt{\Omega_{22}}}, \frac{x_1 - bx_2}{\sqrt{\Omega_{22}}}, \left(\frac{\Omega_{11}}{\Omega_{22}} - \frac{\Omega_{12}^2}{\Omega_{22}^2}\right)^{1/2}, b - \frac{\Omega_{12}}{\Omega_{22}}, p\right).$$

PROOF. Without loss of generality, we can let $\Omega^{1/2}$ be the upper diagonal square root matrix

$$\Omega^{1/2} = \begin{pmatrix} \left(\Omega_{11} - \frac{\Omega_{12}^2}{\Omega_{22}}\right)^{1/2} & \frac{\Omega_{12}}{\sqrt{\Omega_{22}}} \\ 0 & \sqrt{\Omega_{22}} \end{pmatrix}.$$

Then

$$\begin{aligned}&\hat{\beta}_U\big(x + \Omega^{1/2}z, 2\Omega\big)\\&\quad = \hat{\tau}(x_2 + \sqrt{\Omega_{22}}z_2, 2\Omega_{22}) \cdot \left(x_1 + \left(\Omega_{11} - \frac{\Omega_{12}^2}{\Omega_{22}}\right)^{1/2} z_1\right.\\&\qquad \left. + \frac{\Omega_{12}}{\sqrt{\Omega_{22}}} z_2 - \frac{\Omega_{12}}{\Omega_{22}}(x_2 + \sqrt{\Omega_{22}}z_2)\right) + \frac{\Omega_{12}}{\Omega_{22}}\\&\quad = \frac{\hat{\tau}(x_2/\sqrt{\Omega_{22}} + z_2, 2)}{\sqrt{\Omega_{22}}} \cdot \left(x_1 + \left(\Omega_{11} - \frac{\Omega_{12}^2}{\Omega_{22}}\right)^{1/2} z_1 - \frac{\Omega_{12}}{\Omega_{22}} x_2\right) + \frac{\Omega_{12}}{\Omega_{22}}\end{aligned}$$

so that

$$\begin{aligned}&\hat{\beta}_U\big(x + \Omega^{1/2}z, 2\Omega\big) - b\\&\quad = \frac{\hat{\tau}(x_2/\sqrt{\Omega_{22}} + z_2, 2)}{\sqrt{\Omega_{22}}} \cdot \left(x_1 + \left(\Omega_{11} - \frac{\Omega_{12}^2}{\Omega_{22}}\right)^{1/2} z_1 - \frac{\Omega_{12}}{\Omega_{22}} x_2\right) + \frac{\Omega_{12}}{\Omega_{22}} - b\\&\quad = \frac{\hat{\tau}(x_2/\sqrt{\Omega_{22}} + z_2, 2)}{\sqrt{\Omega_{22}}} \cdot \left(x_1 - x_2 b + \left(\Omega_{11} - \frac{\Omega_{12}^2}{\Omega_{22}}\right)^{1/2} z_1 + \left(b - \frac{\Omega_{12}}{\Omega_{22}}\right) x_2\right) + \frac{\Omega_{12}}{\Omega_{22}} - b\\&\quad = \frac{\hat{\tau}(x_2/\sqrt{\Omega_{22}} + z_2, 2)}{\sqrt{\Omega_{22}}} \cdot \left(x_1 - x_2 b + \left(\Omega_{11} - \frac{\Omega_{12}^2}{\Omega_{22}}\right)^{1/2} z_1\right)\\&\qquad + \left(\frac{\hat{\tau}(x_2/\sqrt{\Omega_{22}} + z_2, 2)}{\sqrt{\Omega_{22}}} x_2 - 1\right)\left(b - \frac{\Omega_{12}}{\Omega_{22}}\right)\end{aligned}$$

and the result follows by plugging this in to the definition of R. □

We now give bounds on R and $\tilde{R}$. By the triangle inequality,

$$\begin{aligned}\tilde{R}(t, c_1, c_2, c_3, p)^{1/p} &\le \left(\int\int \hat{\tau}(t+z_2, 2)^p |c_1 + c_2 z_1|^p \phi(z_1)\phi(z_2)\, dz_1\, dz_2\right)^{1/p} \\ &\quad + c_3\left(\int\int \left|\hat{\tau}(t+z_2, 2)t - 1\right|^p \phi(z_1)\phi(z_2)\, dz_1\, dz_2\right)^{1/p} \qquad (17) \\ &= \left[C_1(t, p)\cdot C_2(c_1, c_2, p)\right]^{1/p} + c_3 C_3(t, p)^{1/p},\end{aligned}$$

where $C_1(t, p) = \int \hat{\tau}(t+z, 2)^p \phi(z)\, dz$, $C_2(c_1, c_2, p) = \int |c_1 + c_2 z|^p \phi(z)\, dz$, and $C_3(t, p) = \int |\hat{\tau}(t+z, 2)t - 1|^p \phi(z)\, dz$. Note that, by the triangle inequality, for $t > 0$,

$$\begin{aligned}C_1(t, p)^{1/p} &\le \left(\int \left|\hat{\tau}(t+z, 2) - 1/t\right|^p \phi(z)\, dz\right)^{1/p} + 1/t \\ &= (1/t)\left[C_3(t, p)^{1/p} + 1\right].\end{aligned} \qquad (18)$$

Similarly,

$$C_3(t, p)^{1/p} \le 1 + t\left(\int \hat{\tau}(t+z, 2)^p \phi(z)\, dz\right)^{1/p} = 1 + tC_1(t, p)^{1/p}. \qquad (19)$$

LEMMA B.2. *For $p < 2$, $C_1(t, p)$ is bounded uniformly over t on any compact set, and $\tilde{R}(t, c_1, c_2, c_3, p)$ is bounded uniformly over (t, c_1, c_2, c_3) in any compact set.*

PROOF. We have

$$\begin{aligned}C_1(t, p) &= \int \hat{\tau}(t+z, 2)^p \phi(z)\, dz = \int \left|\frac{1}{\sqrt{2}} \frac{1 - \Phi\big((t+z)/\sqrt{2}\big)}{\phi\big((t+z)/\sqrt{2}\big)}\right|^p \phi(z)\, dz \\ &\le 2^{-p/2} \int \frac{\phi(z)}{\phi\big((t+z)/\sqrt{2}\big)^p}\, dz \le K \int \exp\left(-\frac{1}{2}z^2 + \frac{p}{4}(t+z)^2\right) dz\end{aligned}$$

for a constant K that depends only on p. This is bounded uniformly over t in any compact set as long as $p/4 < 1/2$, giving the first result. Boundedness of $\tilde{R}$ follows from this, (19), and boundedness of $C_2(c_1, c_2, p)$ over c_1, c_2 in any compact set. □

LEMMA B.3. *For $p < 2$, $t\tilde{R}(t, c_1, c_2, c_3, p)^{1/p}$ is bounded uniformly over t, c_1, c_2, c_3 in any set such that t is bounded from below away from zero, and c_1, c_2 and c_3 are bounded.*

PROOF. By (17) and (18), it suffices to bound $tC_3(t, p)^{1/p} = t(\int |\hat{\tau}(t+z, 2)t - 1|^p \times \phi(z)\, dz)^{1/p}$. Let $\varepsilon > 0$ be a constant to be determined later. We split the integral into the regions $t + z < \varepsilon t$ and $t + z \ge \varepsilon t$. We have

$$\begin{aligned}&\int_{t+z<\varepsilon t} \left|\hat{\tau}(t+z, 2)t - 1\right|^p \phi(z)\, dz \\ &\quad = \int_{t+z<\varepsilon t} \left| t\frac{1 - \Phi\big((t+z)/\sqrt{2}\big)}{\sqrt{2}\phi\big((t+z)/\sqrt{2}\big)} - 1\right|^p \phi(z)\, dz\end{aligned} \qquad (20)$$

$$= \int_{t+z<\varepsilon t} \Big| t\big[1-\Phi\big((t+z)/\sqrt{2}\big)\big] - \sqrt{2}\phi\big((t+z)/\sqrt{2}\big)\Big|^p \frac{\phi(z)}{\big[\sqrt{2}\phi\big((t+z)/\sqrt{2}\big)\big]^p}\, dz.$$

This is bounded by a constant times

$$\begin{aligned}
&\max\{t,1\}\int_{t+z\le\varepsilon t}\exp\left(-\frac{1}{2}z^2+\frac{p}{4}(t+z)^2\right)dz\\
&=\max\{t,1\}\int_{t+z\le\varepsilon t}\exp\left(-\frac{1}{2}z^2+\frac{p}{4}\big(z^2+2tz+t^2\big)\right)dz\\
&=\max\{t,1\}\int_{t+z\le\varepsilon t}\exp\left(-\frac{1}{2}\big(z^2(1-p/2)-t^2(p/2)-ptz\big)\right)dz\\
&=\max\{t,1\}\int_{t+z\le\varepsilon t}\exp\left(-\frac{1-p/2}{2}\left(z^2-t^2\frac{p}{2-p}-2\frac{p}{2-p}tz\right)\right)dz\\
&=\max\{t,1\}\int_{t+z\le\varepsilon t}\exp\left(-\frac{1-p/2}{2}\left(\left(z-\frac{p}{2-p}t\right)^2-\left(\frac{p}{2-p}\right)^2t^2-\frac{p}{2-p}t^2\right)\right)dz\\
&=\max\{t,1\}\exp\left(\frac{1-p/2}{2}\left(\frac{p}{2-p}+\left(\frac{p}{2-p}\right)^2\right)t^2\right)\\
&\quad\times\int_{t+z\le\varepsilon t}\exp\left(-\frac{1-p/2}{2}\left(z-\frac{p}{2-p}t\right)^2\right)dz.
\end{aligned}$$

We have

$$\begin{aligned}
&\int_{t+z\le\varepsilon t}\exp\left(-\frac{1-p/2}{2}\left(z-\frac{p}{2-p}t\right)^2\right)dz\\
&\quad=\int_{z-tp/(2-p)\le(\varepsilon-1-p/(2-p))t}\exp\left(-\frac{1-p/2}{2}\left(z-\frac{p}{2-p}t\right)^2\right)dz\\
&\quad=\int_{u\le(\varepsilon-1-p/(2-p))t}\exp\left(-\frac{1-p/2}{2}u^2\right)dz,
\end{aligned}$$

which is bounded by a constant times

$$\Phi\big(t\big(\varepsilon-1-p/(2-p)\big)\sqrt{1-p/2}\big).$$

For $t(\varepsilon-1-p/(2-p))<0$, this is bounded by a constant times $\exp(-\frac{1-p/2}{2}t^2(1+p/(2-p)-\varepsilon)^2)$. Thus, (20) is bounded uniformly over $t>0$ by a constant times $\exp(-\eta t^2)$ for some $\eta>0$ as long as

$$\left(1+\frac{p}{2-p}-\varepsilon\right)^2>\frac{p}{2-p}+\left(\frac{p}{2-p}\right)^2=\frac{p}{2-p}\left(1+\frac{p}{2-p}\right),$$

which can be ensured by choosing $\varepsilon>0$ small enough as long as $p<2$. Thus, $\varepsilon>0$ can be chosen so that (20) is bounded uniformly over t when scaled by t^p.

For the integral over $t+z>\varepsilon t$, we have, by (1.1) in Baricz (2008),

$$\int_{t+z\geq\varepsilon t}\left|t\hat{\tau}(t+z,2)-1\right|^{p}\phi(z)\,dz=t^{p}\int_{t+z\geq\varepsilon t}\left|\hat{\tau}(t+z,2)-\frac{1}{t}\right|^{p}\phi(z)\,dz$$

$$\leq t^{p}\int_{t+z\geq\varepsilon t}\left|\frac{1}{t+z}-\frac{1}{t}\right|^{p}\phi(z)\,dz+t^{p}\int_{t+z\geq\varepsilon t}\left|\frac{1}{(t+z)+2/(t+z)}-\frac{1}{t}\right|^{p}\phi(z)\,dz.$$

The first term is

$$t^{p}\int_{t+z\geq\varepsilon t}\left|\frac{z}{(t+z)t}\right|^{p}\phi(z)\,dz\leq\frac{1}{t^{p}}\int\left|\frac{z}{\varepsilon}\right|^{p}\phi(z)\,dz.$$

The second term is

$$t^{p}\int_{t+z\geq\varepsilon t}\left|\frac{-z-2/(t+z)}{\left[(t+z)+2/(t+z)\right]t}\right|^{p}\phi(z)\,dz\leq\frac{1}{t^{p}}\int\left|\frac{|z|+|2/\varepsilon t|}{\varepsilon}\right|^{p}\phi(z)\,dz.$$

Both are bounded uniformly when scaled by t^{p} over any set with t bounded from below away from zero. □

Appendix C: Relation to Hirano and Porter (2015)

Hirano and Porter (2015) give a negative result establishing the impossibility of unbiased, quantile unbiased, or translation equivariant estimation in a wide variety of models with singularities, including many linear IV models. On initial inspection our derivation of an unbiased estimator for β may appear to contradict the results of Hirano and Porter. In fact, however, one of the key assumptions of Hirano and Porter (2015) no longer applies once we assume that the sign of the first stage is known.

Again consider the linear IV model with a single instrument, where for simplicity we let $\sigma_1^2=\sigma_2^2=1$, and $\sigma_{12}=0$. To discuss the results of Hirano and Porter (2015), it will be helpful to parameterize the model in terms of the reduced-form parameters $(\psi,\pi)=(\pi\beta,\pi)$. For ϕ again the standard normal density, the density of ξ is

$$f(\xi;\psi,\pi)=\phi(\xi_1-\psi)\phi(\xi_2-\pi).$$

Fix some value ψ^*. For any $\pi\neq 0$ we can define $\beta(\psi,\pi)=\frac{\psi}{\pi}$. If we consider any sequence $\{\pi_j\}_{j=1}^{\infty}$ approaching zero from the right, then $\beta(\psi^*,\pi_j)\to\infty$ if $\psi^*>0$ and $\beta(\psi^*,\pi_j)\to-\infty$ if $\psi^*<0$. Thus we can see that β plays the role of the function κ in Hirano and Porter (2015) equation (2.1).

Hirano and Porter (2015) show that if there exists some finite collection of parameter values $(\psi_{l,d},\pi_{l,d})$ in the parameter space and nonnegative constants $c_{l,d}$ such that their Assumption 2.4,

$$f\big(\xi;\psi^*,0\big)\leq\sum_{l=1}^{s}c_{l,d}f(\xi;\psi_{l,d},\pi_{l,d})\quad\forall\xi,$$

holds, then (since one can easily verify their Assumption 2.3 in the present context) there can exist no unbiased estimator of β.

This dominance condition fails in the linear IV model with a sign restriction. For any $(\psi_{l,d}, \pi_{l,d})$ in the parameter space, we have by definition that $\pi_{l,d} > 0$. For any such $\pi_{l,d}$, however, if we fix ξ_1 and take $\xi_2 \to -\infty$,

$$\begin{aligned}\lim_{\xi_2 \to -\infty} \frac{\phi(\xi_2 - \pi_{l,d})}{\phi(\xi_2)} &= \lim_{\xi_2 \to -\infty} \exp\left(-\frac{1}{2}(\xi_2 - \pi_{l,d})^2 + \frac{1}{2}\xi_2^2\right) \\ &= \lim_{\xi_2 \to -\infty} \exp\left(\xi_2 \pi_{l,d} - \frac{1}{2}\pi_{l,d}^2\right) = 0.\end{aligned}$$

Thus, $\lim_{\xi_2 \to -\infty} \frac{f(\xi; \psi_{l,d}, \pi_{l,d})}{f(\xi; \psi^*, 0)} = 0$, and for any fixed ξ_1, $\{c_{l,d}\}_{l=1}^{s}$, and $\{(\psi_{l,d}, \pi_{l,d})\}_{l=1}^{s}$ there exists a ξ_2^* such that $\xi_2 < \xi_2^*$ implies

$$f(\xi; \psi^*, 0) > \sum_{l=1}^{s} c_{l,d} f(\xi; \psi_{l,d}, \pi_{l,d}).$$

Thus, Assumption 2.4 in Hirano and Porter (2015) fails in this model, allowing the possibility of an unbiased estimator. Note, however, that if we did not impose $\pi > 0$, then we would satisfy Assumption 2.4, so unbiased estimation of β would again be impossible. Thus, the sign restriction on π plays a central role in the construction of the unbiased estimator $\hat{\beta}_U$.

Appendix D: Lower bound on risk of unbiased estimators

This appendix gives a lower bound on the attainable risk at a given π, β for an estimator that is unbiased for β for all π, β with π in the positive orthant. The bound is given by the risk in the submodel where $\pi/\|\pi\|$ (the direction of π) is known. While the bound cannot, in general, be obtained, we discuss some situations where it can, which include certain values of π in the case where ξ comes from a model with homoskedastic errors.

Theorem D.1. *Let $\mathcal{U}$ be the set of estimators for β that are unbiased for all $\pi \in (0, \infty)^k$, $\beta \in \mathbb{R}$. For any $\pi^* \in (0, \infty)^k$, $\beta^* \in \mathbb{R}$, and any convex loss function ℓ,*

$$E_{\pi^*, \beta^*} \ell\big(\hat{\beta}_U\big(\xi^*(\pi^*), \Sigma^*(\pi^*)\big) - \beta^*\big) \leq \inf_{\hat{\beta} \in \mathcal{U}} E_{\pi^*, \beta^*} \ell\big(\hat{\beta}(\xi, \Sigma) - \beta^*\big),$$

where $\xi^*(\pi^*) = [(I_2 \otimes \pi^*)' \Sigma^{-1} (I_2 \otimes \pi^*)]^{-1} (I_2 \otimes \pi^*)' \Sigma^{-1} \xi$ and $\Sigma^*(\pi^*) = [(I_2 \otimes \pi^*)' \times \Sigma^{-1} (I_2 \otimes \pi^*)]^{-1}$.

Proof. Consider the submodel with π restricted to $\Pi^* = \{\pi^* t | t \in (0, \infty)\}$. Then $\xi^*(\pi^*)$ is sufficient for (t, β) in this submodel and satisfies $\xi^*(\pi^*) \sim N((\beta t, t)', \Sigma^*(\pi^*))$ in this submodel. To see this, note that, for t, β in this submodel, ξ follows the generalized least squares regression model $\xi = (I_2 \otimes \pi^*)(\beta t, t)' + \varepsilon$, where $\varepsilon \sim N(0, \Sigma)$, and $\xi^*(\pi^*)$ is the generalized least squares estimator of $(\beta t, t)'$.

Let $\tilde{\beta}(\xi(\pi^*), \Sigma(\pi^*))$ be a (possibly randomized) estimator based on $\xi(\pi^*)$ that is unbiased in the submodel where $\pi \in \Pi^*$. By completeness of the submodel, $E[\tilde{\beta}(\xi(\pi^*), \Sigma(\pi^*))|\xi^*(\pi^*)] = \hat{\beta}_U(\xi(\pi^*), \Sigma(\pi^*))$. By Jensen's inequality, therefore,

$$E_{\pi^*,\beta^*}\ell\big(\tilde{\beta}\big(\xi\big(\pi^*\big), \Sigma\big(\pi^*\big)\big) - \beta\big) \geq E_{\pi^*,\beta^*}\ell\big(E\big[\tilde{\beta}\big(\xi\big(\pi^*\big), \Sigma\big(\pi^*\big)\big)|\xi^*(\beta)\big] - \beta\big)$$

(this is just the Rao–Blackwell theorem applied to the submodel with the loss function ℓ). By sufficiency, the set of risk functions for randomized unbiased estimators based on $\xi(\pi^*)$ in the submodel is the same as the set of risk functions for randomized unbiased estimators based on ξ in the submodel. This gives the result with $\mathcal{U}$ replaced by the set of estimators that are unbiased in the submodel, which implies the result as stated, since the set of estimator that are unbiased in the full model is a subset of those that are unbiased in the submodel. $\square$

Theorem D.1 continues to hold in the case where the lower bound is infinite: in this case, the risk of any unbiased estimator must be infinite at β^*, π^*. By Theorem 2.2, the lower bound is infinite for squared error loss $\ell(t) = t^2$ for any π^*, β^*. Thus, unbiased estimators must have infinite variance even in models with multiple instruments.

While in general Theorem D.1 gives only a lower bound on the risk of unbiased estimators, the bound can be achieved in certain situations. A case of particular interest arises in models with homoskedastic reduced-form errors that are independent across observations. In such cases $\mathrm{Var}((U', V')') = \mathrm{Var}((U_1, V_1)') \otimes I_T$, where I_T is the $T \times T$ identity matrix, so that the definition of Σ in (3) gives $\Sigma = \mathrm{Var}((U_1, V_1)') \otimes (Z'Z)^{-1}$. Thus, in models with independent homoskedastic errors, we have $\Sigma = Q_{UV} \otimes Q_Z$ for a 2×2 matrix Q_{UV} and a $k \times k$ matrix Q_Z.

Theorem D.2. *Suppose that* $[(I_2 \otimes \pi^*)'\Sigma^{-1}(I_2 \otimes \pi^*)]^{-1}(I_2 \otimes \pi^*)'\Sigma^{-1} = (I_2 \otimes a(\pi^*)')$ *for some* $a(\pi^*) \in \mathbb{R}^k$. *Then* $\hat{\beta}_U(\xi^*(\pi^*), \Sigma(\pi^*))$ *defined in Theorem D.1 is unbiased at any* π, β *such that* $a(\pi^*)'\pi > 0$. *In particular, if* $a(\pi^*) \in (0, \infty)^k$, *then* $\hat{\beta}_U(\xi^*(\pi^*), \Sigma(\pi^*)) \in \mathcal{U}$ *and the risk bound is attained. Specializing to the case where* $\Sigma = Q_{UV} \otimes Q_Z$ *for a* 2×2 *matrix* Q_{UV} *and a* $k \times k$ *matrix* Q_Z, *the above conditions hold with* $a(\pi^*)' = \pi^{*\prime}Q_Z^{-1}/(\pi^{*\prime}Q_Z^{-1}\pi^*)$, *and the bound is achieved if* $Q_Z^{-1}\pi^* \in (0, \infty)^k$.

Proof. For the first claim, note that under these assumptions $\xi^*(\pi^*) = (a(\pi^*)'\xi_1, a(\pi^*)'\xi_2)'$ is $N((a(\pi^*)'\pi\beta, a(\pi^*)'\pi)', \Sigma^*(\pi))$ distributed under π, β, so $\hat{\beta}_U(\xi^*(\pi^*), \Sigma(\pi^*))$ is unbiased at π, β by Theorem 2.1. For the case where $\Sigma = Q_{UV} \otimes Q_Z$, the result follows by properties of the Kronecker product:

$$\begin{aligned}&\big[\big(I_2 \otimes \pi^*\big)'(Q_{UV} \otimes Q_Z)^{-1}\big(I_2 \otimes \pi^*\big)\big]^{-1}\big(I_2 \otimes \pi^*\big)'(Q_{UV} \otimes Q_Z)^{-1}\\ &\quad = \big[Q_{UV}^{-1} \otimes \pi^{*\prime}Q_Z^{-1}\pi^*\big]^{-1}\big(Q_{UV}^{-1} \otimes \pi^{*\prime}Q_Z^{-1}\big) = I_2 \otimes \big[\pi^{*\prime}Q_Z^{-1}/\big(\pi^{*\prime}Q_Z^{-1}\pi^*\big)\big].\end{aligned}$$ $\square$

The special form of the sufficient statistic in the homoskedastic case derives from the form of the optimal estimator in the restricted seemingly unrelated regression (SUR) model. The submodel for the direction π^* is given by the SUR model

$$\begin{pmatrix} Y \\ X \end{pmatrix} = \begin{pmatrix} Z\pi^* & 0 \\ 0 & Z\pi^* \end{pmatrix} \begin{pmatrix} \beta t \\ t \end{pmatrix} + \begin{pmatrix} U \\ V \end{pmatrix}.$$

Considering this as a SUR model with regressors $Z\pi^*$ in both equations, the optimal estimator of $(\beta t, t)'$ simply stacks the OLS estimator for the two equations, since the regressors $Z\pi^*$ are the same and the parameter space for $(\beta t, t)$ is unrestricted. Note also that, in the homoskedastic case (with $Q_Z = (Z'Z)^{-1}$), $\xi_1^*(\pi^*)$ and $\xi_2^*(\pi^*)$ are proportional to $\pi^{*\prime} Z'Z\xi_1$ and $\pi^{*\prime} Z'Z\xi_2$, which are the numerator and denominator of the 2SLS estimator with ξ_2 replaced by π^* in the first part of the quadratic form.

Thus, for certain parameter values π^* in the homoskedastic case, the risk bound in Theorem D.1 is obtained. In such cases, the estimator that obtains the bound is unique and depends on π^* itself (for the absolute value loss function, which is not strictly concave, uniqueness is shown in Appendix D.1 below). Thus, in contrast to settings such as linear regression, where a single estimator minimizes the risk over unbiased estimators simultaneously for all parameter values, no uniform minimum risk unbiased estimator will exist. The reason for this is clear: knowledge of the direction of $\pi = \pi^*$ helps with estimation of β, even if one imposes unbiasedness for all π.

It is interesting to note precisely how the parameter space over which the estimator in the risk bound is unbiased depends on π^*. Suppose one wants an estimator that minimizes the risk at π^* while still remaining unbiased in a small neighborhood of π^*. In the homoskedastic case, this can always be done as long as $\pi^* \in (0,\infty)^k$, since $\pi^{*\prime} Q_Z^{-1}\pi > 0$ for π close enough to π^*. Where one can expand this neighborhood while maintaining unbiasedness will depend on π^* and Q_Z. In the case where $\pi^{*\prime} Q_Z^{-1}$ is in the positive orthant, the assumption $\pi \in (0,\infty)^k$ is enough to ensure that this estimator is unbiased at π. However, if $\pi^{*\prime} Q_Z^{-1}$ is not in the positive orthant, there is a trade-off between precision at π^* and the range of $\pi \in (0,\infty)^k$ over which unbiasedness can be maintained.

Put another way, in the homoskedastic case, for any $\pi^* \in \mathbb{R}^k \backslash \{0\}$, minimizing the risk of an estimator of β subject to the restriction that the estimator is unbiased in a neighborhood of π^* leads to an estimator that does not depend on this neighborhood, as long as the neighborhood is small enough (this is true even if the restriction $\pi^* \in (0,\infty)^k$ does not hold). The resulting estimator depends on π^*, and is unbiased at π if and only if $\pi^* Q_Z^{-1}\pi > 0$.

D.1 *Uniqueness of the minimum risk unbiased estimator under absolute value loss*

In the discussion above, we used the result that the minimum risk unbiased estimator in the submodel with $\pi/\|\pi\|$ known is unique for absolute value loss. Because the absolute value loss function is not strictly concave, this result does not, to our knowledge, follow immediately from results in the literature. We therefore provide a statement and proof here. In the following theorem, we consider a general setup where a random variable ξ is observed, which follows a distribution P_μ for some $\mu \in M$. The family of distributions $\{P_\mu | \mu \in M\}$ need not be a multivariate normal family, as in the rest of this paper.

Theorem D.3. *Let $\hat{\theta} = \hat{\theta}(\xi)$ be an unbiased estimator of $\theta = \theta(\mu)$, where $\mu \in M$ for some parameter space M and $\Theta = \{\theta | \theta(\mu) = \theta \text{ some } \mu \in M\} \subseteq \mathbb{R}$, and where ξ has the same support for all $\mu \in M$. Let $\tilde{\theta}(\xi, U)$ be another unbiased estimator, based on (ξ, U), where ξ and U are independent, and let $\hat{\theta}(\xi) = E_\mu[\tilde{\theta}(\xi, U)|\xi] = \int \tilde{\theta}(\xi, U)\, dQ(U)$, where Q denotes the probability measure of U, which is assumed not to depend on μ. Suppose that $\hat{\theta}(\xi)$ and $\tilde{\theta}(\xi, U)$ have the same risk under absolute value loss*:

$$E_\mu\big|\tilde{\theta}(\xi, U) - \theta(\mu)\big| = E_\mu\big|\hat{\theta}(\xi) - \theta(\mu)\big| \quad \textit{for all } \mu \in M.$$

Then $\tilde{\theta}(\xi, U) = \hat{\theta}(\xi)$ for almost every ξ with $\hat{\theta}(\xi) \in \Theta$ and almost every U.

Proof. The display can be written as

$$E_\mu\big\{E_\mu\big[\big|\tilde{\theta}(\xi, U) - \theta(\mu)\big| |\xi\big] - \big|\hat{\theta}(\xi) - \theta(\mu)\big|\big\} = 0 \quad \text{for all } \mu \in M.$$

By Jensen's inequality, the term inside the outer expectation is nonnegative for μ-almost every ξ. Thus, the equality implies that this term is zero for μ-almost every ξ (since $EX = 0$ implies $X = 0$ a.e. for any nonnegative random variable X). This gives, noting that $\int |\tilde{\theta}(\xi, U) - \theta(\mu)|\, dQ(U) = E_\mu[|\tilde{\theta}(\xi, U) - \theta(\mu)||\xi]$,

$$\int \big|\tilde{\theta}(\xi, U) - \theta(\mu)\big|\, dQ(U) = \big|\hat{\theta}(\xi) - \theta(\mu)\big| \quad \text{for } \mu\text{-almost every } \xi \text{ and all } \mu \in M.$$

Since the support of ξ is the same under all $\mu \in M$, the above statement gives

$$\int \big|\tilde{\theta}(\xi, U) - \theta\big|\, dQ(U) = \big|\hat{\theta}(\xi) - \theta\big| \quad \text{for almost every } \xi \text{ and all } \theta \in \Theta.$$

Note that, for any random variable X, $E|X| = |EX|$ implies that either $X \geq 0$ a.e. or $X \leq 0$ a.e. Applying this to the above display, it follows that for all $\theta \in \Theta$ and almost every ξ, either $\tilde{\theta}(\xi, U) \leq \theta$ a.e. U or $\tilde{\theta}(\xi, U) \geq \theta$ a.e. U. In particular, whenever $\hat{\theta}(\xi) \in \Theta$, either $\tilde{\theta}(\xi, U) \leq \hat{\theta}(\xi)$ a.e. U or $\tilde{\theta}(\xi, U) \geq \hat{\theta}(\xi)$ a.e. U. In either case, the condition $\int \tilde{\theta}(\xi, U)\, dQ(U) = \hat{\theta}(\xi)$ implies that $\tilde{\theta}(\xi, U) = \hat{\theta}(\xi)$ a.e. U. It follows that, for almost every ξ such that $\hat{\theta}(\xi) \in \Theta$, we have $\tilde{\theta}(\xi, U) = \theta(\xi)$ a.e. U, as claimed. □

Thus, if $\hat{\theta}(\xi) \in \Theta$ with probability 1, we will have $\tilde{\theta}(\xi, U) = \hat{\theta}(\xi)$ a.e. (ξ, U). However, if $\hat{\theta}(\xi)$ can take values outside Θ, this will not necessarily be the case. For example, in the single-instrument case of our setup, if we restrict our parameter space to $(\pi, \beta) \in (0, \infty) \times [c, \infty)$ for some constant c, then forming a new estimator by adding or subtracting 1 from $\hat{\beta}_U$ with equal probability independently of ξ whenever $\hat{\beta}_U \leq c - 1$ gives an unbiased estimator with identical absolute value risk.

In our case, letting $\xi(\pi^*)$ be as in Theorem D.1, the support of $\xi(\pi^*)$ is the same under $\pi^* t$, β for any $t \in (0, \infty)$ and $\beta \in \mathbb{R}$. If $\tilde{\beta}(\xi(\pi^*), U)$ is unbiased in this restricted parameter space, we must have, letting $\hat{\beta}_U(\xi^*(\pi), \Sigma^*(\pi))$ be the unbiased nonrandomized estimator in the submodel, $E[\tilde{\beta}(\xi(\pi^*), U)|\xi(\pi^*)] = \hat{\beta}_U(\xi(\pi^*), \Sigma^*(\pi))$ by completeness for any random variable U with a distribution that does not depend on (t, β). Since $\hat{\beta}_U(\xi(\pi^*), \Sigma^*(\pi)) \in \mathbb{R}$ with probability 1, it follows that if $\tilde{\beta}(\xi(\pi^*), U)$ has the same risk as $\hat{\beta}_U(\xi(\pi^*), \Sigma^*(\pi))$, then $\tilde{\beta}(\xi(\pi^*), U) = \hat{\beta}_U(\xi(\pi^*), \Sigma^*(\pi))$ with probability 1, as long as we impose that $\tilde{\beta}(\xi(\pi^*), U)$ is unbiased for all $t \in (1 - \varepsilon, 1 + \varepsilon)$ and $\beta \in \mathbb{R}$.

Appendix E: Reduction of the parameter space by equivariance

In the appendix, we discuss how we can reduce the dimension of the parameter space using an equivariance argument. We first consider the just-identified case and then note how such arguments may be extended to the overidentified case under the additional assumption of homoskedasticity.

E.1 *Just-identified model*

For comparisons between $(\hat{\beta}_U, \hat{\beta}_{2SLS}, \hat{\beta}_{FULL})$ in the just-identified case, it suffices to consider a two-dimensional parameter space. To see that this is the case, let $\theta = (\beta, \pi, \sigma_1^2, \sigma_{12}, \sigma_2^2)$ be the vector of model parameters and let g_A, for $A = \left[\begin{smallmatrix} a_1 & a_2 \\ 0 & a_3 \end{smallmatrix}\right]$, $a_1 \neq 0$, $a_3 > 0$, be the transformation

$$g_A\xi = \tilde{\xi} = A\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} a_1\xi_1 + a_2\xi_2 \\ a_3\xi_2 \end{pmatrix},$$

which leads to $\tilde{\xi}$ being distributed according to the parameters

$$\tilde{\theta} = (\tilde{\beta}, \tilde{\pi}, \tilde{\sigma}_1^2, \tilde{\sigma}_{12}, \tilde{\sigma}_2^2),$$

where

$$\begin{aligned} \tilde{\beta} &= \frac{(a_1\beta + a_2)}{a_3}, \\ \tilde{\pi} &= a_3\pi, \\ \tilde{\sigma}_1^2 &= a_1^2\sigma_1^2 + a_1a_2\sigma_{12} + a_2^2\sigma_2^2, \\ \tilde{\sigma}_{12} &= a_1a_3\sigma_{12} + a_2a_3\sigma_2^2, \end{aligned}$$

and

$$\tilde{\sigma}_2^2 = a_3^2\sigma_2^2.$$

Define $\mathcal{G}$ as the set of all transformations g_A of the form above. Note that the sign restriction on π is preserved under $g_A \in \mathcal{G}$, and that for each g_A, there exists another transformation $g_A^{-1} \in \mathcal{G}$ such that $g_A g_A^{-1}$ is the identity transformation. We can see that the model (2) is invariant under the transformation g_A. Note further that the estimators $\hat{\beta}_U$, $\hat{\beta}_{2SLS}$, and $\hat{\beta}_{FULL}$ are all equivariant under g_A, in the sense that

$$\hat{\beta}(g_A\xi) = \frac{a_1\hat{\beta}(\xi) + a_2}{a_3}.$$

Thus, for any properties of these estimators (e.g., relative mean and median bias, relative dispersion) that are preserved under the transformations g_A, it suffices to study these properties on the reduced parameter space obtained by equivariance. By choosing A appropriately, we can always obtain

$$\begin{pmatrix} \tilde{\xi}_1 \\ \tilde{\xi}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \tilde{\pi} \end{pmatrix}, \begin{pmatrix} 1 & \tilde{\sigma}_{12} \\ \sigma_{12} & 1 \end{pmatrix}\right)$$

for $\tilde{\pi} > 0$, $\sigma_{12} \geq 0$, and thus reduce to a two-dimensional parameter (π, σ_{12}) with $\sigma_{12} \in [0, 1)$, $\pi > 0$.

E.2 *Overidentified model under homoskedasticity*

As noted in Appendix D, under the assumption of i.i.d. homoskedastic errors, Σ is of the form $\Sigma = Q_{UV} \otimes Q_Z$ for matrix $Q_{UV} = \mathrm{Var}((U_1, V_1)')$ and $Q_Z = (Z'Z)^{-1}$. If we let $\sigma_U^2 = \mathrm{Var}(U_1)$, $\sigma_V^2 = \mathrm{Var}(V_1)$, and $\sigma_{UV} = \mathrm{Cov}(U_1, V_1)$, then by using an equivariance argument as above we can eliminate the parameters σ_U^2, σ_V^2, and β for the purposes of comparing $\hat{\beta}_{2SLS}$, $\hat{\beta}_{FULL}$, and the unbiased estimators. In particular, define $\theta = (\beta, \pi, \sigma_U^2, \sigma_{UV}, \sigma_V^2, Q_Z)$ and again let $A = \left[\begin{smallmatrix} a_1 & a_2 \\ 0 & a_3 \end{smallmatrix}\right]$, $a_1 \neq 0$, $a_3 > 0$, and consider the transformation

$$g_A \xi = \tilde{\xi} = (A \otimes I_k) \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} a_1 \xi_1 + a_2 \xi_2 \\ a_3 \xi_2 \end{pmatrix},$$

which leads to $\tilde{\xi}$ being distributed according to the parameters

$$\tilde{\theta} = \left(\tilde{\beta}, \tilde{\pi}, \tilde{\sigma}_U^2, \tilde{\sigma}_{UV}, \tilde{\sigma}_V^2, \tilde{Q}_Z\right),$$

where

$$\begin{aligned} \tilde{\beta} &= \frac{(a_1 \beta + a_2)}{a_3}, \\ \tilde{\pi} &= a_3 \pi, \\ \tilde{\sigma}_U^2 &= a_1^2 \sigma_U^2 + a_1 a_2 \sigma_{UV} + a_2^2 \sigma_V^2, \\ \tilde{\sigma}_{UV} &= a_1 a_3 \sigma_{UV} + a_2 a_3 \sigma_V^2, \\ \tilde{\sigma}_V^2 &= a_3^2 \sigma_V^2, \end{aligned}$$

and

$$\tilde{Q}_Z = Q_Z.$$

Note that this transformation changes neither the direction of the first stage, $\pi / \|\pi\|$, nor Q_Z. If we again define $\mathcal{G}$ to be the class of such transformations, we again see that the model is invariant under transformations $g_A \in \mathcal{G}$, and that the estimators for β we consider are equivariant under these transformations. Thus, since relative bias and MAD across estimators are preserved under these transformations, we can again study these properties on the reduced parameter space obtained by equivariance. In particular, by choosing A appropriately we can set $\tilde{\sigma}_U^2 = \tilde{\sigma}_V^2 = 1$ and $\tilde{\beta} = 0$, so the remaining free parameters are $\tilde{\pi}$, $\tilde{\sigma}_{UV}$, and $\tilde{Q}_Z$.

Appendix F: Additional simulation results in just-identified case

This appendix gives further results for our simulations in the just-identified case. We first report median bias comparisons for the estimators $\hat{\beta}_U$, $\hat{\beta}_{2SLS}$, and $\hat{\beta}_{FULL}$, and then report further dispersion and absolute deviation simulation results to complement those in Section 4.1.2 of the paper.

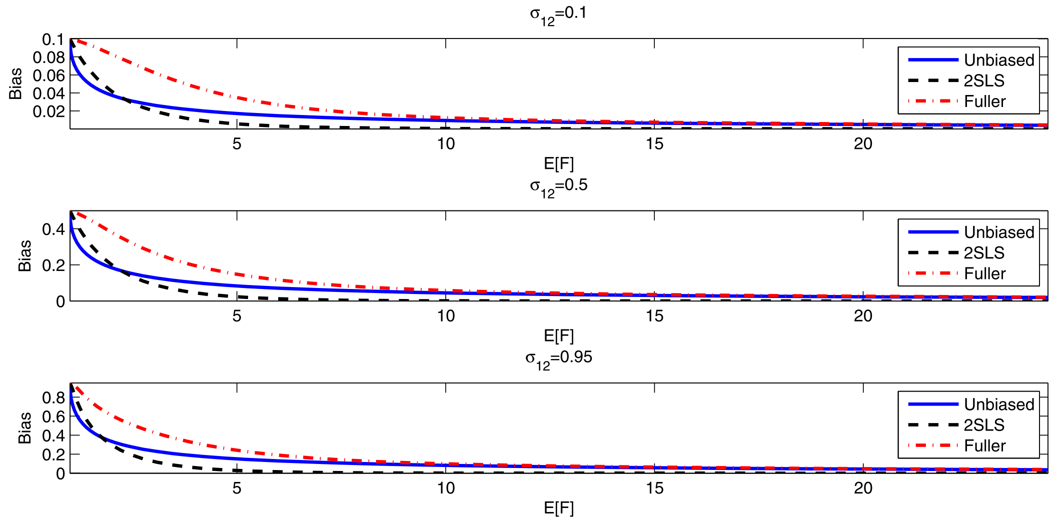

FIGURE 3. Median bias of single-instrument estimators, plotted against mean $E[F]$ of the first-stage F-statistic, based on 10 million simulations.

F.1 *Median bias*

Figure 3 plots the median bias of the single-instrument IV estimators against the mean of the first-stage F-statistic. In all calibrations considered, the unbiased estimator has a smaller median bias than 2SLS when the first stage is very small and a larger median bias for larger values of the first stage. By contrast the median bias of Fuller is larger than that of both the unbiased and the 2SLS estimators, though its median bias is quite close to that of the unbiased estimator once the mean of the first-stage F-statistic exceeds 10.

F.2 *Dispersion simulation results*

The lack of moments for $\hat{\beta}_{2\text{SLS}}$ complicates comparisons of dispersion, since we cannot consider mean squared error or mean absolute deviation, and also cannot recenter $\hat{\beta}_{2\text{SLS}}$ around its mean. As an alternative, we instead consider the full distribution of the absolute deviation of each estimator from its median. In particular, for the estimators $(\hat{\beta}_U, \hat{\beta}_{2\text{SLS}}, \hat{\beta}_{\text{FULL}})$ we calculate the zero-median residuals

$$(\varepsilon_U, \varepsilon_{2\text{SLS}}, \varepsilon_{\text{FULL}}) = \big(\hat{\beta}_U - \text{med}(\hat{\beta}_U), \hat{\beta}_{2\text{SLS}} - \text{med}(\hat{\beta}_{2\text{SLS}}), \hat{\beta}_{\text{FULL}} - \text{med}(\hat{\beta}_{\text{FULL}})\big).$$

Our simulation results suggest a strong stochastic ordering between these residuals (in absolute value). In particular we find that $|\varepsilon_{2\text{SLS}}|$ approximately dominates $|\varepsilon_U|$, which in turn approximately dominates $|\varepsilon_{\text{FULL}}|$, both in the sense of first-order stochastic dominance. This numerical result is consistent with analytical results on the tail behavior of the estimators. In particular, $\hat{\beta}_{2\text{SLS}}$ has no moments, reflecting thick tails in its sampling distribution, while $\hat{\beta}_{\text{FULL}}$ has all moments, reflecting thin tails. As noted in Section 2.3, the unbiased estimator $\hat{\beta}_U$ has a finite first moment, but an infinite $1+\varepsilon$ absolute moment for all $\varepsilon > 0$, and so falls between these two extremes.

To check for stochastic dominance in the distribution of $(|\varepsilon_U|, |\varepsilon_{2SLS}|, |\varepsilon_{FULL}|)$, we simulated 10^6 draws of $\hat{\beta}_U$, $\hat{\beta}_{2SLS}$, and $\hat{\beta}_{FULL}$ on a grid formed by the Cartesian product of $\sigma_{12} \in \{0, (0.005)^{\frac{1}{2}}, (0.01)^{\frac{1}{2}}, \dots, (0.995)^{\frac{1}{2}}\}$ and $\pi \in \{(0.01)^2, (0.02)^2, \dots, 25\}$. We use these grids for σ_{12} and π, rather than a uniformly spaced grid, because preliminary simulations suggested that the behavior of the estimators was particularly sensitive to the parameters for large values of σ_{12} and small values of π.

At each point in the grid we calculate $(\varepsilon_U, \varepsilon_{2SLS}, \varepsilon_{FULL})$, using independent draws to calculate ε_U and the other two estimators, and compute a one-sided Kolmogorov–Smirnov statistic for the hypotheses that (i) $|\varepsilon_{IV}| \geq |\varepsilon_U|$ and (ii) $|\varepsilon_U| \geq |\varepsilon_{FULL}|$, where $A \geq B$ for random variables A and B denotes that A is larger than B in the sense of first-order stochastic dominance. In both cases the maximal value of the Kolmogorov–Smirnov statistic is less than 2×10^{-3}. Conventional Kolmogorov–Smirnov p-values are not valid in the present context (since we use estimated medians to construct ε), but are never below 0.25.

F.3 *Containment of $\hat{\beta}_U$ in Anderson–Rubin confidence set*

As noted in Section 2.4, the Anderson–Rubin test is uniformly most powerful unbiased in the just-identified model. One can show, however, that the unbiased estimator $\hat{\beta}_U$ is not always contained in the Anderson–Rubin confidence set (that is, the confidence set formed by collecting the set of all parameter values not rejected by the Anderson–Rubin test). Specifically, consider the case where ξ_2 is large and negative, ξ_1 is large and positive, and σ_{12} is nonnegative. In this case, the Anderson–Rubin confidence set will consist solely of negative values, while $\hat{\beta}_U$ will be large and positive, and so will necessarily lie outside the Anderson–Rubin confidence set.

While this sort of scenario can easily arise if our sign constraint is violated, it occurs with only low probability when the sign constraint is satisfied. In particular, as in Appendix F.2 we consider a fine grid of values in the parameter space and simulate the frequency with which the unbiased estimator is contained in the Anderson–Rubin confidence set at each point (based on 100,000 simulations). We find that the probability that the 95% Anderson–Rubin confidence set contains the unbiased estimator $\hat{\beta}_U$ is always at least 97%, and exceeds 99.8% when the mean of the first-stage F-statistic is greater than 2. Likewise, the probability that the 90% Anderson–Rubin confidence set contains $\hat{\beta}_U$ is always at least 94.5%, and exceeds 99.3% when the mean of the first-stage F-statistic is greater than 2.

Appendix G: Multi-instrument simulation design

This appendix gives further details for the multi-instrument simulation design used in Section 4.2. We base our simulations on the Staiger and Stock (1997) specifications for the Angrist and Krueger (1991) data. The instruments in all specifications are quarter of birth and quarter of birth interacted with other dummy variables, and in all cases the dummy for the fourth quarter (and the corresponding interactions) are excluded to avoid multicollinearity. The rationale for the quarter of birth instrument in Angrist and Krueger (1991) indicates that the first-stage coefficients on the instruments should therefore be negative.

We first calculate the OLS estimates $\hat{\pi}$. All estimated coefficients satisfy the sign restriction in specification I, but some of them violate it in specifications II, III, and IV. To enforce the sign restriction, we calculate the posterior mean for π conditional on the OLS estimates, assuming a flat prior on the negative orthant and an exact normal distribution for the OLS estimates with variance equal to the estimated variance. This yields an estimate

$$\tilde{\pi}_i = \hat{\pi}_i - \hat{\sigma}_i \phi\left(\frac{\hat{\pi}_i}{\hat{\sigma}_i}\right) \Big/ \left(1 - \Phi\left(\frac{\hat{\pi}_i}{\hat{\sigma}_i}\right)\right)$$

for the first-stage coefficient on instrument i, where $\hat{\pi}_i$ is the OLS estimate and $\hat{\sigma}_i$ is its standard error. When $\hat{\pi}_i$ is highly negative relative to $\hat{\sigma}_i$, $\tilde{\pi}_i$ will be close to $\hat{\pi}_i$, but otherwise $\tilde{\pi}_i$ ensures that our first-stage estimates all obey the sign constraint. We then conduct the simulations using $\tilde{\pi}^* = -\tilde{\pi}$ to cast the sign constraint in the form considered in Section 1.2.

Our simulations fix $\tilde{\pi}^*/\|\tilde{\pi}^*\|$ at its estimated value and fix $Z'Z$ at its value in the data. By the equivariance argument in Appendix E, we can fix $\sigma_U^2 = \sigma_V^2 = 1$ and $\beta = 0$ in our simulations, so the only remaining free parameters are $\|\pi\|$ and σ_{UV}. We consider $\sigma_{UV} \in \{0.1, 0.5, 0.95\}$ and consider a grid of nine values for $\|\pi\|$ such that the mean of the first-stage F-statistic varies between 2 and 11.2. For each pair of these parameters we set

$$\Sigma = \begin{bmatrix} 1 & \sigma_{UV} \\ \sigma_{UV} & 1 \end{bmatrix} \otimes \left(Z'Z\right)^{-1}$$

and draw

$$\xi \sim N\left(\begin{array}{c} 0 \\ \|\pi\| \cdot \dfrac{\tilde{\pi}^*}{\|\tilde{\pi}^*\|} \end{array}, \Sigma\right).$$

References

Angrist, J. D. and A. B. Krueger (1991), "Does compulsory school attendance affect schooling and earnings?" *The Quarterly Journal of Economics*, 106 (4), 979–1014. [23, 24]

Baricz, Á. (2008), "'Mills' ratio: Monotonicity patterns and functional inequalities'." *Journal of Mathematical Analysis and Applications*, 340 (2), 1362–1370. [4, 15]

Hirano, K. and J. Porter (2015), "Location properties of point estimators in linear instrumental variables and related models." *Econometric Reviews*, 34 (6-10), 720–733. [1, 15, 16]

Lehmann, E. L. and J. P. Romano (2005), *Testing Statistical Hypotheses*. Springer. [2]

Small, C. (2010), *Expansions and Asymptotics for Statistics*. Chapman and Hall. [3, 11]

Staiger, D. and J. H. Stock (1997), “Instrumental variables regression with weak instruments.” *Econometrica*, 65 (3), 557–586. [8, 23]

Voinov, V. G. and M. S. Nikulin (1993), *Unbiased Estimators and Their Applications*, Vol. 1. Mathematics and its Applications, Vol. 263. Kluwer Academic Publishers, Dordrecht. Univariate case, Translated from the 1989 Russian original by L. E. Strautman and revised by the authors. [1]

Co-editor Frank Schorfheide handled this manuscript.

Manuscript received 10 April, 2016; final version accepted 9 December, 2016; available online 27 December, 2016.