A Robust Test for Weak Instruments: Supplementary Materials.

José Luis Montiel Olea
Harvard University
Cambridge, MA 02138
montiel@fas.harvard.edu

Carolin Pflueger
University of British Columbia
Vancouver, BC V6T 1Z2, Canada
carolin.pflueger@sauder.ubc.ca.

B. Computational Supplementary Materials

B.1 Numerical Accuracy of Patnaik Methodology

We now compare two slightly different Patnaik (1949) methodologies the exact probabilities for the cases considered in Imhof (1961). While the original Patnaik approximation uses a central χ^2 distribution the modified methodology used in this paper uses a noncentral χ^2. The advantage of using noncentral χ^2 distribution is that the approximation is exact in the conditionally homoskedastic case.

Table B.1 shows that both the central and the noncentral Patnaik methodologies are highly accurate, especially in the tails of the distributions considered. When the exact probability is $\leq 15\%$, the absolute error for both methodologies is at most 0.70% for all quadratic forms considered.

B.2 Monte-Carlo Method

We first replace $\mathbf{W}_2$ by the diagonal matrix of its eigenvalues and normalize the trace of $\mathbf{W}_2$ to 1, which leaves the critical value $c(\alpha, \mathbf{W}_2, x)$ unchanged. Our simulation routine takes as inputs the maximal asymptotic size of the test α, the eigenvalues of the matrix $\mathbf{W}_2$, the

Table B.1 Numerical Accuracy of Patnaik Methodology

Quadratic Form	x	$P(Q > x)$ Exact	Central	Noncentral	Absolute Error Central	Noncentral
$Q_1 = 0.6\chi^2_1 + 0.3\chi^2_1 + 0.1\chi^2_1$	0.1	94.58	91.85	91.85	2.73	2.73
	0.7	50.64	50.79	50.79	0.15	0.15
	2	12.40	13.10	13.10	0.70	0.70
$Q_2 = 0.6\chi^2_2 + 0.3\chi^2_2 + 0.1\chi^2_2$	0.2	99.36	98.68	98.68	0.68	0.68
	2	39.98	40.98	40.98	1.00	1.00
	6	1.61	1.45	1.45	0.16	0.16
$Q_3 = 0.6\chi^2_6 + 0.3\chi^2_4 + 0.1\chi^2_2$	1	99.73	99.61	99.61	0.12	0.12
	5	41.96	44.00	44.00	2.04	2.04
	12	0.87	0.80	0.80	0.07	0.07
$Q_4 = 0.6\chi^2_2 + 0.3\chi^2_4 + 0.1\chi^2_6$	1	96.66	95.22	95.22	1.44	1.44
	3	41.96	43.30	43.30	1.34	1.34
	8	0.87	0.66	0.66	0.21	0.21
$Q_5 = 0.7\chi^2_{6;6} + 0.3\chi^2_{2;2}$	2	99.39	99.54	99.29	0.15	0.10
	10	40.87	40.46	41.09	0.41	0.22
	20	2.21	2.30	2.16	0.09	0.05
$Q_6 = 0.7\chi^2_{1;6} + 0.3\chi^2_{1;2}$	1	95.49	97.19	94.96	1.70	0.53
	6	40.76	39.48	41.12	1.28	0.36
	15	2.23	2.46	2.16	0.23	0.07
$\frac{1}{3}Q_3 + \frac{2}{3}Q_4$	1.5	98.91	98.42	98.42	0.49	0.49
	4	34.53	35.52	35.52	0.99	0.99
	7	1.54	1.31	1.31	0.23	0.23
$\frac{1}{2}Q_5 + \frac{1}{2}Q_6$	3.5	95.63	96.05	95.47	0.42	0.16
	8	41.52	41.01	41.71	0.51	0.19
	13	4.62	4.74	4.58	0.12	0.04
$\frac{1}{4}(Q_3 + Q_4 + Q_5 + Q_6)$	3	98.42	98.37	98.22	0.05	0.20
	6	42.64	42.70	42.99	0.06	0.35
	10	1.17	1.16	1.09	0.01	0.08

NOTE: $P(Q > x)$ in percent, where $Q = \sum_{r=1}^{m} q_r \chi^2_{h_r;\delta^2_r}$ is a positive semidefinite quadratic form in independent normal random variables. The $\chi^2_{h_r;\delta^2_r}$ are independent χ^2 random variables with h_r degrees of freedom and non-centrality parameter δ^2_r. The quadratic forms, thresholds x and the exact probabilities are as in Imhof (1961). We show probabilities for the original central chi-squared Patnaik approximation and for the noncentral chi-square Patnaik approximation. The noncentral Patnaik approximation is used throughout the paper.

threshold x and computes a Monte Carlo critical values $c^m(\alpha, \mathbf{W}_2, x)$.

Draw N independent multivariate normal random variables $\mathbf{z}_v \sim N(\mathbf{0}, \mathbb{I}_K)$. For a given $\mathbf{C}$ and $\mathbf{W}_2$ we use these normal draws to compute N draws from the distribution $\boldsymbol{\gamma}_2'\boldsymbol{\gamma}_2/tr\mathbf{W}_2$, where we set the default to $N = 40,000$. We then compute $F^{m-1}_{\mathbf{C},\mathbf{W}_2}(\alpha)$ as the sample upper α-point from these N draws.

$c(\alpha, \mathbf{W}_2, x)$ is defined as the supremum of $F^{-1}_{\mathbf{C},\mathbf{W}_2}(\alpha)$ over the set $\Lambda = \left\{\mathbf{C} \mid \frac{\mathbf{C}'\mathbf{C}}{tr(\mathbf{W}_2)} \leq x\right\}$. We construct a finite Monte Carlo analogue Λ^m with $10 \times L$ elements with a default value of $L = 50$.

We draw $\boldsymbol{\lambda}_i, i = 1, 2, ..., L$ iid from a multivariate uniform distribution on $[0,1]^K$. Then replace

$$\begin{aligned} \boldsymbol{\lambda}_1 &= [0, .., 0, 1] \qquad (1) \\ \boldsymbol{\lambda}_2 &= [1, 1, ..., 1] \text{ if } K \geq 2 \\ \boldsymbol{\lambda}_3 &= [0, ..., 0, 1, 1] \text{ if } K \geq 3 \\ \boldsymbol{\lambda}_4 &= [0, ..., 0, 1, 1, 1] \text{ if } K \geq 4 \end{aligned}$$

We then use $\boldsymbol{\Lambda}^m = \left\{\mathbf{W}_2^{1/2}\boldsymbol{\lambda}_i \times \sqrt{t}/\sqrt{\boldsymbol{\lambda}_i'\mathbf{W}_2\boldsymbol{\lambda}_i}, i = 1, 2, ..., L; t = x - 9, ...x - 1, x\right\}$.

The Monte-Carlo critical value is then given by

$$c^m(\alpha, \mathbf{W}_2, x) = \max\left\{F^{m-1}_{\mathbf{C},\mathbf{W}_2}(\alpha) \mid \mathbf{C} \in \boldsymbol{\Lambda}^m\right\}.$$

B.3. Levels and Sizes of Patnaik and Monte Carlo Critical Values

We compute Monte Carlo critical values $c^m(5\%, \mathbf{W}_2, 10)$ and Patnaik critical values $c^P(5\%, \mathbf{W}_2, 10)$ for 400 randomly drawn matrices $\mathbf{W}_2$ of size $K = 1, 2, 3, 4, 5$ and compare the size distortions of c^P and c^m.

We can assume wlog that $\mathbf{W}_2$ is diagonal. For each of $K = 1, 2, 3, 4$ we draw 100 vectors of eigenvalues $eig(\mathbf{W}_2)$ iid from a uniform distribution on $[0, 1]$. We then replace the first vector of eigenvalues by $[1, 0, ..., 0]$, the second one by $[1, 1]$, and the third and fourth ones by $[1, 1, 0, ..., 0]$ and $[1, 1, 1, 0]$ provided K is large enough. We normalize the trace of $\mathbf{W}_2$ to equal one. We denote the resulting set of diagonal matrices by $\mathcal{W}_2$.

We obtain $c^P(5\%, \mathbf{W}_2, 10)$ and $c^m(5\%, \mathbf{W}_2, 10)$ for every $\mathbf{W}_2$ in $\mathcal{W}_2$. In computing the Monte Carlo critical values we use $N = 40000$ draws and $L = 50$ for the number of directions of $\mathbf{C}$. We conduct robustness checks with $L = 1$ and $L = 100$.

The supporting web site for Ruud (2000) provides a MATLAB transcription of Imhof (1961)'s algorithm to compute $F_{\mathbf{C},\mathbf{W}_2}(x)$ for a given $\mathbf{C}$ and $\mathbf{W}_2$. This allows us to compute the actual sizes $F_{\mathbf{C},\mathbf{W}_2}(c^P(5\%, \mathbf{W}_2, 10))$ and $F_{\mathbf{C},\mathbf{W}_2}(c^m(5\%, \mathbf{W}_2, 10))$ at an accuracy level of 0.01% for any $\mathbf{C}$, $\mathbf{W}_2$.

We then compute the maximal sizes

$$maxsize^m(\mathbf{W}_2) = \max_{\mathbf{C}\in\Lambda^m} F_{\mathbf{C},\mathbf{W}_2}(c^m(5\%, \mathbf{W}_2, 10))$$

and

$$maxsize^P(\mathbf{W}_2) = \max_{\mathbf{C}\in\Lambda^m} F_{\mathbf{C},\mathbf{W}_2}(c^P(5\%, \mathbf{W}_2, 10)).$$

We find that $|c^P(5\%, \mathbf{W}_2, 10) - c^m(5\%, \mathbf{W}_2, 10)|/c^m(5\%, \mathbf{W}_2, 10) \leq 4.4\%$ for all $\mathbf{W}_2 \in \mathcal{W}_2$. Moreover,

$$\begin{aligned} 4.77\% &\leq maxsize^m(\mathbf{W}_2) \leq 5.26\% \ \forall \mathbf{W}_2 \in \mathcal{W}_2 \\ 5.00\% &\leq maxsize^P(\mathbf{W}_2) \leq 5.02\% \ \forall \mathbf{W}_2 \in \mathcal{W}_2 \end{aligned}$$

B.3.1. Robustness of Size Distortions

One might be concerned that we find artificially small size distortions because we replace the set $\mathbf{\Lambda}$ by a finite set $\mathbf{\Lambda^m}$. We therefore repeat our calculations for a much smaller set $\mathbf{\Lambda}^{m,small}$ with $L = 1$ and one much larger set $\mathbf{\Lambda}^{m,large}$ with $L = 100$ and find that the size distortions of both methodologies are robust. When we use $\mathbf{\Lambda}^{m,small}$ we find that

$4.77\% \leq maxsize^m(\mathbf{W}_2) \leq 5.29\% \ \forall \mathbf{W}_2 \in \mathcal{W}_2$ and

$5.00\% \leq maxsize^P(\mathbf{W}_2) \leq 5.02\% \ \forall \mathbf{W}_2 \in \mathcal{W}_2$.

When we use $\mathbf{\Lambda}^{m,large}$ we find that

$4.77\% \leq maxsize^m(\mathbf{W}_2) \leq 5.26\% \ \forall \mathbf{W}_2 \in \mathcal{W}_2$ and

$5.00\% \leq maxsize^P(\mathbf{W}_2) \leq 5.02\% \ \forall \mathbf{W}_2 \in \mathcal{W}_2$.

B.4. Numerical Implementation of $B_{TSLS}(\mathbf{W}, \mathbf{\Omega})$ and $B_{LIML}(\mathbf{W}, \mathbf{\Omega})$

$$\begin{aligned} B_{TSLS}(\mathbf{W}, \mathbf{\Omega}) &= sup_{\beta \in \mathbb{R}, \mathbf{C}_0 \in S^{K-1}} \left| \frac{N_{TSLS}(\beta, \mathbf{C}, \mathbf{W}, \mathbf{\Omega})}{\mu^{-2} BM(\beta, \mathbf{W})} \right| &(2) \\ &= sup_{\beta \in \mathbb{R}, \mathbf{C}_0 \in S^{K-1}} \left| \frac{tr\mathbf{S}_{12}}{\sqrt{tr\mathbf{S}_2 tr\mathbf{S}_1}} \left[1 - 2\frac{\mathbf{C}_0'\mathbf{S}_{12}\mathbf{C}_0}{tr\mathbf{S}_{12}} \right] \right| &(3) \end{aligned}$$

Now we use that for any given $\beta \in \mathbb{R}$ the maximum and minimum of $\mathbf{C}_0'\mathbf{S}_{12}\mathbf{C}_0$ are $maxeval(\frac{1}{2}\mathbf{S}_{12} + \frac{1}{2}\mathbf{S}_{12}')$ and $mineval(\frac{1}{2}\mathbf{S}_{12} + \frac{1}{2}\mathbf{S}_{12}')$. Denote $\mathbf{S}_{12}^{sym} = \frac{1}{2}\mathbf{S}_{12} + \frac{1}{2}\mathbf{S}_{12}'$. Then

$$B_{TSLS}(\mathbf{W}, \mathbf{\Omega}) = sup_{\beta \in \mathbb{R}} \frac{max\left(|tr\mathbf{S}_{12} - 2mineval(\mathbf{S}_{12}^{sym})|, |tr\mathbf{S}_{12} - 2maxeval(\mathbf{S}_{12}^{sym})|\right)}{\sqrt{tr\mathbf{S}_2 tr\mathbf{S}_1}} \quad (4)$$

The function defined on the real line $g_{TSLS}(\beta) = \frac{max\left(|tr\mathbf{S}_{12} - 2mineval(\mathbf{S}_{12}^{sym})|, |tr\mathbf{S}_{12} - 2maxeval(\mathbf{S}_{12}^{sym})|\right)}{\sqrt{tr\mathbf{S}_2 tr\mathbf{S}_1}}$ converges to $1 - 2\frac{mineval(\mathbf{W}_2)}{tr\mathbf{W}_2}$ as $\beta \to \pm\infty$.

The empirical researcher can specify $\epsilon > 0$, the desired fractional error relative to $\lim_{\beta \to \pm\infty} g_{TSLS}(\beta)$, and the number of starting points *points* for numerical maximization routines. The defaults are set to $\epsilon = 0.001$ and $points = 10000$. The program then finds β_{range} such that $\left|\frac{f_{TSLS}(\pm\beta_{range})}{\lim_{\beta\to\pm\infty} g_{TSLS}(\beta)} - 1\right| \leq \epsilon$. We maximize g_{TSLS} using the MATLAB routine fminsearch using $points$ equally spaced starting points in $[-\beta_{range}, \beta_{range}]$. We also maximize g_{TSLS} over the range $[-\beta_{range}, \beta_{range}]$ using the MATLAB routine fminbnd. Since each of these methodologies might only yield local maxima, we take the maximum over the local maxima to obtain $B_{TSLS}(\mathbf{W}, \mathbf{\Omega})$.

The numerical computation for $B_{LIML}(\mathbf{W}, \mathbf{\Omega})$ is analogous with

$$g_{LIML}(\beta) = max\left(\left|\frac{tr\mathbf{S}_{12} - \frac{\sigma_{12}}{\sigma_1^2} tr\mathbf{S}_1 - maxeval\mathbf{M_B}}{\sqrt{tr\mathbf{S}_1}\sqrt{tr\mathbf{S}_2}}\right|, \left|\frac{tr\mathbf{S}_{12} - \frac{\sigma_{12}}{\sigma_1^2} tr\mathbf{S}_1 - mineval\mathbf{M_B}}{\sqrt{tr\mathbf{S}_1}\sqrt{tr\mathbf{S}_2}}\right|\right) \tag{5}$$

where $\mathbf{M_B} = \frac{1}{2}(2\mathbf{S}_{12} - \frac{\sigma_{12}}{\sigma_1^2}\mathbf{S}_1) + \frac{1}{2}(2\mathbf{S}_{12} - \frac{\sigma_{12}}{\sigma_1^2}\mathbf{S}_1)'$ and

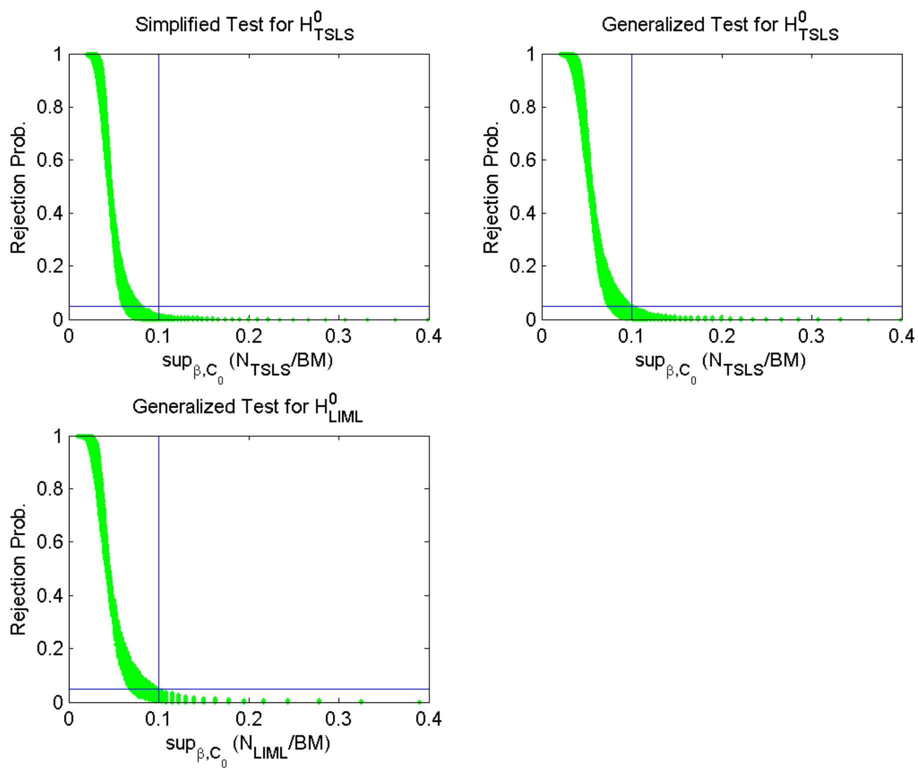
$$g_{LIML}(\beta) \to \frac{maxeval\mathbf{W}_2}{tr\mathbf{W}_2} \text{ as } \beta \to \pm\infty \tag{6}$$

B.5 Power of Tests

We now seek to understand the rejection probabilities of the generalized and simplified testing procedures. We compute each testing procedure's rejection probability against $B_e/\mu^2 = \sup_{\beta\in\mathbb{R}, C_0\in\mathcal{S}^{K-1}}\left(|N_e|/BM\right)$. Ideally we would like the rejection probability to approach zero as B_e/μ^2 approaches zero and one for B_e/μ^2 large. Figure B.1 plots rejection probabilities against $B_e/\mu^2 = \sup_{\beta\in\mathbb{R}, \mathbf{C}_0\in\mathcal{S}^{K-1}}\left(|N_e|/BM\right)$ for $\mathbf{W}$ and $\mathbf{\Omega}$ equal to their estimated counterparts in our empirical example for the USA with the real ex post interest rate as our endogenous regressor. For a fixed direction $\mathbf{C}_0$ we compute rejection probabilities from 1000 independent draws from the distribution $\boldsymbol{\gamma}_2'\boldsymbol{\gamma}_2/tr(\mathbf{W}_2)$. We draw 1000 directions $\mathbf{C}_0$ from a uniform distribution on $\mathcal{S}^{K-1}$ and plot the rejection probability for each $\mathbf{C}_0$ and $\mu^2 = 0, 0.2, 0.4, ..., 40$.

For a given B_e/μ^2 we obtain a range of rejection probabilities corresponding to different directions $\mathbf{C}_0$. We see that the test is asymptotically valid and the rejection probability is less than the size for $B_e/\mu^2 >= 10\%$. For both the generalized and the simplified test, when B_e/μ^2 is large the rejection probability approaches zero and we are very unlikely to falsely reject the null. As B_e/μ^2 becomes small the rejection probability becomes large and we are very likely to correctly reject the null. Figure B.1 also illustrates that the rejection probability is always larger for the simplified procedure than for the generalized procedure but both procedures behave similarly when B_e/μ^2 becomes large or small.

Figure B.1: Rejection Probabilities for Generalized and Simplified Tests with $\tau = 10\%$, $\alpha = 5\%$



B.6 Comparing Robust Critical Values to Stock and Yogo(05)

We now compare the critical values for our testing procedure to those in Stock and Yogo (2005). Assume for now that the errors are conditionally homoskedastic and serially uncorrelated, so that that $W = \mathbf{\Omega} \otimes \mathbb{I}_K$. We then obtain $B_{TSLS}(\mathbf{\Omega} \otimes \mathbb{I}_K, \mathbf{\Omega}) = 1 - 2/K$

and $B_{LIML}(\mathbf{\Omega} \otimes \mathbb{I}_K, \mathbf{\Omega}) = 1/K$. We consider $\alpha = 5\%$ and $\tau = 10\%$. We compare to the Stock and Yogo (2005) critical values for the null hypothesis that the asymptotic estimator bias exceeds 10% of the asymptotic OLS bias with size 5%.

Our generalized and simplified critical values differ from those proposed by Stock and Yogo (2005) for the TSLS bias even when first- and second-stage errors are perfectly conditionally homoskedastic and serially uncorrelated. We consider the Stock and Yogo (2005) 5% critical value for testing the null hypothesis that the TSLS bias exceeds 10% of the OLS bias and generalized and simplified critical values with a threshold of 10% and size 5%. Table B.2 shows that the TSLS critical values critical values are smaller than the Stock and Yogo (2005) critical values for $K = 3, 4$ but larger than the Stock and Yogo (2005) critical values for $K \geq 5$. The difference between the TSLS and Stock and Yogo (2005) critical values is always less than 1. The simplified critical values exceed the Stock and Yogo (2005) critical values but the difference declines in the number of instruments K. The LIML critical values decline more rapidly with the number of instruments than either the TSLS or simplified critical values.

The TSLS critical values for our generalized procedure could differ from Stock and Yogo (2005) for two reasons.

First, we use the "worst-case" benchmark instead of the OLS bias. Denote the asymptotic OLS bias by $Bias_{OLS} = \sigma_{12}/\sigma_2^2$. For any structural error correlation $\rho \in (-1, 1)$ there exists a structural parameter $\beta \in \mathbb{R}$ such that $(\omega_{12} - \beta\omega_2^2)/(\sqrt{\omega_1^2 - \beta\omega_{12} + \beta^2\omega_2^2}\omega_2) = \rho$, provided that $\mathbf{\Omega}$ is nonsingular. In the conditionally homoskedastic serially uncorrelated model, the Nagar bias is independent of the direction $\mathbf{C}_0 \in \mathcal{S}^{K-1}$ with

$$N_{TSLS}(\beta, \mathbf{C}, \mathbf{W}, \mathbf{\Omega}) = Bias_{OLS}\left(1 - \frac{2}{K}\right) \tag{7}$$

$$N_{LIML}(\beta, \mathbf{C}, \mathbf{W}, \mathbf{\Omega}) = -Bias_{OLS}/K \tag{8}$$

Table B.2
Comparing to Stock and Yogo (2005)

K	c_{TSLS}	c_{LIML}	c_{Simp}	c_{TSLS}^{SY}
1	23.11	23.11	23.11	N/A
2	3.00	12.17	19.29	N/A
3	8.53	8.53	17.67	9.08
4	10.23	6.70	16.72	10.27
5	11.06	5.61	16.08	10.83
6	11.52	4.87	15.62	11.12
7	11.80	4.35	15.26	11.29
8	11.99	3.96	14.97	11.39
9	12.11	3.65	14.73	11.46
10	12.19	3.41	14.53	11.49
11	12.25	3.21	14.36	11.51
12	12.29	3.04	14.21	11.52
13	12.32	2.90	14.08	11.52
14	12.33	2.78	13.96	11.52
15	12.35	2.67	13.86	11.51
16	12.35	2.58	13.77	11.50
17	12.35	2.50	13.68	11.49
18	12.35	2.42	13.60	11.48
19	12.35	2.36	13.53	11.46
20	12.35	2.30	13.46	11.45
21	12.34	2.25	13.40	11.44
22	12.34	2.20	13.35	11.42
23	12.33	2.15	13.29	11.41
24	12.32	2.11	13.24	11.40
25	12.31	2.07	13.20	11.38
26	12.31	2.04	13.15	11.37
27	12.30	2.01	13.11	11.36
28	12.29	1.98	13.07	11.34
29	12.28	1.95	13.04	11.33
30	12.27	1.92	13.00	11.32

NOTE: We show simplified, TSLS, and LIML critical values assuming conditional homoskedasticity, no serial autocorrelation, and known $\mathbf{\Omega}$ and $\mathbf{W}$. The null hypothesis is that the Nagar bias exceeds 10% of the benchmark. Critical values have size 5% and are computed with the Patnaik methodology. c_{TSLS}^{SY} denotes Stock and Yogo (2005) 5% critical values of the null hypothesis that the asymptotic TSLS bias exceeds 10% of the asymptotic OLS bias.

and so

$$\sup_{\beta\in\mathbb{R},\mathbf{C}_0\in\mathcal{S}^{K-1}} \left(\left|N_e(\beta,\mathbf{C},\mathbf{W},\mathbf{\Omega})\right|/BM(\beta,\mathbf{W})\right) = \sup_{\beta\in\mathbb{R}} \left(\left|N_e(\beta,\mathbf{C},\mathbf{W},\mathbf{\Omega})\right|/BM(\beta,\mathbf{W})\right) \quad (9)$$

$$= \left|N_e(\beta,\mathbf{C},\mathbf{W},\mathbf{\Omega})\right|/Bias_{OLS} \quad (10)$$

Hence our choice of benchmark is not a source of divergence of our critical values from Stock and Yogo (2005).

Second, we define the null hypothesis in terms of the Nagar bias instead of the asymptotic estimator bias. Therefore, the only source of divergence of our TSLS critical values from Stock and Yogo (2005) critical values in Table B.2 is the Nagar bias approximation.

The simplified procedure also allows for the worst type of heteroskedasticity, serial correlation and/or clustering in the second stage, in contrast to Stock and Yogo (2005). Therefore, simplified critical values are higher than Stock and Yogo (2005) critical values, even when the first-stage errors are estimated to be perfectly conditionally homoskedastic and serially uncorrelated.

C. Additional Results

C.1 Nagar Bias Approximates Asymptotic Bias

Assume that $K \geq 3$. We now prove that

$$\mathbb{E}[\beta^*_{TSLS}] - N_{TSLS} = O(\mu^{-4}) \tag{11}$$

Write $\mathbf{s}_2 = \frac{\mathbf{S}_2}{tr(\mathbf{S}_2)}$ and $\mathbf{s}_1 = \frac{\mathbf{S}_1}{tr(\mathbf{S}_1)}$. Define the normalized statistic $H \equiv \mu \times \beta^*_{TSLS} / \sqrt{tr\left(\mathbf{S}_1\right)/tr\left(\mathbf{S}_2\right)}$ $= \left(\mathbf{C}_0' \mathbf{s}_1^{1/2} \mathbf{z}_u + \mathbf{z}_v{}' \mathbf{s}_2^{1/2} \mathbf{s}_1^{1/2} \mathbf{z}_u / \mu\right) \left(1 + 2\mathbf{z}_v' \mathbf{s}_2^{1/2} \mathbf{C}_0 / \mu + \mathbf{z}_v' \mathbf{s}_2 \mathbf{z}_v / \mu^2\right)$. Denote the numerator of H by a and the denominator by A.

Expand H as a stochastic power series expansion in $1/\mu$ to get

$$\begin{aligned} H &= \mathbf{C}_0' \mathbf{s}_1^{1/2} \mathbf{z}_u + \frac{1}{\mu}\left[\mathbf{z}_v' \mathbf{s}_2^{1/2} \mathbf{s}_1^{1/2} \mathbf{z}_u - 2\left(\mathbf{z}_v' \mathbf{s}_2^{1/2} \mathbf{C}_0\right)\left(\mathbf{C}_0' \mathbf{s}_1^{1/2} \mathbf{z}_u\right)\right] \\ &\quad + \frac{1}{\mu^2}\left[\mathbf{C}_0' \mathbf{s}_1^{1/2} \mathbf{z}_u \left(2\mathbf{z}_v' \mathbf{s}_2^{1/2} \mathbf{C}_0\right)^2 - \mathbf{C}_0' \mathbf{s}_1^{1/2} \mathbf{z}_u \mathbf{z}_v' \mathbf{s}_2 \mathbf{z}_v - 2\mathbf{z}_v' \mathbf{s}_2^{1/2} \mathbf{C}_0 \mathbf{z}_v' \mathbf{s}_2^{1/2} \mathbf{s}_1^{1/2} \mathbf{z}_u\right] + \tilde{R} \end{aligned}$$

Denote the first three terms by d'. The expectation of the first three terms equals $N_{TSLS} \times \mu / \sqrt{tr\left(\mathbf{S}_1\right)/tr\left(\mathbf{S}_2\right)}$.

Write $\Delta A = A - 1$. Both a and $\mu \Delta A$ are finite polynomials in the components of z with $O(1)$ coefficients. A geometric series expansion gives $H \times [1 - (-\Delta A)^3] = [a/(1+\Delta A)]\,[1 - (-\Delta A)^3]$ $= a \sum_{s=0}^{2} (-\Delta A)^s$. We can re-write this as $H = a \sum_{s=0}^{2} (-\Delta A)^s + (-\Delta A)^3 h$. Now show that $E\left[(-\Delta A)^3 h \mu^3\right] = O(1)$ as $\mu \to \infty$. Following the proof in Sargan (1974) write the expectation as an integral. Provided that the expectation exists, we know that

$$\begin{aligned} |E(-\Delta A)^3 h \mu^3| &\leq \frac{1}{(2\pi)^{K/2}} \int_{\mathbf{z} \in \mathbb{R}^K} |(\mu \Delta A(\mu, \mathbf{z}))^3| |h(\mu, \mathbf{z})| \exp\left(-\frac{1}{2}\mathbf{z}'\mathbf{z}\right) d\mathbf{z} \\ &= \frac{1}{(2\pi)^{K/2}} \int_{\mathbf{z} \in \mathbb{R}^K} |(\mu \Delta A(\mu, \mathbf{z}))^3 a| \exp\left(-\frac{1}{4}\mathbf{z}'\mathbf{z}\right) |\frac{1}{A}| \exp\left(-\frac{1}{4}\mathbf{z}'\mathbf{z}\right) d\mathbf{z} \end{aligned}$$

But $(\mu\Delta A(\mu,\mathbf{z}))^3 a$ is a polynomial in $\mathbf{z}$ with coefficients $O(1)$ as $\mu \to \infty$. Hence $\exists$ constant B^* such that $\forall \mathbf{z} \in \mathbb{R}^K$ $|(\mu\Delta A(\mu,\mathbf{z}))^3 a \exp^{-\frac{1}{4}\mathbf{z}'\mathbf{z}}| \leq B^*$. Then,

$$
\begin{aligned}
|E(-\Delta A)^3 h\mu^3| &\leq B^* \frac{1}{(2\pi)^{K/2}} \int_{\mathbf{z}\in\mathbb{R}^K} A^{-1} \exp^{-\frac{1}{4}\mathbf{z}'\mathbf{z}}\, dz \\
&= 2^{K/2} B^* \mathbb{E}\left[\frac{\mu^2/2}{(\mathbf{s}_2^{1/2}\mathbf{z}_v + \mathbf{C}_0\mu/\sqrt{2})'(\mathbf{s}_2^{1/2}\mathbf{z}_v + \mathbf{C}_0\mu/\sqrt{2})}\right] \qquad (12)
\end{aligned}
$$

(12) can be bounded by the inverse moment of a non-central chi-square with noncentrality parameter proportional to μ^2 and 3 degrees of freedom, proving existence. Let $X \sim \chi_3^2(y)$ be a non-central chi-square random variable with non-centrality parameter y and 3 degrees of freedom. Bock, Judge and Yancey (1984) show that $E[X^{-1}] = (\Gamma(1/2)/\sqrt{\pi})(\frac{y}{2})^{-\frac{1}{2}} D\left((y/2)^{\frac{1}{2}}\right)$. Γ is the Gamma function and D is Dawson's integral, tabulated in Abramowitz and Stegun (1964). $D(y) = \frac{y^{-1}}{2} + O\left(y(-\frac{3}{2})\right)$ as $y \to \infty$ proving that (12) is $O(1)$ as $\mu \to \infty$.

By the uniqueness of the Taylor expansion the $O(1)$, $O(1/\mu)$ and $O(1/\mu^2)$ terms of d' and $a\sum_{s=0}^{2}(-\Delta A)^s$ must agree. Moreover, both d' and $a\sum_{s=0}^{2}(-\Delta A)^s$ are finite polynomials in normal random variables and this completes the proof.

C.2 Primitive Conditions for Independent Data

We now specify a set of primitive conditions for independent (not necessarily identically distributed) data that imply Assumption HL. While Assumption HL is more general and can allow for serially autocorrelated data, this case encompasses cross-sectional heteroskedastic models with independent observations and linear panel data models with fixed effects and independent clusters. Assumption HL is implied by standard results for independent processes.

The main results of this section are summarized as follows. First, we show that a class of cross-sectional models satisfy Assumption HL and:

$$\begin{aligned}
\mathbf{W} &\equiv \lim_{S\to\infty}(1/S)\sum_{s=1}^{S}\mathbb{E}[\mathbf{V}_s\mathbf{V}_s' \otimes \mathbf{Z}_s\mathbf{Z}_s'] \\
\mathbf{\Omega} &\equiv \lim_{S\to\infty}(1/S)\sum_{s=1}^{S}\mathbb{E}[\mathbf{V}_s\mathbf{V}_s'] \\
\widehat{\mathbf{W}} &\equiv (1/S)\sum_{s=1}^{S}\widehat{\mathbf{V}}_s\widehat{\mathbf{V}}_s' \otimes \mathbf{Z}_s\mathbf{Z}_s'
\end{aligned}$$

where $\widehat{\mathbf{V}}_s$ are OLS estimates of the reduced form errors of the model.

Second, we verify assumption Assumption HL in a class of linear panel data models with fixed effects and clustered data. Suppose $\{\mathbf{Z}_s, v_{1s}, v_{2s}\}_{s=1}^{S}$ corresponds to the within transformation (Wooldridge (2002)) of the instrumental variables and the endogenous regressors in a linear panel data model with additive fixed effects. Assume that the data is partitioned according to L independent clusters and that the sample size $(S \equiv L \times M)$ grows as the number of observations per cluster (M) stays constant and L grows to infinity. Write $s \in S_l$ if observation s is in cluster l and allow for an arbitrary correlation structure within clusters. In this case, Assumption HL is satisfied with

$$\begin{aligned}
\mathbf{W} &\equiv \lim_{L\to\infty}(1/L)\sum_{l=1}^{L}\mathbb{E}\Big[(1/M)\Big(\sum_{s\in S_l}(\mathbf{V}_s \otimes \mathbf{Z}_s)\Big)\Big(\sum_{s\in S_l}(\mathbf{V}_s \otimes \mathbf{Z}_s)\Big)'\Big] \\
\mathbf{\Omega} &\equiv \lim_{L\to\infty}(1/L)\sum_{l=1}^{L}\Big((1/M)\sum_{s\in S_l}\mathbb{E}[\mathbf{V}_s\mathbf{V}_s']\Big) \\
\widehat{\mathbf{W}} &\equiv (1/L)\sum_{l=1}^{L}\mathbb{E}\Big[(1/M)\Big(\sum_{s\in S_l}(\hat{\mathbf{V}}_s \otimes \mathbf{Z}_s)\Big)\Big(\sum_{s\in S_l}(\hat{\mathbf{V}}_s \otimes \mathbf{Z}_s)\Big)'\Big]
\end{aligned}$$

where $\widehat{\mathbf{V}}_s$ are the OLS estimates of the reduced form errors based of the model, and $\{\mathbf{y}_s, \mathbf{Y}_s, \mathbf{Z}_s\}$ correspond to the within transformations of the variables. This is equivalent to estimating the reduced form errors with the fixed-effects estimator applied to the original panel data, clustering at the S_l level.

C.2.1 Primitive Conditions for HL.1

Let $\{\mathbf{X}_s\}_{s=1}^S$ be an independent $\mathbb{R}^p$-valued process. Let

1. $\mathbb{E}[\mathbf{X}_s] = 0$ for $i = 1 \dots S$
2. $\overline{\mathbf{W}}_S \equiv (1/S)\sum_{s=1}^S \mathbb{E}[\mathbf{X}_s\mathbf{X}_s']$ is positive definite for S sufficiently large.
3. $\mathbf{W} \equiv \lim_{S\to\infty} \overline{\mathbf{W}}_S < \infty$
4. There exists $\delta > 0$ such that for all $\boldsymbol{\lambda} \in \mathbb{R}^p$

$$\sum_{s=1}^S \mathbb{E}\left|\boldsymbol{\lambda}\overline{\mathbf{W}}_S^{-1/2}\Big(\mathbf{X}_s\Big)\right|^{2+\delta} \times S^{-(2+\delta)/2} \to 0$$

Theorem 3.1 in White (1980), pg. 729, imply that $\sum_{s=1}^S \mathbf{X}_s/\sqrt{S} \xrightarrow{d} N_p(\mathbf{0}, \mathbf{W})$.

HL.1 for Cross-Sectional Heteroskedastic Models: Let $\{\mathbf{Z}_s, v_{1s}, v_{2s}\}_{s=1}^S$ be an independent process. Let $\mathbf{X}_s \equiv (\mathbf{V}_s \otimes \mathbf{Z}_s)$ where $\mathbf{V}_s = (v_{1s}, v_{2s})'$. If $\mathbf{X}_s$ satisfies assumptions 1-4 then $\Big((\mathbf{Z}'\mathbf{v}_1)'/\sqrt{S}, (\mathbf{Z}'\mathbf{v}_2)'/\sqrt{S}\Big)' \xrightarrow{d} \mathcal{N}_{2K}(\mathbf{0}, \mathbf{W})$ where $\mathbf{W} = \lim_{S\to\infty}(1/S)\sum_{s=1}^S \mathbb{E}[\mathbf{V}_s\mathbf{V}_s' \otimes \mathbf{Z}_s\mathbf{Z}_s']$

HL.1 for Linear Panel Data Models with Fixed-Effects and Clustering: Note that:

$$\mathbf{Z}'\mathbf{v}_i/\sqrt{S} = \frac{1}{\sqrt{L}}\sum_{l=1}^L \left(\frac{1}{\sqrt{M}}\sum_{s\in S_l} \mathbf{Z}_s v_{is}\right)$$

Let $\mathbf{X}_l \equiv (1/\sqrt{M})\sum_{s\in S_l}(\mathbf{V}_s \otimes \mathbf{Z}_s)$. Since observations are independent across clusters, the process $\{\mathbf{X}_l\}$ is independent. Therefore, the primitive conditions 1-4 yield the CLT $\Big((\mathbf{Z}'\mathbf{v}_1)'/\sqrt{S}, (\mathbf{Z}'\mathbf{v}_2)'/\sqrt{S}\Big)' \xrightarrow{d} \mathcal{N}_{2K}(\mathbf{0}, \mathbf{W})$, where the asymptotic covariance matrix $\mathbf{W} = \lim_{L\to\infty}(1/L)\sum_{l=1}^L \mathbb{E}\Big[(1/M)\Big(\sum_{s\in S_l}(\mathbf{V}_s \otimes \mathbf{Z}_s)\Big)\Big(\sum_{s\in S_l}(\mathbf{V}_s \otimes \mathbf{Z}_s)\Big)'\Big]$

C.2.2 Primitive Conditions for Assumption HL.2

Let $\{\mathcal{Z}_s\}$ be a sequence of independent random variables. Let

1. $\mu_s = \mathbb{E}[\mathcal{Z}_s] < \infty$ for all s.

2. For some $\delta > 0$, $\lim_{S\to\infty} \sum_{s=1}^{S} \left(|\mathcal{Z}_s - \mu_s|^{1+\delta} \right) / s^{1+\delta} < \infty$.

3. $\Omega \equiv \lim_{S\to\infty}(1/S) \sum_{s=1}^{S} \mu_s < \infty$

By Corollary 3.9 in White (2001), pg. 35, $(1/S) \sum_{s=1}^{S} (\mathcal{Z}_s - \mu_s) \xrightarrow{a.s.} 0$.

HL.2 for Cross-Sectional Heteroskedastic Models: Let $\{\mathbf{Z}_s, v_{1s}, v_{2s}\}_{s=1}^{S}$ be an independent process. For $i, j = 1, 2$, let $\mathcal{Z}_{ijs} \equiv v_{is} v_{js}$. If $\{\mathcal{Z}_{ijs}\}_{s=1}^{\infty}$ satisfy 1-2 for all $i, j = 1, 2$, then $(1/S) \sum_{s=1}^{S} \mathbf{V}_s \mathbf{V}_s' \xrightarrow{p} \mathbf{\Omega}$ where $\mathbf{\Omega} = \lim_{S\to\infty}(1/S) \sum_{s=1}^{S} \mathbb{E}[\mathbf{V}_s \mathbf{V}_s']$

HL.2 for for Linear Panel Data Models with Fixed-Effects and Clustering: Note that

$$\frac{1}{S} \mathbf{v}_i' \mathbf{v}_j = \frac{1}{L} \sum_{l=1}^{L} \left(\frac{1}{M} \sum_{s \in S_l} v_{is} v_{js} \right)$$

Let $\mathcal{Z}_{ijs} = \left((1/M) \sum_{s \in S_l} v_{is} v_{js} \right)$. Since observations are independent across clusters, it follows that $\{\mathcal{Z}_{ijs}\}$ is an independent sequence. If the sequence satisfies 1-3, then it follows that $(1/S) \sum_{s=1}^{S} \mathbf{V}_s \mathbf{V}_s' \xrightarrow{p} \mathbf{\Omega}$ where $\mathbf{\Omega} = \lim_{L\to\infty}(1/L) \sum_{l=1}^{L} \left((1/M) \sum_{s \in S_l} \mathbb{E}[\mathbf{V}_s \mathbf{V}_s'] \right)$.

C.2.3 Primitive Conditions for Assumption HL.3

Exercise 6.8, pg. 146, in White (2001) provides sufficient conditions for consistent estimation of the asymptotic variance in a multivariate linear model. Let:

1. $\boldsymbol{Y}_s = \boldsymbol{X}_s' \boldsymbol{\beta} + \boldsymbol{\epsilon}_s, \quad s = 1, 2, \ldots \boldsymbol{\beta} \in \mathbb{R}^P, \boldsymbol{Y}_s \in \mathbb{R}^N, \boldsymbol{X}_s \in \mathbb{R}^{P \times N}$

2. Let $\{\boldsymbol{X}_s, \boldsymbol{\epsilon}_s\}$ be an independent sequence (so that ϕ is of size -1, with $r = 1$, see White (2001) page. 146).

3. $\mathbb{E}[\boldsymbol{X}_s \boldsymbol{\epsilon}_s] = 0$ for all s.

4. $\mathbb{E} \left| X_{spn} \epsilon_{sn} \right|^{2(1+\delta)} < \Delta < \infty$ for some $\delta > 0$ and all $n = 1 \ldots N, p = 1 \ldots P$ and s.

5. $\boldsymbol{V}_n \equiv \text{var}\left((1/S) \sum_{s=1}^{S} \boldsymbol{X}_s \boldsymbol{\epsilon}_s \right)$ is uniformly positive definite.

6. $\mathbb{E}\Big|X_{sp_1n}X_{sp_2n}\Big|^{2(1+\delta)} < \Delta < \infty$ for some $\delta > 0$ and all $n = 1 \ldots N, p_1, p_2 = 1 \ldots P$ and s.

7. $\mathbb{E}[(1/S)\sum_{s=1}^{S} \boldsymbol{X}_s \boldsymbol{X}_s']$ has uniformly full column rank and is uniformly positive definite.

Define

$$V_n \equiv (1/S)\sum_{s=1}^{S} \mathbb{E}[\boldsymbol{X}_s \boldsymbol{\epsilon}_s \boldsymbol{\epsilon}_s' \boldsymbol{X}_s']$$

and

$$\widehat{V}_n \equiv (1/S)\sum_{s=1}^{S} \boldsymbol{X}_s \widehat{\boldsymbol{\epsilon}}_s \widehat{\boldsymbol{\epsilon}}_s' \boldsymbol{X}_s'$$

where $\widehat{\boldsymbol{\epsilon}}_s = \widehat{\boldsymbol{Y}}_s - \widehat{\boldsymbol{X}}_s' \widehat{\boldsymbol{\beta}}_{OLS}$, and $\widehat{\boldsymbol{\beta}}_{OLS} = (\sum_{s=1}^{S} \boldsymbol{X}_s \boldsymbol{X}_s')^{-1} \sum_{s=1}^{S} \boldsymbol{X}_s \boldsymbol{Y}_s$. Exercise 6.8, pg. 146 (with $\boldsymbol{Z}_s = \boldsymbol{X}_s$) in White (2001) implies that $\widehat{V}_n - V_n \xrightarrow{p} 0$. In fact, the result holds for any $\widehat{\boldsymbol{\beta}}$ such that $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \xrightarrow{p} 0$

HL.3 for Cross-Sectional Heteroskedastic Models: Let $\{Z_s, v_{1s}, v_{2s}\}$ be an independent sequence. Let $\boldsymbol{Y}_s = (\mathbf{y}_s', \mathbf{Y}_s')', \boldsymbol{X}_s = (\mathbb{I}_2 \otimes \mathbf{Z}_s)$, $\boldsymbol{\beta} = (\boldsymbol{\Gamma}_1', \boldsymbol{\Gamma}_2')'$, $\boldsymbol{\epsilon}_s = V_s$, $\widehat{\boldsymbol{\epsilon}_s} \equiv \widehat{Y}_s - \boldsymbol{X}_s' \widehat{\boldsymbol{\beta}}_{OLS} = \widehat{V}_s$. Note that $\widehat{\boldsymbol{\beta}}_{OLS}$ corresponds to the reduced form OLS estimates for $\boldsymbol{\Gamma}$. If 1-7 holds then

$$(1/S)\sum_{s=1}^{S}\Big[\widehat{\mathbf{V}}_s \widehat{\mathbf{V}}_s' \otimes \mathbf{Z}_s \mathbf{Z}_s'\Big] \xrightarrow{p} \mathbf{W} \equiv \lim_{S\to\infty} (1/S)\sum_{s=1}^{S} \mathbb{E}\Big[\mathbf{V}_s \mathbf{V}_s' \otimes \mathbf{Z}_s \mathbf{Z}_s'\Big]$$

HL.3 for Linear Panel Data Models with Fixed Effects and Clustering: Let $\{Z_s, v_{1s}, v_{2s}\}$ corresponds to the within transformation of the instrumental variables and the reduced form errors in a linear panel data model with fixed effects. Define $\boldsymbol{Y}_l = (y_{l1}, \ldots y_{lM}, Y_{l1}, \ldots Y_{lM})'$, $\boldsymbol{X}_l = (\mathbb{I}_2 \otimes (Z_{l1}, \ldots Z_{lM}))$, $\boldsymbol{\beta} = (\boldsymbol{\Gamma}_1', \boldsymbol{\Gamma}_2')'$, and the innovations $\boldsymbol{\epsilon}_s = (v_{1l1}, \ldots v_{1lM}, v_{2l1}, \ldots v_{2lM})'$. Since clusters are independent, the sequence $\{\boldsymbol{X}_s, \boldsymbol{\epsilon}_s\}$ is independent as well. In this case, $\widehat{\boldsymbol{\beta}}_{OLS}$ corresponds to the fixed effects estimator for $\boldsymbol{\Gamma}$ in the reduced form model. If 1-7 holds then:

$$(1/L)\sum_{l=1}^{L}\Big[(1/M)\Big(\sum_{s\in S_l}(\widehat{\mathbf{V}}_s\otimes\mathbf{Z}_s)\Big)\Big(\sum_{s\in S_l}(\widehat{\mathbf{V}}_s\otimes\mathbf{Z}_s)\Big)'\Big]\xrightarrow{p}\mathbf{W}$$

where $\mathbf{W}\equiv\lim_{L\to\infty}(1/L)\sum_{l=1}^{L}\mathbb{E}\Big[(1/M)\Big(\sum_{s\in S_l}(\mathbf{V}_s\otimes\mathbf{Z}_s)\Big)\Big(\sum_{s\in S_l}(\mathbf{V}_s\otimes\mathbf{Z}_s)\Big)'\Big]$

C.3 Uniformity

Let Γ be a parameter space. We say that a testing procedure with test statistic T_S has asymptotic size α in the uniform sense if

$$\lim\sup_{S\to\infty}\sup_{\gamma\in\Gamma}\mathbb{P}_\gamma(T_S>c)\leq\alpha$$

Equivalently, the testing procedure T has size α (in a uniform sense) if under any sequence of parameter values $\{\gamma_S\}\in\Gamma$:

$$\lim\sup_{S\to\infty}\mathbb{P}_{\gamma_S}(T_S>c)\leq\alpha$$

See Guggenberger (2010*a*), Guggenberger (2010*b*).

C.3.1 Uniformity problems with pretests

Guggenberger (2010*a*) studies tests for a structural parameter β that follow a Hausman exogeneity pretest in an IV set-up. The Hausman pretest looks at a properly scaled difference between the TSLS and OLS estimators and tests the null hypothesis that their difference is zero.

Let x denote a sample of size S. Guggenberger (2010a) shows that the test T for a structural parameter β that follows a pretest

$$T(x) = \begin{pmatrix} \text{Reject } \beta = \beta_0 & \text{if} & \text{Pretest rejects and } W_{\text{TSLS}}(x) > 3.84 \\ \\ \text{Reject } \beta = \beta_0 & \text{if} & \text{Pretest does not reject and } W_{\text{OLS}}(x) > 3.84 \end{pmatrix}$$

does not have asymptotic size α in the uniform sense. The argument goes as follows: if the correlation between the second-stage structural error and the first-stage error is very small (local to zero), then the Hausman test will not reject the null hypothesis of exogeneity (with high probability). Therefore, the pretest will be followed by a Wald OLS statistic to test $\beta = \beta_0$. The problem is that for small values of the correlation parameter, the size of the Wald can be larger than 5%

Although it is not surprising that a pretest does not have the right size in the uniform sense (think about finite-sample size distortion of pretests), Guggenberger (2010a) shows that the problem with the Hausman test is very important: there is a sequence of parameter values (local-to-zero correlations) for which the rejection probability under the null is close to 1.

C.3.2 Uniformity and Tests for Weak Instruments

It is true that a two-stage test that selects between a standard procedure (like the Wald) and a robust procedure (like the Anderson and Rubin test) following a test for weak instruments will in general lack uniformity. That is, a nominal α-level test for weak instruments followed by a nominal α-level test in a second-stage need not deliver an overall α-level test under a weak instrument sequence. However, the size distortion need not be large.

To illustrate this point consider the following just-identified model with arbitrary heteroskedasticity, serial correlation and/or clustering. Suppose we test the null hypothesis:

$$\mathbf{H}_0 : \beta = \beta_0 \quad vs. \quad \mathbf{H}_0 : \beta \neq \beta_0$$

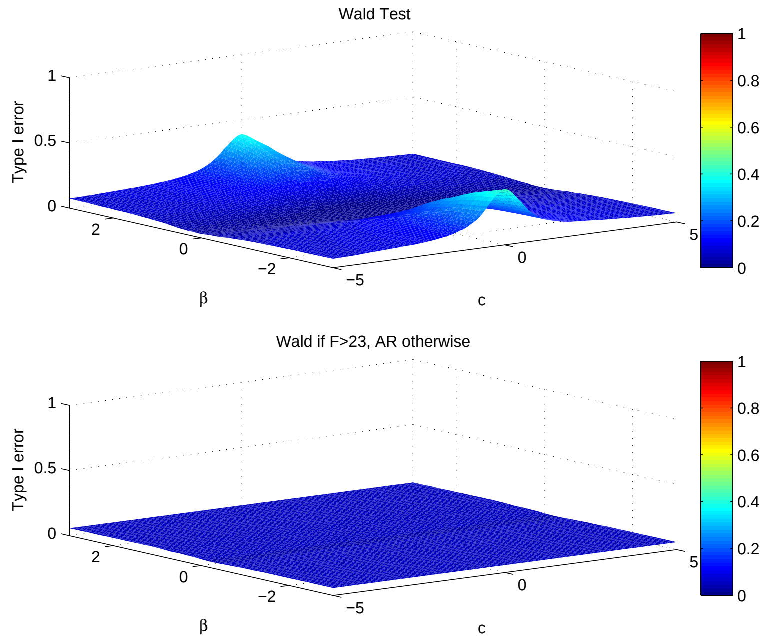
Consider the test that follows the pretest that uses the critical value of 23:

$$T(\sqrt{S}\widehat{\beta\pi}, \sqrt{S}\pi) = \begin{pmatrix} \text{Reject } \beta = \beta_0 \quad \text{if} \quad \widehat{F}_{eff} < 23 \text{ and Anderson-Rubin } AR_S(\beta_0) > 3.84 \\ \\ \text{Reject } \beta = \beta_0 \quad \text{if} \quad \widehat{F}_{eff} > 23 \text{ and Wald } W_S(\beta_0) > 3.84 \end{pmatrix}$$

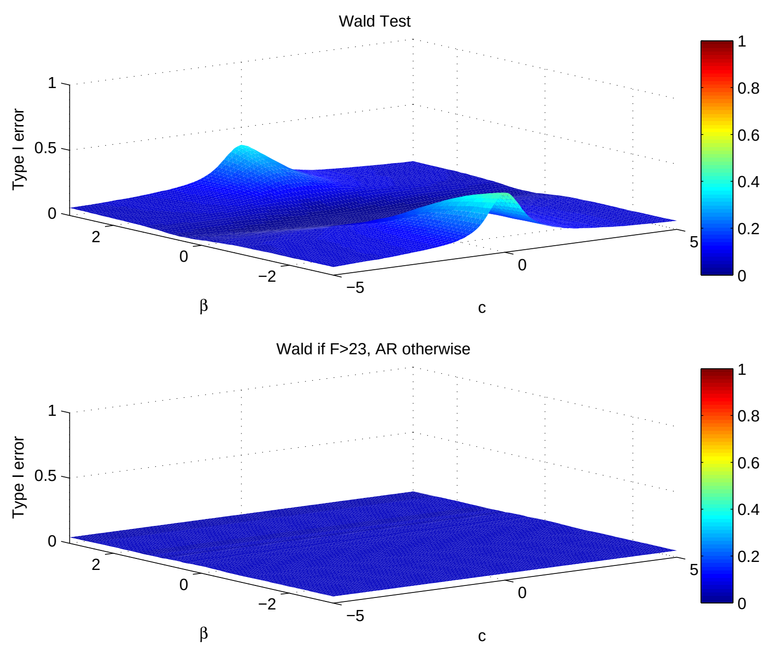
We report the size of the two-stage test T in a Monte-Carlo simulation. We consider a covariance matrix $\mathbf{\Omega}$ with unit variance and correlation parameter ρ. We report the size of the test T for values $\rho = .2, .4, .6, .8$ and uniform grids for $c \in [-5, 5]$ and a $\beta \in [-3, 3]$.

The number of Monte-Carlo draws is 5,000. The simulation shows that the test for weak instruments has a size that is close to α, at least in the region of the parameter space under consideration. The size only gets close to 10% in the last figure. For the parameter values considered, uniformity problems are therefore not a first order concern.

Figure C.3.3.1: Rejection Probabilities for the Wald and T Tests ($\alpha = 5\%$)

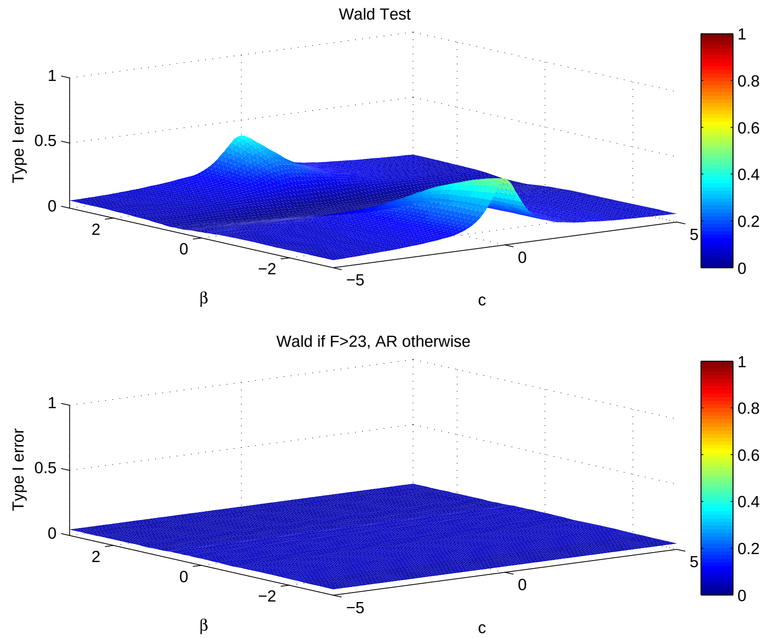


(a) $\rho = .2$

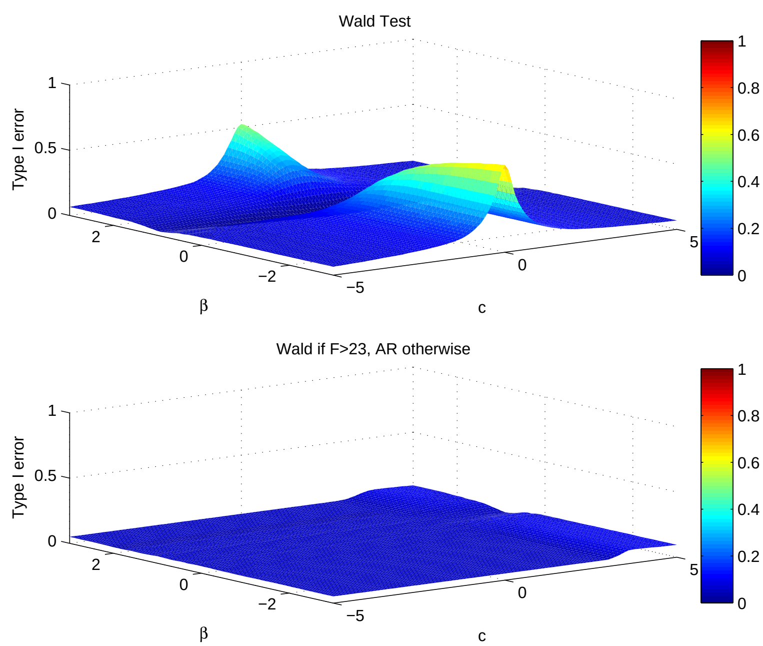


(b) $\rho = .4$

Figure C.3.3.1: Rejection Probabilities for the Wald and T Tests ($\alpha = 5\%$), ctd.



(a) $\rho = .6$



(b) $\rho = .8$

References

Abramowitz, M., and Stegun, I. (1964), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Vol. 55 Dover.

Bock, M., Judge, G., and Yancey, T. (1984), "A Simple Form For the Inverse Moments of Non-Central χ^2 and F Random Variables and Certain Confluent Hypergeometric Functions," *Journal of Econometrics*, 25(217-234), 217–234.

Guggenberger, P. (2010*a*), "The impact of a Hausman pretest on the asymptotic size of a hypothesis test," *Econometric Theory*, 26, 369–382.

Guggenberger, P. (2010*b*), "The impact of a Hausman pretest on the size of a hypothesis test: The panel data case," *Journal of Econometrics*, 156, 337–343.

Imhof, J. (1961), "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48(3-4), 419.

Patnaik, P. (1949), "The Non-Central Chi-Square and F-Distributions and Their Applications," *Biometrika*, 36(1/2), 202.

Ruud, P. (2000), *An Introduction to Classical Econometric Theory* Oxford University Press New York.

Sargan, J. (1974), "The validity of Nagar's expansion for the moments of econometric estimators," *Econometrica*, 42(1), 169–176.

Stock, J., and Yogo, M. (2005), "Testing for weak instruments in linear IV regression, Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg," *ed. D.W. Andrews and J.H. Stock*, pp. 80–108.

White, H. (1980), "Nonlinear Regression on Cross-Section Data," *Econometrica*, 48(4), 721–746.

White, H. (2001), *Asymptotic Theory for Econometricians* Academic Press New York.

Wooldridge, J. (2002), *Econometric Analysis of Cross Section and Panel Data* The MIT Press.