Supplementary Materials for Weak Instruments in IV Regression: Theory and Practice

Isaiah Andrews, James Stock, and Liyang Sun

August 2, 2018

This supplement accompanies the review article "Weak Instruments in IV Regression: Theory and Practice" by Isaiah Andrews, James Stock, and Liyang Sun. Section A describes the collection of *American Economic Review* (AER) articles and specifications for tabulations reported in the review article. Section B discusses the calibration of simulations to these results, with an emphasis on estimation of variance-covariance matrix for the reduced-form and first-stage estimates from our collected linear instrumental variables (IV) specifications. Section (C) provides details on size simulations in calibrations to AER specifications. Section (D) presents the results underlying the discussion of AR confidence sets in Section 5.1 of the review article. Section E overviews available Stata implementations of the weak-IV robust procedures discussed in the main text.

A Publication Selection Criterion

To examine the practical relevance of weak instrument issues, we select recent publications in the *American Economic Review* (AER). We first find articles published in the AER between January 2014 to June 2018 with the keyword "instrument" in their abstract, excluding those which did not relate to instrumental variables. We exclude articles published in the May issue. There are 22 articles that meet such criteria. We then exclude five articles that do not estimate linear instrumental variables (IV) model. Table 1 lists the resulting 17 articles. From these 17 articles, we collect all IV specifications reported in their main text (12 IV specifications are excluded because only first-stage and reduced-form estimates are reported). This yields a total of 230 specifications, which is the "AER sample" in the main text.

For each specification, we record the number of endogenous regressors p and the number of instruments k. In Table 2, we report summary statistics on p and k for these specifications. In one article only, all specifications have multiple endogenous regressors ($p > 1$). While two other articles contain a few specifications with multiple endogenous regressors, 211 of the 230 specifications collected have a single endogenous regressor ($p = 1$), which are the focus of our discussion. Among

specifications with a single endogenous regressor, there are 101 just-identified specifications ($p = k = 1$), found in 12 articles.

For simulations, we rely on the subset of 8 articles and 124 specifications for which we could obtain a full and non-singular variance-covariance matrix for the reduced-form and first-stage estimates, either from the published results or from posted replication data and code. All specifications used for simulation happen to have only a single endogenous regressor ($p = 1$). For details on our calibration of simulations to these data, see Section B.

For our tabulations on reported first-stage F-statistics, we rely on the subset of 14 articles and 108 specifications that report F-statistics and have a single endogenous regressor. Among these specifications, 56 specifications are just-identified, found in 10 articles.

We recognize that not all specifications are important. To distinguish specifications that are most important in each article, we code specifications as "main specifications" for each paper based on the following criteria:

1. The specification is in the first table of IV estimates.

2. If in this table the outcome and endogenous variable(s) are the same across specifications i.e. the only difference is sample restriction / control variables, then we code the last specification as "main specification". If in this table outcome and endogenous variable(s) differ across specifications, restrict specifications to those with outcome and endogenous variables referenced in the abstract or introduction of the article, then code the last specification in this restricted subset as "main specification".

3. If 1) and 2) do not give a unique specification, then we code all specifications at the end of 2) as "main specifications."

We do not use the "main specification" variable in our analysis in the main text, but report this variable in our replication files in case it is of interest to other researchers.

B Details on Simulation Calibration

Consider the IV model

$$\begin{aligned} Y_i &= X_i\beta + W_i'\kappa + \varepsilon_i, \quad (1) \\ X_i &= Z_i'\pi + W_i'\gamma + V_i, \quad (2) \end{aligned}$$

for a scalar outcome Y_i, a scalar endogenous regressors X_i, a $k \times 1$ vector of instrumental variables Z_i, and an $r \times 1$ vector of exogenous regressors W_i. A linear transformation of (1) and (2) is

$$\begin{aligned} Y_i &= Z_i'\delta + W_i'\tau + U_i, \quad (3) \\ X_i &= Z_i'\pi + W_i'\gamma + V_i, \end{aligned}$$

with $\delta = \pi\beta$.

Define $(\hat{\delta}, \hat{\pi})$ as the coefficients on Z_i from the reduced-form and first-stage regressions of Y_i and X_i, respectively, on (Z_i, W_i). By the Frisch-Waugh theorem these are the same as the coefficients from regressing Y_i and X_i on $Z_i^{\perp}$, the part of Z_i orthothogonal to W_i. Under mild regularity conditions (and, in the time-series case, stationarity), $(\hat{\delta}, \hat{\pi})$ are consistent and asymptotically normal in the sense that

$$\sqrt{n}\begin{pmatrix} \hat{\delta} - \delta \\ \hat{\pi} - \pi \end{pmatrix} \to_d N(0, \Sigma^*) \tag{4}$$

for

$$\Sigma^* = \begin{pmatrix} \Sigma^*_{\delta\delta} & \Sigma^*_{\delta\pi} \\ \Sigma^*_{\pi\delta} & \Sigma^*_{\pi\pi} \end{pmatrix} = \begin{pmatrix} Q_{Z^\perp Z^\perp} & 0 \\ 0 & Q_{Z^\perp Z^\perp} \end{pmatrix}^{-1} \Lambda^* \begin{pmatrix} Q_{Z^\perp Z^\perp} & 0 \\ 0 & Q_{Z^\perp Z^\perp} \end{pmatrix}^{-1}$$

where $Q_{Z^\perp Z^\perp} = E[Z_i^\perp Z_i^{\perp\prime}]$ and

$$\Lambda^* = \lim_{n\to\infty} Var\left(\left(\frac{1}{\sqrt{n}}\sum_i U_i Z_i^{\perp\prime}, \frac{1}{\sqrt{n}}\sum_i V_i Z_i^{\perp\prime}\right)'\right).$$

Under standard assumptions, the sample-analog estimator $\hat{Q}_{Z^\perp Z^\perp} = \frac{1}{n}\sum Z_i^\perp Z_i^{\perp\prime}$ will be consistent for $Q_{Z^\perp Z^\perp}$, and we can construct consistent estimators $\hat{\Lambda}^*$ for Λ^* depending on the assumptions imposed on the data generating process (for example whether we allow heteroskedasticity, clustering, or time-series dependence). One can then form consistent estimators for the asymptotic variance matrix Σ^*.

These results imply that for the two stage least square estimator,

$$\hat{\beta}_{2SLS} = \left(\hat{\pi}'\hat{Q}_{Z^\perp Z^\perp}\hat{\pi}\right)^{-1}\hat{\pi}'\hat{Q}_{Z^\perp Z^\perp}\hat{\delta},$$

assuming that IV model is correctly specified and the instruments are strong, so $\delta = \pi\beta$ and π takes a fixed non-zero value,

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \to_d N(0, \Sigma^*_{\beta,2SLS})$$

for

$$\Sigma^*_{\beta,2SLS} = \frac{\pi' Q_{Z^\perp Z^\perp}(\Sigma^*_{\delta\delta} - \beta(\Sigma^*_{\delta\pi} + \Sigma^*_{\pi\delta}) + \beta^2\Sigma^*_{\pi\pi})Q_{Z^\perp Z^\perp}\pi}{(\pi' Q_{Z^\perp Z^\perp}\pi)^2}, \tag{5}$$

which simplifies to $\frac{1}{\pi^2}\Sigma^*_{\delta\delta} - 2\frac{\delta}{\pi^3}\Sigma^*_{\delta\pi} + \frac{\delta^2}{\pi^4}\Sigma^*_{\pi\pi}$ for $p = k = 1$. If we plug in consistent estimators for all the terms in the expression we obtain a consistent estimator $\hat{\Sigma}^*_{\beta,2SLS}$ for $\Sigma^*_{\beta,2SLS}$.

For reasons discussed in the main text, motivated by the asymptotic approximation (4), we consider the case where the reduced-form and first-stage regression coefficients are jointly normal

$$\begin{pmatrix} \hat{\delta} \\ \hat{\pi} \end{pmatrix} \sim N\left(\begin{pmatrix} \delta \\ \pi \end{pmatrix}, \Sigma\right) \tag{6}$$

for

$$\Sigma = \begin{pmatrix} \Sigma_{\delta\delta} & \Sigma_{\delta\pi} \\ \Sigma_{\pi\delta} & \Sigma_{\pi\pi} \end{pmatrix} = \frac{1}{n}\begin{pmatrix} \Sigma^*_{\delta\delta} & \Sigma^*_{\delta\pi} \\ \Sigma^*_{\pi\delta} & \Sigma^*_{\pi\pi} \end{pmatrix}.$$

The estimated variance matrix for $(\hat{\delta}, \hat{\pi})$ is thus $\hat{\Sigma} = \frac{1}{n}\hat{\Sigma}^*$. For our simulation exercise, we calibrate the normal model (6) to IV specifications in the AER sample, with π set to the estimate $\hat{\pi}$ in the data, δ set to $\hat{\pi}\hat{\beta}_{2SLS}$, and Σ set to the estimated variance matrix for $(\hat{\delta}, \hat{\pi})$ under the same assumptions used by the original authors. Most articles report estimates $(\hat{\delta}, \hat{\pi})$ and their standard error, whose squared terms are variance estimates of $\hat{\delta}$ and $\hat{\pi}$. However, none of the articles reports the covariance estimate between $\hat{\delta}$ and $\hat{\pi}$. So we replicate the covariance estimate for as many specifications as possible, as well as any parameter estimate that is not reported in the article.

B.1 Calibration by Direct Calculation

With replication data and code, we can directly estimate Σ by jointly estimating $(\hat{\delta}, \hat{\pi})$ in a seemingly unrelated regression under the same assumptions used by the original authors. Appropriate estimates $\hat{\Sigma}$ are then generated automatically by standard statistical software e.g. `suest` or `avar` in Stata.[1] We are able to obtain $(\hat{\delta}, \hat{\pi})$ and its estimated variance matrix this way for 116 specifications from 7 articles. We exclude one specification from simulation because $\hat{\Sigma}$ is singular. For overidentified specifications, we also obtain $\hat{Q}_{Z^\perp Z^\perp}$.

B.2 Calibration Based on 2SLS Standard Error

For specifications with $p = k = 1$, the 2SLS squared standard error is estimated as

$$\hat{\Sigma}_{\beta,2SLS} = \frac{1}{\hat{\pi}^2}\hat{\Sigma}_{\delta\delta} - 2\frac{\hat{\delta}}{\hat{\pi}^3}\hat{\Sigma}_{\delta\pi} + \frac{\hat{\delta}^2}{\hat{\pi}^4}\hat{\Sigma}_{\pi\pi}.$$

Thus, we can solve for $\hat{\Sigma}_{\delta\pi}$ using $\hat{\Sigma}_{\beta,2SLS}$, $\hat{\delta}$, $\hat{\pi}$, $\hat{\Sigma}_{\delta\delta}$ and $\hat{\Sigma}_{\pi\pi}$. We are able to obtain $(\hat{\delta}, \hat{\pi})$ and its estimated variance matrix for 12 specifications found in Lundborg et al. (2017). We exclude three specifications from simulation because $\hat{\Sigma}$ is singular.

B.3 First-stage F Statistics

To facilitate the discussion on usage of first-stage F-statistics in detecting weak instruments, for each specification we record the following features of each specification:

1. whether first-stage F-statistic is reported;
2. what type of the reported first-stage F-statistics is computed; We can infer this based on labels used by authors in text, or by replicating the reported F-statistics to infer authors' choice of F-statistics. If no explicit discussion on F-statistic is found in text or replication data is not available, we calculate $F = \frac{1}{k}\hat{\pi}'\hat{\Sigma}_{\pi\pi}^{-1}\hat{\pi}$ based on reported first-stage estimates $\hat{\pi}$ and $\hat{\Sigma}_{\pi\pi}$. For an estimate $\hat{\Sigma}_{\pi\pi}$ that does not assume homoskedasticity, this would match with non-homoskedasticity-robust F-statistic F^R. For an estimate $\hat{\Sigma}_{\pi\pi}$ that assumes homoskedasticity,

[1] Some articles report first-stage and reduced-form variances based on Stata's `regress` routine, which applies a degree-of-freedom adjustment to the variance estimate of $\hat{\delta}$ and $\hat{\pi}$ by default. In contrast, Stata's linear IV regression routines do not apply such adjustment by default. For these articles, we replicate the variance matrix estimate for $(\hat{\delta}, \hat{\pi})$ with the degree-of-freedom adjustment for consistency.

this would match with the traditional, non-robust F-statistic F^N. We categorize specifications that do not match with our calculation as "unknown". For specifications specifications that report robust $\hat{\Sigma}_{\pi\pi}$, but computed F^R does not match with reported F-statistic, it could be either that reported F-statistic is F^N or authors use a different degree-of-freedom adjustment in calculating F^R.

3. what label is used by authors in text for their reported first-stage F-statistics; If there is no mention of the type of first-stage F-statistics, we categorize these specifications as "no discussion"

4. whether any weak-instruments robust methods are used.

In Table 3, we summarize the distribution of first-stage F statistics.

As an alternative to the robust and non-robust F-statistics in non-homoskedastic settings, Olea and Pflueger (2013) proposed the effective F statistic. The effective F statistic can be written

$$F^{Eff} = \frac{\hat{\pi}'\hat{Q}_{Z^\perp Z^\perp}\hat{\pi}}{tr\left(\hat{\Sigma}_{\pi\pi}\hat{Q}_{Z^\perp Z^\perp}\right)} \tag{7}$$

The effective F-statistic is equivalent to non-robust F-statistic in homoskedastic settings. The effective F-statistic is equivalent to robust F-statistic in non-homoskedastic settings only when $k = 1$.

Under the normal model (6) with known variance, the average value of the effective F-statistic is $\frac{\pi' Q_{Z^\perp Z^\perp}\pi}{tr\left(\Sigma_{\pi\pi}Q_{Z^\perp Z^\perp}\right)} + 1$.

C Details on Size Simulations

In this section we describe how we calculate sizes of various tests using simulations based on our AER sample. Based on the normal model (6) with known variance, under a given null hypothesis $H_0 : \beta = \beta_0$, we have $g(\beta_0) = \hat{\delta} - \hat{\pi}\beta_0 \sim N(0, \Omega(\beta_0))$ for

$$\Omega(\beta_0) = \Sigma_{\delta\delta} - \beta_0\Sigma_{\delta\pi} - \beta_0\Sigma_{\pi\delta} + \beta_0^2\Sigma_{\pi\pi}.$$

Denote by $\hat{\beta}_{2SLS}$ the 2SLS estimate and $\Sigma^*_{\beta,2SLS}$ its asymptotic variance estimate based on Expression (5). Define the t-statistic as $t(\beta_0) = \sqrt{n}\frac{\hat{\beta}_{2SLS}-\beta_0}{\hat{\Sigma}^{*1/2}_{\beta,2SLS}}$. The size-$\alpha$ t-test of $H_0 : \beta = \beta_0$ is

$$\phi_t(\beta_0) = \mathbf{1}\{t(\beta_0) > c_{1-\alpha}\} \tag{8}$$

where $c_{1-\alpha}$ is the $1-\alpha$ quantile of standard normal distribution. Besides the t-test, we also consider t-test after screening on the first-stage effective F-statistic. In this case, we only evaluate the t-test when $F^{Eff} \geq 10$,

$$\phi_{screen}(\beta_0) = \begin{cases} \phi_t(\beta_0), & F^{Eff} > 10 \\ \text{undefined} & F^{Eff} < 10 \end{cases}. \tag{9}$$

Define the AR statistic as $AR(\beta_0) = g(\beta_0)'\Omega(\beta_0)^{-1}g(\beta_0)$. The size-$\alpha$ AR test of $H_0 : \beta = \beta_0$ is

$$\phi_{AR}(\beta_0) = \mathbf{1}\{AR(\beta_0) > \chi^2_{k,1-\alpha}\} \tag{10}$$

where $\chi^2_{k,1-\alpha}$ is the $1-\alpha$ quantile of χ^2_k distribution. Lastly, we consider the two-step test

$$\phi_{twostep}(\beta_0) = \begin{cases} \phi_t(\beta_0), & F^{Eff} > 10 \\ \phi_{AR}(\beta_0), & F^{Eff} < 10 \end{cases}. \tag{11}$$

In other results (not reported, but available in replication files) we also considered cutoffs based on the Olea and Pflueger (2013) critical values.

We are also interested in testing correct specification $H_0 : \delta - \pi\beta_0 = 0$. Define the J-statistic as $J = g(\hat{\beta}_{2SGMM})'\Omega(\hat{\beta}_{2SGMM})^{-1}g(\hat{\beta}_{2SGMM})$ where $\hat{\beta}_{2SGMM} = \left(\hat{\pi}'\Omega(\hat{\beta}_{2SLS})^{-1}\hat{\pi}\right)^{-1}\hat{\pi}'\Omega(\hat{\beta}_{2SLS})^{-1}\hat{\delta}$ is the efficient two-step GMM estimator, with first-step estimator being the 2SLS estimator $\hat{\beta}_{2SLS}$. The size-α over-identification test is

$$\phi_{overid} = \mathbf{1}\{J > \chi^2_{k-1,1-\alpha}\} \tag{12}$$

where $\chi^2_{k-1,1-\alpha}$ is the $1-\alpha$ quantile of χ^2_{k-1} distribution.

We calculate the size of tests with respect to $H_0 : \beta = \beta_0$ for all 124 specifications that we replicate based on simulations as described in Section C.1 and C.2. We report these results in the main text. For over-identification test, we focus on 90 out of these 124 specifications that are over-identified. We calculate the size based on simulations calibrated to AER sample as described in Section C.1 and report the results in Figure 1 and 2. These 90 specifications are collected from three articles.

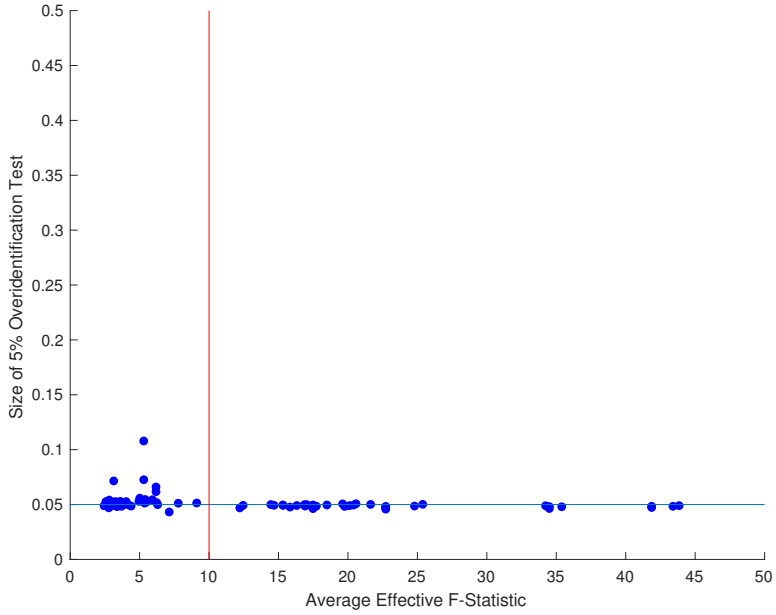

Figure 1: Rejection probability for nominal 5% over-identification test, plotted against average first-stage effective F-statistic in calibrations to AER sample. Limited to the 81 out of 90 over-identified specifications with average F smaller than 50.

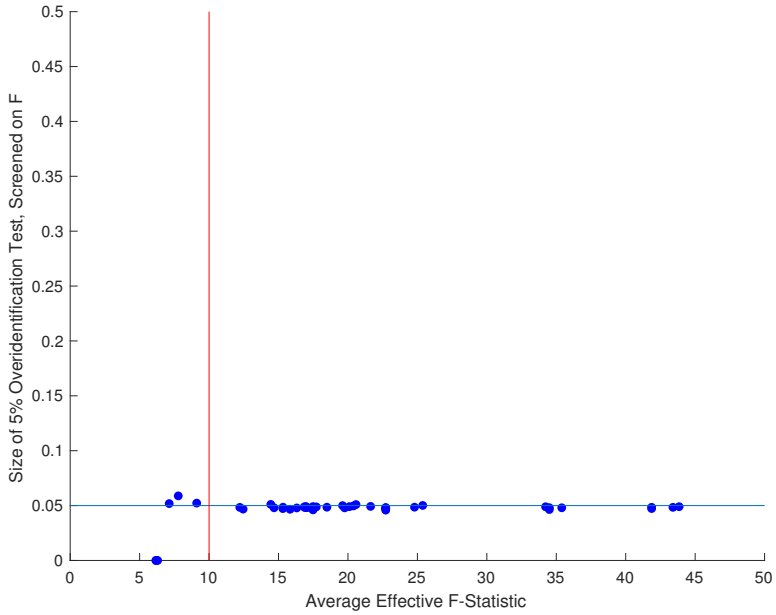

Figure 2: Rejection probability for nominal 5% over-identification test after screening on the first-stage effective F-statistic, plotted against average first-stage effective F-statistic in calibrations to AER sample. Limited to the 81 out of 90 over-identified specifications with average F smaller than 50.

C.1 Simulations calibrated to AER sample

In simulation runs $s = 1, \ldots, S$ for $S = 10000$, we draw $(\hat{\delta}^*, \hat{\pi}^*)$ from the normal model (6) calibrated to IV specifications from the AER sample, with π set to the estimate $\hat{\pi}$ in the data, δ set to $\hat{\pi}\hat{\beta}_{2SLS}$, and Σ set to the estimated variance matrix for $(\hat{\delta}, \hat{\pi})$ under the same assumptions used by the original authors. The null is thus set to $\beta_0 = \hat{\beta}_{2SLS}$. We calculate the t-statistic $t^*(\beta_0)$ based on the 2SLS estimate at each draw $\hat{\beta}^*_{2SLS} = \left(\hat{\pi}^{*'}\hat{Q}_{Z^\perp Z^\perp}\hat{\pi}^*\right)^{-1}\hat{\pi}^{*'}\hat{Q}_{Z^\perp Z^\perp}\hat{\delta}^*$ and its asymptotic variance estimate $\hat{\Sigma}^*_{\beta,2SLS}$. We calculate the AR statistic $AR^*(\beta_0)$ based on the moment function at each draw $g(\beta_0)^*$ and its variance $\Omega(\beta_0)$. We calculate the J-statistic J^* based on $g(\hat{\beta}^*_{2SGMM})^*$ and $\Omega(\hat{\beta}^*_{2SGMM})$. The size of each test is calculated as the average of $\phi^*_j(\beta_0)$ across simulation runs. We also record the median of $t^*(\beta_0)$ across simulation runs.

C.2 Bayesian exercise

To account for uncertainty in estimating (β, π), we adopt a Bayesian approach consistent with the normal model (6). Specifically, we calculate the posterior distribution on (δ, π) after observing estimates $(\hat{\delta}, \hat{\pi})$ based on the normal likelihood from (6) with Σ set to the estimated variance matrix for $(\hat{\delta}, \hat{\pi})$ and a flat prior. Then in $d = 1, \ldots, D$ for $D = 1000$, we draw $(\tilde{\beta}, \tilde{\pi})$ based on the posterior distribution

$$N\left(\begin{pmatrix}\hat{\delta}\\ \hat{\pi}\end{pmatrix}, \Sigma\right) \tag{13}$$

for (δ, π) under a flat prior, setting $\beta = \left(\pi' Q_{Z^\perp Z^\perp}\pi\right)^{-1}\pi' Q_{Z^\perp Z^\perp}\delta$. To obtain a posterior distribution on the size of each test, at each posterior draw $(\tilde{\delta}, \tilde{\pi})$ where $\tilde{\delta} = \tilde{\pi}\tilde{\beta}$, we calculate the size of each test by simulations as described above where the null is set to $\beta_0 = \tilde{\beta}$. Note that we set $\tilde{\delta} = \tilde{\pi}\tilde{\beta}$ to ensure that our simulation designs are consistent with the IV model.

D AR Confidence Sets in Applications

In this section, we provide the underlying results for the discussion of AR test in Section 5.1. For $k = p = 1$, the 5% AR confidence set can be calculated analytically by solving the following quadratic inequality

$$\begin{aligned}&(\hat{\delta} - \hat{\pi}\beta_0)^2 \leq \chi^2_{1,1-\alpha}(\hat{\Sigma}_{\delta\delta} - \beta_0\hat{\Sigma}_{\delta\pi} - \beta_0\hat{\Sigma}_{\pi\delta} + \beta_0^2\hat{\Sigma}_{\pi\pi})\\ \Rightarrow\ &(\hat{\pi}^2 - \chi^2_{1,1-\alpha}\hat{\Sigma}_{\pi\pi})\beta_0^2 + (2\chi^2_{1,1-\alpha}\hat{\Sigma}_{\delta\pi} - 2\hat{\delta}\hat{\pi})\beta_0 + \hat{\delta}^2 - \chi^2_{1,1-\alpha}\hat{\Sigma}_{\delta} \leq 0\end{aligned} \tag{14}$$

where $\chi^2_{1,1-\alpha}$ is the $1-\alpha$ quantile of χ^2_1 distribution. If $(\hat{\pi}^2 - \chi^2_{1,1-\alpha}\hat{\Sigma}_{\pi}) < 0$, the confidence set is unbounded, which happens when we cannot reject $\pi = 0$ based on a 5% t-test.

In Table 4, we list the AR confidence set for 34 just-identified specifications out of 124 specifications that we replicate.

E Weak-IV Robust Procedures in Stata

In replicating IV specifications in our AER sample, we noticed that the `ivreg2` suite, described in Baum et al. (2007), remains a common toolkit for linear IV estimation. `ivreg2` implements the Stock and Yogo (2005) weak instrument test and reports confidence sets for coefficients on endogenous regressors based on the t-test.

As discussed in the main text, the Stock and Yogo (2005) weak instrument test is only valid in the homoskedastic case. The weak instrument test of Olea and Pflueger (2013) is robust to heteroskedasticity, autocorrelation, and clustering. It is thus the preferred test for detecting weak instruments in the over-identified, non-homoskedastic setting. A recent Stata package `weakivtest` by Pflueger and Wang (2015) implements this test. It computes the effective F-statistic and tabulates critical values based on Olea and Pflueger (2013).

The weak instrument test of Olea and Pflueger (2013) concerns only the Nagar (1959) bias approximation, not size distortions in conventional inference procedures (t-tests), though as discussed in the text, in the $k = 1$ case one can use the Olea and Pflueger (2013) effective F-statistic, or the Kleibergen-Paap statistic reported by `ivreg2`, along with the Stock and Yogo (2005) critical values to test for size distortions. Our review paper discusses several tests robust to weak instruments and explains how to construct a level $1-\alpha$ confidence set based on test inversion. In just-identified models, the AR test is efficient and thus recommended. In over-identified models with a single endogenous regressor and homoskedastic errors, the CLR test has good properties. Except for the AR test and the Kleibergen score test, these robust tests require simulations to calculate their critical values in many cases, which can be computationally costly. Below we describe several recent Stata packages that augment `ivreg2` in terms of inference with weak instruments.[2]

For the $p = 1$ and homoskedastic case, the Stata package `condivreg` by Mikusheva and Poi (2006) computes the AR, Kleibergen score, and CLR confidence sets. This routine implements algorithms proposed by Mikusheva (2010) that allow one to construct confidence sets by quickly and accurately inverting these tests without having to use grid search.

For the non-homoskedastic and $p \geq 1$ case, the Stata package `weakiv` computes the AR, Kleibergen score, and CLR confidence sets based on grid search. Finlay and Magnusson (2009) describe a previous version of this package called `rivtest`. Together with simulations for critical values of the CLR test, the grid search can be computationally demanding.

As an alternative to a two-step confidence set based on first-stage F-statistic, Andrews (2018) propose a two-step weak-instruments-robust confidence set. The Stata package `twostepweakiv` by Sun (forthcoming) computes such confidence sets based on grid search. While `twostepweakiv` does not need to simulate critical values for most cases, the grid search alone can be computationally demanding.

For the $p > 1$ cases, one may be interested in subvector inference. `weakiv` implements the traditional projection method and `twostepweakiv` implements the refined projection method based on Chaudhuri and Zivot (2011).

To summarize, in homoskedastic settings, `ivreg2` conducts valid weak instrument tests and

[2] `ivreg2` performs the Anderson-Rubin (AR) test for the null $H_0 : \beta = 0$, but does not calculate a confidence set.

`condivreg` calculates weak-instrument-robust confidence sets. In non-homoskedastic settings, `weakivtest` conducts valid weak instrument tests, but does not guarantee valid inference. For $k = 1$ one can use the effective F-statistic reported by `weakivtest`, or the Kleibergen-Paap statistic reported by `ivreg2`, along with the Stock and Yogo (2005) critical values to test for size distortions. To construct robust confidence sets, `weakiv` or `twostepweakiv` should be used.[3]

References

Andrews, Isaiah, "Valid Two-Step Identification-Robust Confidence Sets for GMM," *Review of Economics and Statistics*, 2018, *100* (2), 337–348.

Baum, Christopher F, Mark E. Schaffer, and Steven Stillman, "Enhanced routines for instrumental variables/generalized method of moments estimation and testing," *Stata Journal*, December 2007, *7* (4), 465–506.

Chaudhuri, Saraswata and Eric Zivot, "A new method of projection-based inference in GMM with weakly identified nuisance parameters," *Journal of Econometrics*, 2011, *164* (2), 239–251.

Finlay, Keith and Leandro Magnusson, "Implementing weak-instrument robust tests for a general class of instrumental-variables models," *Stata Journal*, 2009, *9* (3), 398–421.

Lundborg, Petter, Erik Plug, and Astrid Würtz Rasmussen, "Can Women Have Children and a Career? IV Evidence from IVF Treatments," *American Economic Review*, June 2017, *107* (6), 1611–1637.

Mikusheva, Anna, "Robust confidence sets in the presence of weak instruments," *Journal of Econometrics*, 2010, *157* (2), 236 – 247.

_ **and Brian P. Poi**, "Tests and confidence sets with correct size when instruments are potentially weak," *Stata Journal*, September 2006, *6* (3), 335–347.

Nagar, A. L., "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 1959, *27* (4), 575–595.

Olea, José Luis Montiel and Carolin Pflueger, "A Robust Test for Weak Instruments," *Journal of Business & Economic Statistics*, July 2013, *31* (3), 358–369.

Pflueger, Carolin E. and Su Wang, "A robust test for weak instruments in Stata," *Stata Journal*, March 2015, *15* (1), 216–225.

Stock, James H. and Motohiro Yogo, "Testing for Weak Instruments in Linear IV Regression," in Donald W. K. Andrews and James H. Stock, eds., *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge University Press, 2005, pp. 80–108.

Sun, Liyang, "Implementing valid two-step identification-robust confidence sets for linear instrumental-variables models," *Stata Journal*, forthcoming.

[3]Unfortunately, `weakivtest` and `twostepweakiv` are only compatible with the syntax of `ivreg2`. `weakiv` is also compatible with the syntax of other State IV model packages including `xtivreg2`, which accommodates fixed effects.

Table 2: Summary statistics

(a) unweighted

p	# specifications	% just-identified	% over-identified	Avg. k | over-identified
1	211	48%	52%	21.95
2	5	40%	60%	48.67
3	2	0%	100%	58
4	12	0%	100%	8

(b) weighted by the inverse of the number of specifications in each article

p	# articles	% just-identified	% over-identified	Avg. k | over-identified
1	15.63	74.23%	25.78%	23.01
2	0.19	26.19%	73.80%	72.68
3	0.18	0%	100%	58
4	1	0%	100%	8

Notes: In panel (a), we tabulate the distribution of p and k by specifications. The sample consists of 230 specifications. In panel (b), we tabulate the distribution of p and k by specifications, weighted by the inverse of the number of specifications in each article so that each article receives the same weight.

Table 1: Selected publications

Article ID	Issue	Title	# of specifications	Replication files available
1	104 (1)	Immigration and the Diffusion of Technology: The Huguenot Diaspora in Prussia	11	1
2	104 (10)	German Jewish Emigres and US Invention	10	1
3	104 (11)	Structural Transformation, the Mismeasurement of Productivity Growth, and the Cost Disease of Services	40	1
4	104 (6)	Compulsory Education and the Benefits of Schooling	30	1
5	104 (6)	The Wage Effects of Offshoring: Evidence from Danish Matched Worker-Firm Data	12	0
6	104 (7)	Mafia and Public Spending: Evidence on the Fiscal Multiplier from a Quasi-experiment	20	1
7	105 (12)	Media Influences on Social Outcomes: The Impact of MTV's 16 and Pregnant on Teen Childbearing	8	0
8	105 (2)	The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South	6	0
9	105 (3)	Credit Supply and the Price of Housing	3	1
10	105 (4)	Medicare Part D: Are Insurers Gaming the Low Income Subsidy Design	6	1
11	106 (3)	How Do Electricity Shortages Affect Industry? Evidence from India	8	0
12	106 (6)	Charters without Lotteries: Testing Takeovers in New Orleans and Boston	40	0
13	107 (2)	The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit: Comment	4	0
14	107 (6)	Can Women Have Children and a Career? IV Evidence from IVF Treatments	12	0
15	107 (9)	Virtual classrooms: how online college courses affect student success	4	0
16	108 (2)	Export Destinations and Input Prices	11	0
17	108 (3)	Disentangling the effects of a banking crisis: evidence from German firms and counties	5	0

Notes: we record whether replications files are available to replicate all specifications contained in an article. For a complete list of specifications, please see replication files.

Table 3: Summary statistics on F statistics for specifications with $p = 1$

(a) unweighted; level of observation is specification

k	# spec	% report F	Avg. F	% F $>$ 10	% F $>$ SY cutoffs	SY cutoffs	% use robust tests	Avg. F
1	101	55.44%	3483.23	89.29%	92.86%	8.96	0%	
2	28	78.57%	33.75	86.36%	68.18%	11.59	0%	
3	30	60%	49.27	88.89%	88.89%	12.83	100%	49.27
16	4	100%	3.96	0%	0%	27.99	100%	3.96
21	4	0%				33.97	0%	
23	4	100%	57.1	100%	100%	36.37	0%	
28	16	0%				42.37	0%	
59	20	0%				44.78	0%	
100	4	100%	2.88	0%	0%	44.78	100%	2.88
total	211	51.18%	1823.572	82.41%			18%	35.16

(b) weighted by the inverse of the number of specifications in each article

k	# articles	% report F	Avg. F	% F $>$ 10	% F $>$ SY cutoffs	SY cutoffs	% use robust tests	Avg. F
1	11.6	64.99%	2228.85	89.31%	91.96%	8.96	0%	
2	1.2	87.5%	24.02	85.71%	66.67%	11.59	0%	
3	1	60%	49.27	88.89%	88.89%	12.83	100%	49.27
16	0.36	100%	3.96	0%	0%	27.99	100%	3.96
21	0.1	0%				33.97	0%	
23	0.1	100%	57.1	100%	100%	36.37	0%	
28	0.4	0%				42.37	0%	
59	0.5	0%				44.78	0%	
100	0.36	100%	2.88	0%	0%	44.78	100%	2.88
total	15.63	64.09%	1683.87	82.53%			11.06%	24.15

Notes: Column (3), (4), (5) are conditional on reporting F statistics. Column (8) is conditional on using robust tests. There are four specifications report F statistics being $>$614. We take them to be 614. The SY cutoffs ensure the maximal size is 15% of a 5% Wald test. Tabulation is only available for $k \leq 30$. For $k > 30$, we use the cutoff for $k = 30$, which is 44.78. Among the 16 articles with any specifications with $p = 1$, there are 14 articles that report some F statistics, one reports p-value and the other none.

Table 4: AR confidence sets for just-identified specifications

Article ID	Table reference	Reported F statistic	Constructed F statistic	IV estimate	IV standard error	5% AR confidence set
14	Table 3 Panel A (1)	38427	30102.25	-70088	2054	[-74104.73 , -66052.54]
14	Table 3 Panel A (2)	38427	30102.25	-0.07	0.01	[-0.08 , -0.06]
14	Table 3 Panel A (3)	38427	30102.25	-5.91	0.19	[-6.19 , -5.44]
14	Table 3 Panel B (1)	6281	6400	-29378	5285	[-39726.2 , -19002.77]
14	Table 3 Panel B (2)	6281	6400	-0.04	0.01	[-0.06 , -0.02]
14	Table 3 Panel B (3)	6281	6400	1.47	0.36	[0.79 , 2.18]
14	Table 3 Panel B (4)	6281	6400	-26.85	4.45	[-35.87 , -18.41]
14	Table 3 Panel C (1)	2273	2061.16	-30675	10546	[-51361.01 , -9982.12]
14	Table 3 Panel C (2)	2273	2061.16	-0.02	0.02	[-0.06 , 0.03]
10	Table 8 (1)	33.91	10.5	0.54	0.24	[0.22 , 1.58]
10	Table 8 (2)	26.68	7.99	0.62	0.32	[0.05 , 1.98]
10	Table 8 (3)	46.35	16.21	0.69	0.24	[0.33 , 1.5]
10	Table 8 (4)	43.38	7.99	0.63	0.34	[0.12 , 2.26]
10	Table 8 (5)	44.28	16.75	0.75	0.25	[0.38 , 1.62]
10	Table 8 (6)	40.18	8.17	0.7	0.31	[0.22 , 2.2]
9	Table 5 (1)	10.13	9.97	0.14	0.07	[0.04 , 0.45]
9	Table 5 (2)	10.31	10.31	0.13	0.06	[0.04 , 0.4]
9	Table 5 (3)	8.59	8.6	0.12	0.06	[0.04 , 0.41]
2	Table 4 (1)	20.8	20.81	218.71	60.61	[107.4 , 373.95]
2	Table 4 (2)	18.25	18.25	170.14	57.99	[63.91 , 324.19]
2	Table 4 (3)	9.79	9.79	25.72	8.75	[14.07 , 67.3]
2	Table 4 (4)	8.99	8.99	17.14	6.91	[7.18 , 49.31]
2	Table 6 (4)	21.65	21.85	-0.02	0.01	[-0.04 , -0.01]
1	Table 4 (3)	3.668	3.67	3.48	1.16	(-∞, -58.44]$\cup$[1.55,∞)
1	Table 4 (5)	4.792	4.79	3.38	1.14	[1.64 , 19.16]
1	Table 5 (3)	5.736	5.74	1.67	0.85	[-0.46 , 5.94]
1	Table 5 (6)	15.35	15.35	0.07	0.04	[-0.01 , 0.16]
1	Table 6 Panel C (1)	NA	4.2	1.63	1.01	[-1.33 , 18.2]
1	Table 6 Panel C (2)	NA	5.48	1.11	1.21	[-2.57 , 6.87]
1	Table 6 Panel C (3)	NA	5.8	1.53	0.83	[-0.63 , 5.62]
1	Table 6 Panel C (4)	NA	6.09	1.7	0.75	[-0.34 , 4.84]
1	Table 6 Panel C (5)	NA	6.7	2.07	0.52	[0.95 , 4.52]
1	Table 6 Panel C (6)	NA	2.88	3.08	1.9	(-∞ , -6.14]$\cup$[-2.29 , ∞)
1	Table 6 Panel C (7)	NA	5.76	1.84	0.97	[-0.71 , 6.44]

Notes: Article ID refers to article identification in Table 1. Table reference refers to the table containing IV estimates in the corresponding article. We construct F statistic for each specification by $\pi^{\intercal}(\Sigma_{\pi\pi}/N)^{-1}\pi/k$.