Two-sample Instrumental Variables Regression with Potentially Weak Instruments

Jaerim Choi
University of Hawaii at Manoa
Honolulu, Hawaii, United States
choijm@hawaii.edu

Shu Shen
University of California, Davis
Davis, California, United States
shushen@ucdavis.edu

Abstract. We develop a Stata package for two-sample instrumental variables regression models with one endogenous regressor and potentially weak instruments. The package includes the classic two-sample 2SLS estimator whose inference is only valid under the assumption of strong instruments, as well as statistical tests and confidence sets with correct size and coverage probabilities even when the instruments are weak.

Keywords: st0001, two-sample 2SLS, weak IV, inference

1 Introduction

Conventional instrumental variables (IV) regression requires that the dependent variable, the endogenous regressor, and the instruments come from the same dataset. But in many cases, researchers can only observe the dependent variable and the endogenous regressor in two separate data samples (see Björklund and Jäntti 1997; Miguel 2005; Feldman 2010; Brunner et al. 2012; Siminski 2013; Olivetti and Paserman 2015, among many others). Angrist and Krueger (1992, 1995) propose two estimation strategies, two-sample instrumental variables (TSIV) and two-sample two-stage least squares (TS2SLS) for such two-sample IV regression models. Under the assumption of strong instruments, both TSIV and TS2SLS estimators are consistent. Inoue and Solon (2010) provide valid inference formula for TSIV and TS2SLS under the assumption of strong instruments and show that TS2SLS is more efficient. When the first stage is weak, however, neither estimation strategy is valid following arguments similar to the famous Bound et al. (1995) critiques in the classic (one-sample) two-stage IV literature.

In a recent study, Choi et al. (2018) develop weak-instrument robust inference for the two-sample IV regression model with one single endogenous regressor. This article develops a Stata command companion to this newly proposed inference method. Specifically, the new method extends the classic Anderson-Rubin (AR, c.f. Dufour 1997; Staiger and Stock 1997; Dufour and Jasiak 2001), Kleibergen (K, c.f. Kleibergen 2002), and conditional likelihood ratio (CLR, c.f. Moreira 2003; Andrews et al. 2006, 2008; Moreira 2009) tests and confidence sets to the two-sample setting. The proposed command also reports for completeness the classic TS2SLS estimates and associated standard errors. Both cases of homoskedasticity and heteroskedasticity are considered in the proposed command.

The next section provides background on the two-sample IV regression model. Section 3 discusses the weak-instrument robust inference methods developed in Choi et al. (2018). Section 4 introduces the new Stata command **weaktsiv**. Section 5 gives illustration examples.

2 Model and Background

Let subscript j, $j = 1, 2$, denote random variables in the first and second dataset with sample size n_j. Assume that $n_1/n_2 \to \tau$ for some fixed $\tau > 0$. In this article we consider the following two-sample IV regression model with independent and identically distributed data and a single endogenous regressor:

$$
\begin{aligned}
y_1 &= w_1\beta + X_1\gamma + \epsilon_1, \\
w_j &= Z_j\pi + X_j\psi + \varepsilon_j, \quad j = 1, 2,
\end{aligned} \tag{1}
$$

where y_1, w_1, ϵ_1 and ε_1 are $n_1 \times 1$; w_2, ε_2 are $n_2 \times 1$; Z_j is $n_j \times k$ for $j = 1, 2$; and X_j is $n_j \times p$ for $j = 1, 2$. All variables in the above model are observed except for w_1. Researchers are primarily interested in the parameter β in the outcome equation.

TS2SLS follows the idea of classic two-stages least squares (2SLS) estimation by regressing the outcome variable y_1 on a predicted endogneous regressor, named $\hat{w}_1$. Let $\hat{w}_1 = Z_1\hat{\pi} + X_1\hat{\psi}$. Unlike classic 2SLS, $\hat{\pi}$ and $\hat{\psi}$ in TS2SLS are estimated using information from the second data sample because w_1 is not observed. Specifically, the TS2SLS estimator for β is defined as

$$
\hat{\beta}_{TS2SLS} = (\hat{w}_1' M_{X_1} \hat{w}_1)^{-1} \hat{w}_1' M_{X_1} y_1,
$$

where $\hat{w}_1 = Z_1(Z_2' M_{X_2} Z_2)^{-1} Z_2' M_{X_2} w_2 + X_1(X_2' M_{Z_2} X_2)^{-1} X_2' M_{Z_2} w_2$. For any matrix X, P_X is used to denote $X(X'X)^{-1}X'$ and $M_X = I - P_X$.

Under the assumption that the first stage correlation between the endogenous regressor and instruments is strong, the TS2SLS estimator is consistent and asymptotic normal. Inoue and Solon (2010) provide inference for TS2SLS under the additional assumptions of homoskedasticity and equal moments of $[Z_j \; X_j]$ across the two samples with $j = 1, 2$. In our model with one endogenous regressor, the Inoue and Solon (2010) formula inflates the second stage standard errors (i.e. standard errors from regressing y_1 on $\hat{w}_1$ and X_1) by a factor of $\left[1 + (n_1/n_2)\hat{\beta}^2_{TS2SLS}\left(\hat{\sigma}_{\varepsilon_2}/\hat{\sigma}_{u_1}\right)^2\right]^{1/2}$, where $\hat{\sigma}^2_{u_1} = y_1' M_{[Z_1:X_1]} y_1/(n_1 - k - p)$ and $\hat{\sigma}^2_{\varepsilon_2} = w_2' M_{[Z_2:X_2]} w_2/(n_2 - k - p)$.

The additional assumptions on homoskedasticity and equal moments required in Inoue and Solon (2010) could be restricted in applications. Pacini and Windmeijer (2016) provide TS2SLS inference that is robust to heteroskedasticity and unequal moments of excluded instruments and exogenous regressors, although their results are still not robust to weak instruments. In general, TS2SLS is only valid when the first stage correlation between instruments and the endogenous regressor is strong.

Table 1 illustrates limitations of the TS2SLS strategy. The DGP is taken from Choi et al. (2018) where $Z_{1i} \sim N(0, I_k)$, $(\epsilon_{1i}, e_{1i}) \sim N(0, I_2)$, $\epsilon_{1i} = 0.1\varepsilon_{1i} + \sqrt{1 - 0.1^2} e_{1i}$, $y_{1i} = w_{1i}\beta + \epsilon_{1i}$, $w_{1i} = Z_{1i}\pi + \varepsilon_{1i}$, $Z_{2i} \sim N(0, I_k)$,$\varepsilon_{2i} \sim N(0, 1)$, and $w_{2i} = Z_{2i}\pi + \varepsilon_{2i}$. The sample sizes are $n_1 = 5{,}000$ and $n_2 = 1{,}000$. The number of instruments k is set to 1, 5, or 10. The first stage coefficient vector π is set to $\sqrt{\lambda/(n_2 k)} \cdot \iota$ where ι is a vector of k ones and λ/k is the concentration parameter capturing the strength of instruments. λ/k is set to 1, 4, or 16.

Table 1 reports the coverage rate of the 95% Inoue and Solon (2010) confidence interval among 5,000 simulation repetitions as well as the bias and Root Mean Squared Error (RMSE) of the TS2SLS estimator. The simulation results show that TS2SLS produces large biases and unreliable confidence interval when the instruments are weak. The 95% confidence interval could have coverage rate as low as 13.7% when there are many weak instruments ($k = 10$, $\lambda/k = 1$). The bias is generally negative when $\beta = 2$ and positive when $\beta = -2$ because TS2SLS suffers from a classic attenuation bias which is approximately proportional to the true value of β. The attenuation bias is also inversely proportional to the strength of instruments. See Choi et al. (2018) for more details.

Table 1: Properties of TS2SLS Under Weak Instruments

	Coverage of 95% CI			Bias of Estimator			RMSE of Estimator		
λ/k	1	4	16	1	4	16	1	4	16
(a): $\beta = -2$									
$k = 1$	0.767	0.833	0.806	-4.442	0.297	-0.160	180.971	31.051	0.834
$k = 5$	0.358	0.610	0.715	0.889	0.288	0.076	1.058	0.508	0.249
$k = 10$	0.137	0.453	0.662	0.955	0.348	0.092	1.018	0.443	0.190
(b): $\beta = 0$									
$k = 1$	0.990	0.974	0.956	-0.427	-0.036	-0.001	37.121	3.567	0.128
$k = 5$	0.961	0.954	0.947	0.002	0.002	0.001	0.168	0.097	0.050
$k = 10$	0.961	0.950	0.949	0.004	0.001	-0.000	0.109	0.067	0.035
(c): $\beta = 2$									
$k = 1$	0.788	0.840	0.819	3.588	-0.370	0.159	160.209	30.845	0.832
$k = 5$	0.389	0.628	0.735	-0.885	-0.284	-0.074	1.065	0.517	0.251
$k = 10$	0.165	0.473	0.685	-0.947	-0.346	-0.093	1.018	0.443	0.193

Notes: Sample sizes are $n_1 = 5,000$ and $n_2 = 1,000$. Results are based on 5,000 simulation repetitions. The coverage results of the 95% Inoue and Solon (2010) confidence intevals are also reported in Choi et al. (2018).

3 Weak-instrument Robust Methods

In the following, we introduce the weak-instrument robust inference discussed in Choi et al. (2018). There are two versions of the method. A benchmark strategy is developed under the same set of assumptions as in Angrist and Krueger (1992, 1995) and Inoue and Solon (2010), except for allowing for instruments to be potentially weak. The benchmark strategy requires both homoskedasticity and equal moments of excluded instruments and exogenous regressors across the two data samples. In contrast, a fully robust strategy that makes the two-sample IV inference robust to weak instruments as well as heteroskedasticity and unequal moments is also considered. In empirical applications, researchers might want to adopt the fully robust method for its generality and the benchmark method for a direct comparison with the classic Inoue and Solon (2010) results. Starting from this section, we follow the weak inference literature and assume without loss of generality that $Z_j'X_j = 0$, for both $j = 1, 2$. The orthogonality assumption is without loss of generality because one can always define new excluded instruments as residuals from the regression of original instruments on exogenous regressors.

Consider the weak IV asymptotic where the first stage parameter π is a local sequence converging to zero.

$$\pi = C/\sqrt{n_1} \text{ for some non-stochastic } k\text{-vector } C. \tag{2}$$

Under this asymptotic, the TS2SLS estimator is no longer consistent, and alternative weak-instrument robust method needs to be considered. In applications, researchers shall consider adopting weak-instrument robust inference methods when the instruments are expected to have weak correlations with the endogenous regressor or when the sample size is small.

3.1 Benchmark Weak-instrument Robust Tests and Confidence Sets

Let $Y_1 = [y_1 \;\; \hat{w}_1]$, $a = [\beta \;\; 1]'$, $\eta = [\gamma \;\; \psi]$, and $V_1 = [u_1 \;\; v_1]$, where $\gamma = \gamma_1 + \psi\beta$, $u_1 = \epsilon_1 + \beta\varepsilon_1$, and $v_1 = Z_1(\hat{\pi} - \pi) + X_1(\hat{\psi} - \psi)$. The simultaneous equation model described in the last section could be rewritten as:

$$Y_1 = Z_1 \pi a' + X_1 \eta + V_1.$$

In this section, we follow Inoue and Solon (2010) and assume homoskedasticity as well as equal moments of $[Z_j \; X_j]$ for $j = 1, 2$. Let $\sigma^2_{u_1} = E[u_1^2 | Z_1, X_1]$, $\sigma^2_{\varepsilon_2} = E[\varepsilon_2^2 | Z_2, X_2]$, and Σ_{ZZ} be the probability limit of both $Z_1'Z_1/n_1$ and $Z_2'Z_2/n_2$.

First, consider the two-sided null hypothesis $H_0 : \beta = \beta_0$ with some pre-determined significance level α. Let $b_0 = [1 \;\; -\beta_0]'$, $a_0 = [\beta_0 \;\; 1]'$. Define statistics

$$\hat{S}_n = (Z_1'Z_1)^{-1/2} Z_1' Y_1 b_0 / (b_0' \hat{\Omega} b_0)^{1/2}, \quad \hat{T}_n = (Z_1'Z_1)^{-1/2} Z_1' Y_1 \hat{\Omega}^{-1} a_0 / (a_0' \hat{\Omega}^{-1} a_0)^{1/2},$$

$$\hat{Q}_S = \hat{S}_n' \hat{S}_n, \quad \hat{Q}_T = \hat{T}_n' \hat{T}_n, \quad \hat{Q}_{ST} = \hat{S}_n' \hat{T}_n, \quad \hat{Q} = \begin{pmatrix} \hat{Q}_S & \hat{Q}_{ST} \\ \hat{Q}_{ST} & \hat{Q}_T \end{pmatrix},$$

where $\hat{\Omega} = \begin{pmatrix} \hat{\sigma}^2_{u_1} & 0 \\ 0 & \hat{\sigma}^2_{\varepsilon_2} n_1/n_2. \end{pmatrix}$ with $\hat{\sigma}^2_{u_1} = y_1' M_{[Z_1:X_1]} y_1 / (n_1 - k - p)$ and $\hat{\sigma}^2_{\varepsilon_2} = w_2' M_{[Z_2:X_2]} w_2 / (n_2 - k - p)$ are consistent estimators of $\sigma^2_{u_1}$ and $\sigma^2_{\varepsilon_2}$, respectively.

Further define test statistics

$$\mathcal{T}_1(\beta_0) = \hat{Q}_S, \quad \mathcal{T}_2(\beta_0) = \hat{Q}_{ST}^2 / \hat{Q}_T,$$

$$\mathcal{T}_3(\beta_0) = \frac{1}{2}\left[\hat{Q}_S - \hat{Q}_T + \left[\left(\hat{Q}_S + \hat{Q}_T \right)^2 - 4\left(\hat{Q}_S \hat{Q}_T - \hat{Q}_{ST}^2 \right) \right]^{1/2} \right].$$

Under the weak IV asymptotic in (2) and when the null condition $\beta = \beta_0$ holds, one can show that in the limit $\mathcal{T}_1(\beta_0)$ follows a $\chi^2(k)$ distribution and $\mathcal{T}_2(\beta_0)$ follows a $\chi^2(1)$ distribution. When $k = 1$, both $\mathcal{T}_2(\beta_0)$ and $\mathcal{T}_3(\beta_0)$ reduce to $\mathcal{T}_1(\beta_0)$. When $k \geq 2$, in the limit the probability of $\mathcal{T}_3(\beta_0)$ exceeding m is

$$p(m; q_T) = 1 - 2K \int_0^1 P\left[\chi_k^2 < \frac{q_T + m}{1 + q_T s_2^2 / m} \right] (1 - s_2^2)^{(k-3)/2} ds_2$$

under the null, where $K = \Gamma(k/2) / \left[\pi^{1/2} \Gamma((k-1)/2) \right]$ and χ_k^2 is a random variable following a χ^2 distribution with k degrees of freedom (Andrews et al. 2007).

Let $q_{1-\alpha}(k)$ be the $(1 - \alpha)$ quantile of the $\chi^2(k)$ distribution. Define the decision rules of the three statistics as "reject the null if $\mathcal{T}_1(\beta_0) > q_{1-\alpha}(k)$", "reject the null if $\mathcal{T}_2(\beta_0) > q_{1-\alpha}(1)$", and "reject the null if $\mathcal{T}_3(\beta_0) > q_{1-\alpha}(1)$ when $k = 1$, and reject the null if $p\left(\mathcal{T}_3(\beta_0); \hat{Q}_T \right) < \alpha$ when $k \geq 2$", respectively. All three tests have asymptotic size control under the weak IV asymptotic. When the null is violated, all three tests have nontrivial power dependent on C when the first stage π satisfies (2). When the instruments are strong, all three tests have power approaching one.

We call the test based on $\mathcal{T}_1(\beta_0)$ the TSAR test, the one based on $\mathcal{T}_2(\beta_0)$ the TSK test, and the one based on $\mathcal{T}_3(\beta_0)$ the TSCLR test. Note that when $k = 1$, all three tests give identical results. When $k \geq 2$, TSCLR generally has better power performances than the other two methods, but there are also some data generating processes where TSAR can outperform. See Choi et al. (2018) for details.

Given the proposed tests, the $(1-\alpha) \times 100\%$ confidence sets for β can be obtained by inverting the corresponding tests. Define

$$\mathcal{CI}_1(\alpha) = \{\beta_0 : \mathcal{T}_1(\beta_0) \leq q_{1-\alpha}(k)\}, \quad \mathcal{CI}_2(\alpha) = \{\beta_0 : \mathcal{T}_2(\beta_0) \leq q_{1-\alpha}(1)\},$$
$$\mathcal{CI}_3(\alpha) = \{\beta_0 : p\left(\mathcal{T}_3(\beta_0); \hat{Q}_T(\beta_0)\right) \geq \alpha\}.$$

The confidence sets, since inverted from asymptotic valid tests under the weak IV asymptotics, have correct coverage in the limit. When the instruments are weak, the confidence sets could be unbounded, which is an essential property for confidence sets to have correct coverage with arbitrarily weak instruments (Dufour 1997). The benchmark confidence sets are computed analytically following the fast computation method proposed by Mikusheva and Poi (2006) for the classic (one-sample) AR, K and CLR confidence sets.

Similar to the classic K test, the TSK test also has irregular non-monotonic power curve when $k \geq 2$, resulting in power loss with some data generating processes. For confidence sets, the TSK confidence set can also take the form of a union of two finite intervals, i.e., $[x_1, x_2] \cup [x_3, x_4]$, while the TSAR and TSCLR confidence sets, conditional on boundedness, only take the usual form of finite interval, or $[x_1, x_2]$. Therefore, similar to the classic one-sample case (see, e.g., Mikusheva and Poi (2006)), the TSK method is generally not recommended in practice.

Table 2 is taken from Panel A and B of Table 1 in Choi et al. (2018). It uses the same DGP as the one discussed in Section 2. Compared to the TS2SLS results reported in Table 1, the proposed TSAR, TSCLR, TSK confidence sets have targeted coverage rates regardless of instrument strength. Panel B of Table 2 provides a rough idea about how often the proposed weak-instrument robust confidence sets could be unbounded given various instrument strengths. The panel also shows the excellent power property of TSCLR in general and the irregular power performance of TSK under some DGPs.

3.2 Fully Robust Tests and Confidence Sets

This section relaxes the assumptions of homoskedasticity and equal moments of excluded instruments and exogenous regressors. Let Σ_{z,u_1} and Σ_{z,ε_2} are probability limits of $V[Z_1'u_1/\sqrt{n_1}]$ and $V[Z_2'\varepsilon_2/\sqrt{n_2}]$, respectively, and that $\Sigma_{l,ZZ}$ is the probability limit of $Z_l'Z_l/n_l$ for $l = 1, 2$. Replace the first equation in two-sample IV regression model by its reduced form $y_1 = Z_1\zeta + X_1\gamma + u_1$. Let $\delta = [\zeta \quad \pi]'$, we know that

$$r(\delta, \beta) = \zeta - \pi\beta = 0.$$

Let $\hat{\zeta} = (Z_1'Z_1)^{-1}Z_1'y_1$, $\hat{\pi} = (Z_2'Z_2)^{-1}Z_2'w_2$, and $\hat{\delta} = [\hat{\zeta} \quad \hat{\pi}]'$. It is easy to see that, for any β_0,

$$\sqrt{n_1}\left(r(\hat{\delta}, \beta_0) - r(\delta, \beta_0)\right) \Rightarrow N\left(0, \Sigma_{\beta_0}\right),$$

Table 2: Properties of Benchmark Weak-instrument Robust Confidence Sets

	TSAR			TSCLR			TSK		
λ/k	1	4	16	1	4	16	1	4	16
Panel A: coverage of 95% confidence sets									
(a): $\beta = -2$									
$k = 1$	0.947	0.955	0.947	0.947	0.955	0.947	0.947	0.955	0.947
$k = 5$	0.951	0.950	0.952	0.958	0.950	0.954	0.957	0.949	0.954
$k = 10$	0.948	0.944	0.943	0.946	0.944	0.950	0.947	0.945	0.950
(b): $\beta = 0$									
$k = 1$	0.947	0.946	0.948	0.947	0.946	0.948	0.947	0.946	0.948
$k = 5$	0.947	0.955	0.952	0.950	0.949	0.946	0.953	0.950	0.946
$k = 10$	0.948	0.945	0.949	0.954	0.948	0.949	0.957	0.948	0.948
(c): $\beta = 2$									
$k = 1$	0.946	0.951	0.950	0.946	0.951	0.950	0.946	0.951	0.950
$k = 5$	0.951	0.948	0.950	0.959	0.951	0.954	0.960	0.950	0.953
$k = 10$	0.952	0.945	0.944	0.949	0.949	0.946	0.949	0.949	0.946
Panel B: number of bounded 95% confidence sets among 5,000 simulations									
(a): $\beta = -2$									
$k = 1$	854	2,648	4,901	854	2,648	4,901	854	2,648	4,901
$k = 5$	1,730	4,635	4,872	2,670	4,963	5,000	2,649	4,964	5,000
$k = 10$	2,525	4,818	4,811	4,075	5,000	5,000	4,032	5,000	5,000
(b): $\beta = 0$									
$k = 1$	854	2,648	4,901	854	2,648	4,901	854	2,648	4,901
$k = 5$	1,804	4,682	4,874	1,732	4,712	5,000	843	2,116	3,419
$k = 10$	2,633	4,849	4,838	2,500	4,983	5,000	857	2,025	3,275
(c): $\beta = 2$									
$k = 1$	854	2,648	4,901	854	2,648	4,901	854	2,648	4,901
$k = 5$	1,734	4,639	4,879	2,667	4,960	5,000	2,629	4,962	5,000
$k = 10$	2,546	4,816	4,812	4,066	4,999	5,000	4,014	4,999	5,000

Notes: Sample sizes are $n_1 = 5,000$ and $n_2 = 1,000$. Results are based on 5,000 simulation repetitions.

where $\Sigma_{\beta_0} = \Sigma_{1,ZZ}^{-1}\Sigma_{z,u_1}\Sigma_{1,ZZ}^{-1} + \tau\beta_0^2\Sigma_{2,ZZ}^{-1}\Sigma_{z,\varepsilon_2}\Sigma_{2,ZZ}^{-1}$ is a $k \times k$ variance-covariance matrix. Let $\hat{\Sigma}_{\zeta,\beta_0} = \frac{n_1^2}{n_1-k-p}(Z_1'Z_1)^{-1}\left(\sum_{i=1}^{n_1}\hat{u}_{1i}^2 Z_{1i}Z_{1i}'\right)(Z_1'Z_1)^{-1}$ and $\hat{\Sigma}_{\pi,\beta_0} = \frac{n_2^2}{n_2-k-p}(Z_2'Z_2)^{-1}\left(\sum_{i=1}^{n_2}\hat{\varepsilon}_{2i}^2 Z_{2i}Z_{2i}'\right)(Z_2'Z_2)^{-1}$, where $\hat{u}_{1i}$ is the i-th entry of $M_{[Z_1:X_1]}y_1$ and $\hat{\varepsilon}_{2i}$ is the i-th entry of $M_{[Z_2:X_2]}w_2$. Then Σ_{β_0} could be consistently estimated by $\hat{\Sigma}_{\beta_0} = \hat{\Sigma}_{\zeta,\beta_0} + \frac{n_1}{n_2}\beta_0^2\hat{\Sigma}_{\pi,\beta_0}$.

Following Magnusson (2010), the robust TSAR, TSK, and TSCLR test statistics for $H_0 : \beta = \beta_0$ can be written as,

$$\mathcal{T}_{1,robust}(\beta_0) = n_1(\hat{\zeta} - \hat{\pi}\beta_0)'\hat{\Sigma}_{\beta_0}^{-1}(\hat{\zeta} - \hat{\pi}\beta_0),$$

$$\mathcal{T}_{2,robust}(\beta_0) = n_1\left[\hat{\Sigma}_{\beta_0}^{-1/2}(\hat{\zeta} - \hat{\pi}\beta_0)\right]' P_{\hat{\Sigma}_{\beta_0}^{-1/2}\hat{D}_{\beta_0}}\left[\hat{\Sigma}_{\beta_0}^{-1/2}(\hat{\zeta} - \hat{\pi}\beta_0)\right],$$

$$\mathcal{T}_{3,robust}(\beta_0) = \frac{1}{2}\left[\mathcal{T}_{1,robust} - \hat{q}_{\beta_0} + \left[(\mathcal{T}_{1,robust} + \hat{q}_{\beta_0})^2 - 4\left(\mathcal{T}_{1,robust}\hat{q}_{\beta_0} - \mathcal{T}_{2,robust}\hat{q}_{\beta_0}\right)\right]^{1/2}\right]$$

where $-\hat{D}_{\beta_0} = \hat{\pi} + \frac{n_1}{n_2}\beta_0\hat{\Sigma}_{\pi,\beta_0}\hat{\Sigma}_{\beta_0}^{-1}(\hat{\zeta} - \hat{\pi}\beta_0)$, and $\hat{q}_{\beta_0} = n_1\hat{D}_{\beta_0}'(\frac{n_1}{n_2}\hat{\Sigma}_{\pi,\beta_0} - (\frac{n_1}{n_2})^2\beta_0^2\hat{\Sigma}_{\pi,\beta_0}\hat{\Sigma}_{\beta_0}^{-1}\hat{\Sigma}_{\pi,\beta_0})^{-1}\hat{D}_{\beta_0}$. When the robust variance-covariance matrix $\hat{\Sigma}_{\beta_0}$ is replaced with $\bar{\Sigma}_{\beta_0} = \hat{\sigma}_{u_1}^2(\frac{Z_1'Z_1}{n_1})^{-1} + \frac{n_1}{n_2}\beta_0^2(\frac{Z_1'Z_1}{n_1})^{-1}\hat{\sigma}_{\epsilon_2}^2$, the three robust test statistics reduce to the benchmark counterparts.

Under the null hypothesis, $\mathcal{T}_{1,robust}(\beta_0)$ and $\mathcal{T}_{2,robust}(\beta_0)$ have limit distributions $\chi^2(k)$ and $\chi^2(1)$, respectively, and $\mathcal{T}_{3,robust}(\beta_0) \Rightarrow \frac{1}{2}\left[\chi^2(1) + \chi^2(k-1) - q_{\beta_0} + \left[\left(\chi^2(1) + \chi^2(k-1) + q_{\beta_0}\right)^2 - 4\chi^2(k-1)q_{\beta_0}\right]^{1/2}\right]$, where $\chi^2(1)$ and $\chi^2(k-1)$ are independent chi-squared distributed random variables with 1 and $k-1$ degrees of freedom, respectively, given that $\hat{q}_{\beta_0} = q_{\beta_0}$. Therefore, we reject the null if $\mathcal{T}_{1,robust}(\beta_0)$ is larger than $q_{1-\alpha}(k)$, if $\mathcal{T}_{2,robust}(\beta_0)$ is larger than $q_{1-\alpha}(1)$, or if $p(\mathcal{T}_{3,robust}(\beta_0); \hat{q}_{\beta_0})$ is smaller than α, where $p(.,.)$

is defined in Section 3.1.

Similar to the benchmark case, robust TSAR, TSK, and TSCLR confidence sets of β can be constructed by inverting the robust TSAR, TSK, and TSCLR tests. In the proposed Stata command, these fully robust confidence sets are computed using a grid search. Specifically, our grid search codes are written based on the **ivtest** command developed in Finlay and Magnusson (2009). Simulation results for the robust TSAR, TSK, and TSCLR tests and confidence sets are omitted. Interested readers could refer to the simulation section in Choi et al. (2018) for details. Similar to the benchmark case, the robust TSAR, TSK, and TSCLR methods have good small sample properties regardless of the strength of instruments.

4 Stata Implementation

By default, the **weaktsiv** command generates two output tables. The first table reports TS2SLS estimates together with Inoue and Solon (2010) standard errors. The second table calculates the benchmark TSAR, TSK, and TSCLR tests and confidence sets discussed in Section 3.1. If the **robust** option is used, the **weaktsiv** command provides TS2SLS estimates together with Pacini and Windmeijer (2016) standard errors, as well as the robust TSAR, TSK, and TSCLR tests and confidence sets discussed in Section 3.2.

The two-sample IV regression model requires the use of two data samples. **weaktsiv** distinguishes the two samples based on missing values of outcome and endogenous variables. If the dataset happened to have non-missing values in both outcome and endogenous variables, **weaktsiv** would drop these observations.

4.1 Syntax

`weaktsiv` *depvar varlist_exog (varlist_endog =varlist_iv)* [*if*] [*in*] [`, noconstant robust level(#)`
`test(#) point(#) grid(#(#)#)`]

depvar is the outcome variable.

varlist_exog is the list of exogenous variables.

varlist_endog is the endogenous regressor of the model.

varlist_iv is the list of exogenous variables used togehter with *varexoglist* as instruments for *varendog*.

4.2 Options

Options of the **weaktsiv** command include:

noconstant suppresses the constant term in the regression model.

robust provides versions of two-sample weak IV robust tests that are also robust to heteroskedasticity and unequal moments of excluded instruments and exogenous regressors across the two data samples. The option also reports the Pacini and Windmeijer (2016)'s robust standard error following the TS2SLS estimation.

level(#) sets confidence level; default is *level(95)*.

test(#) sets the hypothesized value of the endogenous variable's coefficient; default is *test(0)*.

grid specifies the grid used for confidence region calculation. *grid* could only be used together with the *robust* option because the benchmark confidence regions are calculated analytically. If not specified, the default uses the TS2SLS estimator plus/minus two times the Inoue and Solon (2010) standard error and 100 grid points.

point specifies the number of points used to create the grid for confidence region calculation. *point* could only be used together with the *robust* option, and could not be used together with the *grid* option. The default uses 100 grid points.

4.3 Saved Results

weaktsiv saves the following in e():

Scalars

`e(p_TSAR)`	TSAR test p-value	`e(p_TSK)`	TSK test p-value
`e(p_TSCLR)`	TSCLR test p-value	`e(TSAR_xi)`	endpoints of benchmark TSAR confidence sets
`e(TSK_xi)`	endpoints of benchmark TSK confidence sets	`e(TSCLR_xi)`	endpoints of benchmark TSCLR confidence sets
`e(level)`	confidence level for weak-iv robust inference	`e(H0_b)`	value of β under null for weak-iv robust inference
`e(n1)`	# of obs in the sample 1 (the outcome sample)	`e(n2)`	# of obs in the sample 2 (the endog. var. sample)
`e(chi2)`	TS2SLS Wald statistic	`e(F_first)`	TS2SLS first-stage F
`e(numinst)`	number of instruments	`e(df_m_first)`	TS2SLS first-stage residual degrees of freedom
`e(df_m)`	TS2SLS model degrees of freedom	`e(df_r)`	TS2SLS residual degrees of freedom
`e(r2)`	R^2	`e(r2_a)`	adjusted R^2
`e(mss)`	TS2SLS model sum of squares	`e(rss)`	TS2SLS residual sum of squares
`e(rmse)`	TS2SLS root mean squared errors	`e(points)`	# of grid points for robust confidence sets

Macros

`e(cmd)`	`weaktsiv`	`e(robust)`	whether robust inference methods are used
`e(TSAR_type)`	type of benchmark TSAR confidence set	`e(TSK_type)`	type of benchmark TSK confidence set
`e(TSCLR_type)`	type of benchmark TSCLR confidence set	`e(TSAR_cset)`	robust TSAR confidence set
`e(TSK_cset)`	robust TSK confidence set	`e(TSCLR_cset)`	robust TSK confidence set
`e(grid)`	grid range for robust confidence set	`e(cons)`	whether constants are used
`e(instd)`	instrumented variable	`e(insts)`	instruments
`e(exog)`	exogenous variables	`e(depvar)`	dependent variable

Matrices

`e(b)`	coefficient vector	`e(V)`	variance-covariance matrix of the estimators

Functions

`e(sample)`	marks estimation sample

For the benchmark mehtods, the saved type (i.e. `e(TSAR_type)`, `e(TSK_type)`, and `e(TSCLR_type)`) and endpoint (i.e. `e(TSAR_xi)`, `e(TSK_xi)`, and `e(TSCLR_xi)`) information could be used together to retrieve the exact confidence sets using the relationship in Table 3.

Table 3: Benchmark TSCLR, TSAR, TSK Confidence sets, Analytical Solution

Test	Result Type	Interval
TSCLR	1	Empty set
	2	[x1, x2]
	3	(-∞, ∞)
	4	(-∞, x1] U [x2, ∞)
TSAR	1	Empty set
	2	[x1, x2]
	3	(-∞, ∞)
	4	(-∞, x1] U [x2, ∞)
TSK	1	Not used (not possible)
	2	[x1, x2]
	3	(-∞, +∞)
	4	(-∞, x1] U [x2, ∞)
	5	(-∞, x1] U [x2, x3] U [x4, ∞)
	6	[x1, x2] U [x3, x4]

5 Example

5.1 The Case With Just-identification

We use the dataset of Currie and Yelowitz (2000) to illustrate implementing the **weaktsiv** command in the case of just-identification. The example estimates the effects of public housing on monthly rental payments in equations (2) and (3) of Currie and Yelowitz (2000). The outcome variable is household monthly rental payments ($ry1$). The endogenous regressor is a dummy variable indicating whether a household participates in the public housing project ($ry2$). The excluded instrument is the sex composition of children, a dummy variable equaling one if the family has a boy and a girl (z). The exogenous regressors include information on the household head's age and its square, marital status, sex, race, education level, MSA-level controls for public housing supply, and children's sex.

The default **weaktsiv** command gives a table reporting the TS2SLS estimation results together with Inoue and Solon (2010) standard errors and another table reporting benchmark TSCLR results. Only TSCLR is reported here because TSAR, TSK, and TSCLR are equivalent when the model is just-identified.

For the effects of public housing on monthly rental payments, the confidence interval based on Inoue and Solon (2010) standard errors is $[.151, .592]$. The weak-instrument robust confidence interval is $[.214, .784]$, wider than the non-robust one. The TSCLR confidence interval is also centered farther away from zero than the TS2SLS confidence interval, likely due to the fact that TS2SLS suffers from an attenuation bias and is biased towards zero. These confidence interval results are also reported in column (1) of Table 3 in Choi et al. (2018). The TSCLR confidence interval reported here is slightly different from the one reported in Choi et al. (2018) due to different ways of rounding implemented in the two papers. – R codes for the emprical applications in Choi et al. (2018) only keep two significant digits after the decimal point for the fully robust confidence sets.

```
. use sample1.dta

. weaktsiv ry1 h* p* b* (ry2=z)

Two-sample Instrumental variables (TS2SLS) regression
```

```
First-stage F Results                                  Number of obs =  116901
-----------------------                                Wald chi2( 17) = 5611.78
F(  1, 10364) =    15.03                               Prob > chi2    =  0.0000
Prob > F      =   0.0001                               R-squared      =  0.1489
                                                       Adj R-squared  =  0.1488
                                                       Root MSE       =  0.2205
```

ry1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ry2	.3716513	.1123615	3.31	0.001	.1514244	.5918781
hdage	.0192549	.0015888	12.12	0.000	.016141	.0223689
hdage2	-.0198632	.0020119	-9.87	0.000	-.0238065	-.0159198
hdmarr	.0198959	.0062194	3.20	0.001	.007706	.0320859
hdfemale	-.0976462	.0107003	-9.13	0.000	-.1186186	-.0766738
hdblack	-.1211384	.0133036	-9.11	0.000	-.1472133	-.0950634
hdother	-.0163956	.00506	-3.24	0.001	-.0263131	-.0064781
hdhisp	.0005148	.0040625	0.13	0.899	-.0074478	.0084773
hded0911	.0336821	.0052502	6.42	0.000	.0233919	.0439723
hded1212	.0768111	.0054963	13.97	0.000	.0660384	.0875838
hded1315	.1345398	.0068719	19.58	0.000	.1210711	.1480085
hded16p	.20465	.0078281	26.14	0.000	.1893071	.2199928
pctlihtc	-.7054565	.1697396	-4.16	0.000	-1.038143	-.3727696
pctprj	-.8651579	.1218703	-7.10	0.000	-1.104022	-.6262939
pctrehab	-1.026055	.160297	-6.40	0.000	-1.340235	-.7118757
pctvch	-2.920268	.5245491	-5.57	0.000	-3.948376	-1.89216
boys	-.004395	.0019092	-2.30	0.021	-.008137	-.0006529
_cons	.1192835	.0300903	3.96	0.000	.060307	.17826

```
Instrumented:  ry2
Instruments:   z
Confidence set and p-value for ry2 are based on normal approximation,
thus not robust to weak instruments.
```

Weak IV Robust 95% confidence set and p-value
for H0: _b[ry2] = 0

Test	95% Confidence Set	p-value
Benchmark TSCLR	[.2137159, .7837229]	0.0000

Note: In the just identifed case, TSCLR = TSAR = TSK.

We would like to emphasize that both confidence intervals reported by the default command require the assumptions of homoskedasticity and equal moments. Next, we illustrate the use of the ", robust" option. With this option, the **weaktsiv** command reports inference results that are robust to both homoskedasticity and unequal moments of excluded instruments and exogenous regressors. In the TS2SLS output table, Pacini and Windmeijer (2016) standard errors are now reported. In the weak IV robust output table, results from the fully robust version of TSAR, TSK, and TSCLR are reported. Again, in this example, only TSCLR is reported due to just-identification. The robust TSCLR confidence interval is [.214, .784].

```
. use sample1.dta

. weaktsiv ry1 h* p* b* (ry2=z), robust grid(0(0.001)1)

Two-sample Instrumental variables (TS2SLS) regression

                                                       Number of obs =  116901
                                                       Wald chi2( 17) = 4838.19
                                                       Prob > chi2   =  0.0000
                                                       R-squared     =  0.1489
                                                       Root MSE      =  0.2205
```

ry1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ry2	.3716513	.1111589	3.34	0.001	.1537816	.5895209
hdage	.0192549	.0016301	11.81	0.000	.0160599	.02245
hdage2	-.0198632	.0020866	-9.52	0.000	-.0239529	-.0157735
hdmarr	.0198959	.0053553	3.72	0.000	.0093997	.0303921
hdfemale	-.0976462	.0103521	-9.43	0.000	-.1179361	-.0773562
hdblack	-.1211384	.0132088	-9.17	0.000	-.1470274	-.0952493
hdother	-.0163956	.0056157	-2.92	0.004	-.0274023	-.0053889
hdhisp	.0005148	.0033995	0.15	0.880	-.0061483	.0071778
hded0911	.0336821	.0050553	6.66	0.000	.0237737	.0435905
hded1212	.0768111	.0049421	15.54	0.000	.0671248	.0864975
hded1315	.1345398	.0064849	20.75	0.000	.1218295	.1472501
hded16p	.20465	.0071545	28.60	0.000	.1906273	.2186727
pctlihtc	-.7054565	.1903181	-3.71	0.000	-1.078477	-.332436
pctprj	-.8651579	.1222525	-7.08	0.000	-1.104771	-.625545
pctrehab	-1.026055	.1683943	-6.09	0.000	-1.356105	-.696005
pctvch	-2.920268	.5080081	-5.75	0.000	-3.915956	-1.92458
boys	-.004395	.0018605	-2.36	0.018	-.0080415	-.0007485
_cons	.1192835	.0304138	3.92	0.000	.0596729	.178894

```
Instrumented:  ry2
Instruments:   z
Confidence set and p-value for ry2 are based on normal approximation,
thus not robust to weak instruments.
```

Weak IV Robust 95% confidence set and p-value
for H0: _b[ry2] = 0

Test	95% Confidence Set	p-value
Robust TSCLR	[.214, .784]	0.0000

```
Note: In the just identifed case, TSCLR = TSAR = TSK.
Confidence sets estimated for 1001 points in [0,1].

.
```

5.2 The Case With Over-identification

Now we illustrate the **weaktsiv** command in two-sample IV regression models in the case of over-identification. We use the dataset of Olivetti and Paserman (2015), who examine historical intergenerational income elasticity in the U.S. We consider the specification in column (1) - row (5) in Table 3 of Olivetti and Paserman (2015) for father-son-in-law elasticity in 1950-1970. The outcome variable of interest is son-in-law's log earnings ($ry1$). The endogenous regressor is father's log earnings ($ry2$). The excluded instruments are daughters' first name dummies ($z1 - z726$). There are no exogenous regressors in the model except for the intercepts in both stages of TS2SLS regression.

The following outputs are from the default setting of the **weaktsiv** command, which reports TS2SLS estimates with Inoue and Solon (2010) standard errors as well as benchmark TSAR, TSK, and TSCLR tests and confidence sets. The results are also reported in column (1) - row (5) in Table 2 of Choi et al. (2018). The Inoue and Solon (2010) confidence interval is $[0.307, 0.401]$, while the benchmark TSCLR confidence interval is $[0.571, 0.731]$. The robust TSCLR confidence interval lies above the TS2SLS confidence interval likely due to the large attenuation bias of TS2SLS, as TS2SLS only has a first stage F-statistic equal to 1.98.

```
. use sample2.dta

. weaktsiv ry1 (ry2=z*), level(90)
note: z726 omitted because of collinearity

Two-sample Instrumental variables (TS2SLS) regression

First-stage F Results                                 Number of obs =    16650
-----------------------                               Wald chi2(  1) =  153.60
F(726, 18771) =     1.98                              Prob > chi2   =  0.0000
Prob > F      =   0.0000                              R-squared     =  0.0098
                                                      Adj R-squared =  0.0097
                                                      Root MSE      =  0.4317

------------------------------------------------------------------------------
         ry1 |      Coef.   Std. Err.       t     P>|t|     [90% Conf. Interval]
-------------+----------------------------------------------------------------
         ry2 |    .354363   .0285928    12.39    0.000     .3073294    .4013965
       _cons |    1.92788   .0830654    23.21    0.000     1.791242    2.064518
------------------------------------------------------------------------------
Instrumented:  ry2
Instruments:   z1 z2 ... z725 z726
Confidence set and p-value for ry2 are based on normal approximation,
thus not robust to weak instruments.
------------------------------------------------------------------------------


------------------------------------------------------------------------------
               Weak IV Robust 90% confidence sets and p-values
                         for H0: _b[ry2] = 0
------------------------------------------------------------------------------
 Test                           90% Confidence Set                     p-value
------------------------------------------------------------------------------
 Benchmark TSCLR             [ .5706445,  .7306929]                     0.0000
 Benchmark TSAR                        empty                            0.0000
 Benchmark TSK    [-2.803132, -2.362719] U [ .5633191,  .7388234]       0.0000
------------------------------------------------------------------------------
```

As is discussed in Choi et al. (2018), the empirical example of Olivetti and Paserman (2015) is not suitable for heteroskedasticity-robust inference because the original specifications in Olivetti and Paserman (2015) result in perfect fit for a number of individuals in either the first stage (regressing w_2 on z_2) or the reduced-form (regressing y_1 on z_1) regressions. Heteroskedasticity-robust inference, therefore, cannot be calculated for either TS2SLS or the proposed weak-instrument robust methods. To illustrate the use of our ", robust" option in models with over-identification, we turn back to the Currie and Yelowitz (2000) example discussed in Section 5.1. We randomly generate a normally distributed instrument and use it as the second instrument ($z2$). The TS2SLS results are much noisier than in Section 5.1 because the second instrument is completely uninformative and therefore brings in additional prediction error to the first stage of TS2SLS. In contrast, the robust TSCLR and TSK are very close to the results in Section 5.1. Robust TSAR confidence interval is a little wider in length because TSAR is generally inefficient when the number of instruments exceeds the number of endogenous regressors.

```
. use sample1.dta

. set seed 12345

. generate z2 = rnormal()

.
. weaktsiv ry1 h* p* b* (ry2=z z2), robust

Two-sample Instrumental variables (TS2SLS) regression

                                                Number of obs =  116901
                                                Wald chi2( 17) = 5064.72
                                                Prob > chi2   =  0.0000
                                                R-squared     =  0.1489
                                                Root MSE      =  0.2206
```

ry1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ry2	.3596925	.1059215	3.40	0.001	.1520881	.5672969
hdage	.019132	.0015721	12.17	0.000	.0160506	.0222134
hdage2	-.0197176	.0020157	-9.78	0.000	-.0236685	-.0157668
hdmarr	.0196077	.0052261	3.75	0.000	.0093647	.0298508
hdfemale	-.09669	.0099068	-9.76	0.000	-.1161072	-.0772728
hdblack	-.119754	.0125937	-9.51	0.000	-.1444374	-.0950707
hdother	-.0160799	.0054359	-2.96	0.003	-.0267342	-.0054257
hdhisp	.0003739	.0033182	0.11	0.910	-.0061296	.0068774
hded0911	.0338316	.0049277	6.87	0.000	.0241733	.0434898
hded1212	.0765785	.0047991	15.96	0.000	.0671723	.0859847
hded1315	.1340538	.0062587	21.42	0.000	.1217868	.1463207
hded16p	.2041276	.0069173	29.51	0.000	.1905697	.2176854
pctlihtc	-.6924238	.1834473	-3.77	0.000	-1.051978	-.3328699
pctprj	-.8568538	.1180531	-7.26	0.000	-1.088236	-.6254715
pctrehab	-1.016239	.163189	-6.23	0.000	-1.336087	-.696391
pctvch	-2.874294	.4874553	-5.90	0.000	-3.829699	-1.91889
boys	-.0043023	.0018107	-2.38	0.018	-.0078513	-.0007533
_cons	.1215828	.029348	4.14	0.000	.0640613	.1791043

```
Instrumented:  ry2
Instruments:   z z2
```

```
Confidence set and p-value for ry2 are based on normal approximation,
thus not robust to weak instruments.
```

Weak IV Robust 95% confidence set and p-value
for H0: _b[ry2] = 0

Test	95% Confidence Set	p-value
Robust TSCLR	[.221291, .732957]	0.0000
Robust TSAR	[.187739, .774897]	0.0000
Robust TSK	[.221291, .724569]	0.0000

Confidence sets estimated for 100 points in [-.055512, .774897].

6 References

Andrews, D. W., M. J. Moreira, and J. H. Stock. 2006. Optimal Two-sided Invariant Similar Tests for Instrumental Variables Regression. *Econometrica* 715–752.

———. 2007. Performance of Conditional Wald Tests in IV Regression with Weak Instruments. *Journal of Econometrics* 139(1): 116–132.

———. 2008. Efficient Two-sided Nonsimilar Invariant Tests in IV Regression with Weak Instruments. *Journal of Econometrics* 146(2): 241–254.

Angrist, J. D., and A. B. Krueger. 1992. The Effect of Age at School Entry on Educational Attainment: an Application of Instrumental Variables with Moments from Two Samples. *Journal of the American Statistical Association* 87(418): 328–336.

———. 1995. Split-sample Instrumental Variables Estimates of the Return to Schooling. *Journal of Business & Economic Statistics* 13(2): 225–235.

Björklund, A., and M. Jäntti. 1997. Intergenerational Income Mobility in Sweden Compared to the United States. *The American Economic Review* 1009–1018.

Bound, J., D. Jaeger, and R. Baker. 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association* 90(430): 443–450.

Brunner, E. J., S.-W. Cho, and R. Reback. 2012. Mobility, Housing Markets, and Schools: Estimating the Effects of Inter-district Choice Programs. *Journal of Public Economics* 96(7): 604–614.

Choi, J., J. Gu, and S. Shen. 2018. Weak-instrument Robust Inference for Two-sample Instrumental Variables Regression. *Journal of Applied Economics* 33(1): 109–125.

Currie, J., and A. Yelowitz. 2000. Are Public Housing Projects Good for Kids? *Journal of Public Economics* 75(1): 99–124.

Dufour, J.-M. 1997. Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica* 1365–1387.

Dufour, J.-M., and J. Jasiak. 2001. Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors. *International Economic Review* 42(3): 815–844.

Feldman, N. E. 2010. Mental Accounting Effects of Income Tax Shifting. *The Review of Economics and Statistics* 92(1): 70–86.

Finlay, K., and L. M. Magnusson. 2009. Implementing weak-instrument robust tests for a general class of instrumental-variables. *The Stata Journal* 9(3): 398–421.

Inoue, A., and G. Solon. 2010. Two-sample Instrumental Variables Estimators. *The Review of Economics and Statistics* 92(3): 557–561.

Kleibergen, F. 2002. Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression. *Econometrica* 70: 1781–1803.

Magnusson, L. M. 2010. Inference in Limited Dependent Variable Models Robust to Weak Identification. *Econometrics Journal* 13: S56–S79.

Miguel, E. 2005. Poverty and Witch Killing. *The Review of Economic Studies* 72(4): 1153–1172.

Mikusheva, A., and B. P. Poi. 2006. Tests and Confidence Sets with Correct Size when Instruments are Potentially Weak. *Stata Journal* 6(3): 335–347.

Moreira, M. J. 2003. A Conditional Likelihood Ratio Test for Sructural Models. *Econometrica* 71: 1027–1048.

———. 2009. Tests with Correct Size When Instruments Can Be Arbitrarily Weak. *Journal of Econometrics* 152(2): 131–140.

Olivetti, C., and M. D. Paserman. 2015. In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850 - 1940. *The American Economic Review* 105(8): 2695–2724.

Pacini, D., and F. Windmeijer. 2016. Robust inference for the Two-Sample 2SLS estimator. *Economics letters* 146: 50–54.

Siminski, P. 2013. Employment Effects of Army Service and Veterans' Compensation: Evidence from the Australian Vietnam-era Conscription Lotteries. *Review of Economics and Statistics* 95(1): 87–97.

Staiger, D., and J. H. Stock. 1997. Instrumental Variables Regression with Weak Instruments. *Econometrica: Journal of the Econometric Society* 557–586.

About the authors

Jaerim Choi is Assistant professor in the department of economics at the University of Hawaii at Manoa. Shu Shen is Associate professor in the department of economics at the University of California, Davis.