The Stata Journal (*yyyy*) *vv*, Number *ii*, pp. 1–21

Implementing valid two-step identification-robust confidence sets for linear instrumental-variables models

Liyang Sun
MIT
Cambridge, MA
lsun20@mit.edu

Abstract. We consider inference in the linear instrumental variables models with one or more endogenous variables and potentially weak instruments. We develop a Stata package, `twostepweakiv`, to implement the two-step identification-robust confidence sets proposed by Andrews (2018) based on the Wald tests and the linear combination tests (Andrews 2016). Unlike popular procedures based on first-stage F statistics (Stock and Yogo 2005), the two-step identification-robust confidence sets control coverage distortion without assuming the data are homoskedastic. We demonstrate the use of `twostepweakiv` with an example of analyzing the effect of wages on married female labor supply. For inference on subsets of parameters, `twostepweakiv` also implements the refined projection method (Chaudhuri and Zivot 2011). We illustrate that the refined projection method is more powerful than the conventional projection method using Monte Carlo simulations.

Keywords: coverage, first stage F-statistic, pretesting, weak instruments

1 Introduction

Linear instrumental variable (IV) regression is commonly used to estimate the causal effect of a potentially endogenous regressor X on an outcome variable Y. For linear IV estimates to be consistent, the instruments Z should be relevant for X and satisfy the exclusion restriction. Weak identification arises when this relevance requirement is close to being violated. One defining characteristic of weak instruments is that the conventional level-α Wald tests (t-tests) based on the IV estimates and their standard errors have true size larger than α. Thus, the usual (Wald) confidence sets, formed by adding and subtracting a multiple of the standard errors from the IV estimates, are not robust to weak identification. When the instruments are weak, they may cover the true parameter value less often than we intend. To conduct reliable inference, a sensible solution would be to always report robust confidence sets based on tests with correct sizes regardless of identification strength. However, this method offers no formal assessment on identification strength.

To measure the identification strength, it is common practice to conduct a pretest based on the first-stage F statistic. Researchers compare their F statistic to some critical value to gauge the degree of weak identification. Stock and Yogo (2005) tabulates critical values for a 5%-level pretest that ensure the actual coverage probability of a nominal

95% Wald confidence set is at least 90%, under the assumption of homoskedastic data. The critical values are a function of the number of endogenous regressors and the number of instruments.[1]

If the F statistic is greater than the aforementioned critical value, the pretest rejects the null of weak identification and researchers report the usual Wald confidence set. Otherwise, researchers report a robust confidence set. We can view this practice as constructing a confidence set in two steps: the first step conducts a pretest, and the second step reports a confidence set based on the result in the first step. To obtain the coverage probability of the resulting two-step confidence set, we need to account for the type I error due to the pretest itself. Taking this error into account via a Bonferroni correction gives an 85% lower bound on the coverage probability of the two-step confidence set: error rate of the pretest (5%) plus the error rate of the Wald test (10%) gives an 85% coverage probability overall.

While such two-step procedures are popular in practice, researchers commonly overlook the homoskedasticity assumption in conducting the pretest based on the F statistic. When the data are heteroskedastic, clustered, or serially correlated, there is no formal justification for comparing F statistic to critical values from Stock and Yogo (2005). Two-step confidence sets based on such pretest can have severe coverage distortions. We illustrate this coverage distortion using simulations later in Section 2: for confidence sets intended to have coverage 95%, the probability that confidence sets chosen based on whether first-stage F statistic exceeds the critical value from Stock and Yogo (2005) cover the true parameter value is much less than 95% when data is heteroskedastic.

To address this gap between empirical practice and the theoretical econometrics literature, Andrews (2018) develops an alternative method to gauge identification strength. We develop a Stata package, `twostepweakiv`, to implement such method for linear IV models.[2] `twostepweakiv` allows the researchers to specify their desired level of coverage distortion γ, which parameterizes their tolerance for weak identification. Researchers can then compare γ with a data-driven cutoff $\hat{\gamma}$, which we call distortion cutoff, returned by `twostepweakiv`, to gauge identification strength. Based on such comparison, researchers can form two-step confidence sets with coverage level at least $1-\alpha-\gamma$, even under heteroskedastic, clustered, or serially correlated data. The idea behind this method is the follows: under strong identification, some robust test statistics and the Wald statistic are equivalent local to the true parameter value. Using this equivalence, the corresponding robust confidence set with coverage level $1-\alpha-\gamma$ should be contained in the Wald (non-robust) confidence set with coverage level $1-\alpha$ when instruments are strong. We can thus assess the strength of identification by checking how larger γ needs to be for this containment to hold, and bound coverage distortions accordingly. We

1. The commonly-used critical value 10 is calculated by Stock and Yogo (2005) for a 5%-level pretest that ensures the maximum bias of the 2SLS estimator not to exceed 10% of the bias of the OLS estimator. This rule of thumb critical value applies to linear IV models with one endogenous (instrumented) regressor under the assumption of homoskedastic data.
2. To be exact, instead of a linear GMM model, we implement this valid two-step procedure using a Classical Minimum Distance (MD) model in Stata following prior work by Finlay and Magnusson (2009). We establish validity of two-step procedure using the MD approach in the online appendix.

explain this two-step procedure in more detail in Section 2.

Building on the existing Stata package `weakiv`, the new Stata package `twostepweakiv` adds the following features. It first adds to `weakiv` the linear combination (LC) test, which is used to form the robust confidence set mentioned in the previous paragraph. The LC statistic is a linear combination of the S (i.e. Anderson-Rubin) and K statistics, which yields more powerful tests in some cases when identification is weak. We discuss more of its properties in Section 3.

`twostepweakiv` also supports multiple endogenous regressors and instruments. Currently available approaches to construct confidence set for subsets of parameters when there are multiple endogenous variables can be inefficient when identification is strong. `twostepweakiv` improves power of the robust confidence sets for subsets of parameters when the model is well-identified. It is also less computationally demanding than the existing Stata command `weakiv`. We describe our approach in Section 3.1.

The next section illustrates the two-step procedures using simulations and show how the two-step procedure based on Andrews (2018) can bound coverage distortions. Section 3 provides the details needed for implementing Andrews (2018) for linear IV models for both the full set of parameters and for a single parameter. To learn how to implement the two-step confidence sets using `twostepweakiv`, readers can skip Section 3 and proceed directly to Section 4. Section 4 describes the syntax of `twostepweakiv` and demonstrate its usage with an example. Section 5 details the simulation design and presents additional simulation results to illustrate the improved performance of `twostepweakiv`. We refer interested readers to the online appendix for more details and relevant proofs.[3]

2 Constructing valid two-step confidence sets

Validity of two-step confidence sets means bounded (asymptotic) coverage distortions. To develop intuition, we first revisit the two-step procedure based on the first-stage F statistic and illustrate how it fails to bound coverage distortions when data is non-homoskedastic based on a simulation exercise.

We focus on constructing a 2SLS confidence sets for a linear IV model with a single endogenous regressor and ten instruments under heteroskedasticity.[4] In each simulation, we vary the identification strength as measured by $\|\pi_0\|$ in the following linear IV model

$$\begin{aligned} Y &= X\beta_0 + \epsilon \\ X &= Z\pi_0 + V. \end{aligned}$$

Specifically, we fix $\frac{\pi_0}{\|\pi_0\|}$ and $\beta_0 = 0$ while varying $\|\pi_0\| \in [0, 0.11]$. For each value of $\|\pi_0\|$, we construct a nominal 95% Wald confidence set and a nominal 95% AR

3. The relevant ado file, proofs, as well as simulation replication do files are available on https://github.com/lsun20/TwoStep

4. We replicate the simulation design for moderate endogeneity under heteroskedasiticty with one endogenous variable and ten instruments presented in the supplementary appendix to Andrews (2018). The results are based on 2500 simulations, each with 10,000 observations.

confidence set. The AR confidence set is robust to weak identification, whereas the Wald confidence set is not. We also construct a nominal 95% two-step confidence set based on the F statistic together with critical value $c = 38.54$. For F statistic we use the Kleibergen and Paap (2006) *rk* Wald F statistic returned by `ivreg2`. [5]The critical value $c = 38.54$ is tabulated from Stock and Yogo (2005). If the first-stage F statistic is less than c, then results of Stock and Yogo (2005) imply we cannot reject the null of weak identification or more accurately, the Wald confidence set has coverage less than 90% under homoskedasticity. In this case, we use the AR confidence set for the two-step confidence set. If the first-stage F statistic is larger than c, then we reject the null and use the Wald confidence set. As a result, under homoskedasticity, after taking into account for 5% errors due to the pretest itself, this choice of critical value ensures coverage no less than 85%, for a two-step confidence set with nominal coverage 95%.

Under non-homoskedasticity, however, it is unclear whether the two-step confidence set based on comparing the Kleibergen and Paap (2006) *rk* Wald F statistic with c can control coverage. While it is one option in `ivreg2`, Baum et al. (2007) suggests users apply with caution the critical values from Stock and Yogo (2005), which are intended for homoskedastic data.

We study the coverage probability for each of the three confidence sets under heteroskedasticity at different values of $\|\pi_0\|$ by calculating the probability that the confidence set includes the true parameter value $\beta = 0$ based on simulations. We expect the non-robust Wald confidence set to have poor coverage for small values of $\|\pi_0\|$ and the robust AR confidence set to have coverage at least 95% for all values of $\|\pi_0\|$.

In Figure 1, we plot the coverage of these confidence sets at each $\|\pi_0\|$ against the mean of the F statistic at the same $\|\pi_0\|$ as we vary $\|\pi_0\| \in [0, 0.11]$. As expected, the nominal 95% Wald confidence sets have actual coverage probability much smaller than 95% when identification is weak, which corresponds to small values of $\|\pi_0\|$ and small mean F statistic. In contrast, the nominal 95% AR confidence set has coverage at 95% for all values of $\|\pi_0\|$.

The coverage of the two-step confidence set (dashed line) drops quickly as $\|\pi_0\|$ exceeds 0.06, which corresponds to the mean of the F statistic exceeding c. Above this value, we are very likely to reject the null of weak identification in the pretest and use the Wald confidence set for the two-step confidence set. Therefore, the coverage of the Wald confidence set and the coverage of the two-step confidence set gradually coincide. However, coverage of the two-step confidence set is far below 85% above this value of $\|\pi_0\|$.

The issue is that F statistic used with existing critical values is not a reliable indicator of identification strength under heteroskedasticity: even when the mean F statistic is greater than $c = 38.54$, e.g. at 90, many nominal 95% Wald confidence sets exhibit coverage distortions exceeding 10%. The two-step confidence set thus has coverage distortion larger than 15%. Our simulation results suggest that the critical value $c = 38.54$

5. With one endogenous variable, the Kleibergen and Paap (2006) *rk* Wald F statistic is equivalent to the heteroskedasticity-robust F statistic testing the null $H_0 : \pi_0 = 0$.

fail to bound its coverage distortion at 15% with heteroskedastic data.

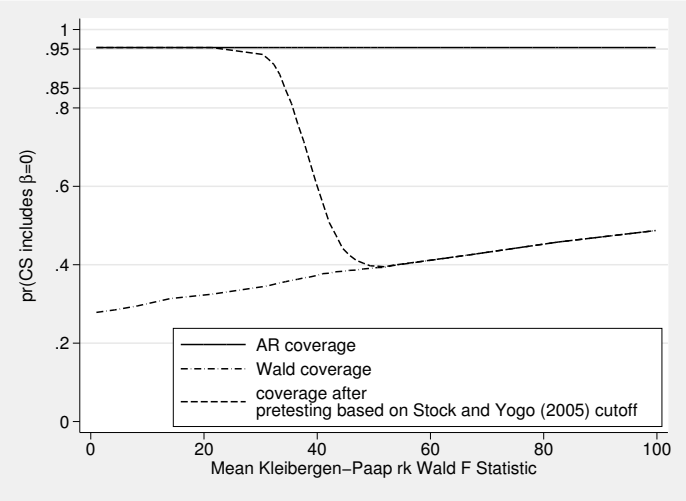

Figure 1: Coverage of two-step confidence set based on the F statistic

Since the pretest based on the F statistic tests the null hypothesis that the nominal Wald confidence set has actual coverage probability not less than $1-\alpha-\gamma\%$, we can also form pretest by directly comparing the Wald confidence set with a nominal $1-\alpha-\gamma\%$ robust confidence set as the robust confidence set would have an actual coverage probability not less than $1-\alpha-\gamma\%$.

This means that to construct a valid two-step confidence set, we need three inputs. A level $1-\alpha$ non-robust (Wald) confidence set CS_N, a level $1-\alpha$ robust confidence set CS_R, and a preliminary robust confidence set $CS_P(\gamma)$, where γ is a specified maximal coverage distortion.

Assumption 1. *The preliminary robust confidence set $CS_P(\gamma)$ needs to satisfy the following assumptions*

$CS_P(\gamma)$ has coverage at least $1-\alpha-\gamma$ when identification is weak;

$CS_P(\gamma)$ is contained in CS_R with probability one regardless of identification strength;

$CS_P(\gamma)$ is contained in CS_N with probability tending to one when identification is strong.

For a more formal discussion of the above assumptions, see Assumption (1) of Andrews (2018). While he shows how to construct CS_R and CS_P satisfying this assumption in GMM models, we extend his results to minimum distance (MD) models; Section 3 details the construction and proofs are provided in the online appendix.

Given such CS_N, CS_R, $CS_p(\gamma)$, the indicator $\mathbf{1}\{CS_p(\gamma) \not\subseteq CS_N\}$ is an indicator of weak identification because when the identification is strong, the non-robust CS_N

should contain the robust $CS_P(\gamma)$. We can construct a two-step confidence set $CS_2(\gamma)$ accordingly by

$$CS_2(\gamma) = \begin{cases} CS_N & \text{if } CS_P(\gamma) \subseteq CS_N \\ CS_R & \text{if } CS_P(\gamma) \nsubseteq CS_N \end{cases} \tag{1}$$

The two-step confidence set $CS_2(\gamma)$ is CS_N when $CS_P(\gamma)$ is contained in CS_N, and CS_R otherwise. Furthermore, since $CS_P(\gamma) \subseteq CS_2(\gamma)$ by construction, $CS_2(\gamma)$ has coverage at least $1-\alpha-\gamma$ regardless of identification strength. Thus the coverage distortion, which is the difference between nominal coverage $1-\alpha$ and the true coverage of $CS_2(\gamma)$, is at most γ.

2.1 Specify coverage distortion γ

The approach outlined above involves the maximal coverage distortion γ. While the researchers can specify a γ in forming the $CS_2(\gamma)$, different readers may prefer different choices for the maximal coverage distortion. To avoid making a choice on γ, we report both CS_R and CS_N, together with some indication of the reliability of CS_N, and readers can decide which confidence sets to focus on based on their preference for coverage distortion.

Start with some γ_{min}. `twostepweakiv` allows several levels of γ_{min}, and we will discuss the choice of γ_{min} later in this section. Note that we can construct robust confidence sets $CS_P(\gamma)$ that decrease in γ in the sense that $CS_P(\tilde{\gamma}) \subseteq CS_p(\gamma)$ for $\tilde{\gamma} \geq \gamma \geq \gamma_{min}$. We can also construct CS_R such that $CS_P(\gamma_{min}) \subseteq CS_R$ and thus $CS_P(\gamma) \subset CS_R$ for $\gamma \geq \gamma_{min}$. Define $\hat{\gamma}$ to be the smallest coverage distortion $\gamma \geq \gamma_{min}$ such that $CS_P(\gamma) \subseteq CS_N$. Intuitively, $\hat{\gamma}$ is the smallest distortion γ such that we would report CS_N for $CS_2(\gamma)$, and we can use $\hat{\gamma}$ to guide the choice of confidence set to focus on.

If a reader prefers a maximal coverage distortion γ, then he can adopt the following decision rule on which confidence set to focus on,

$$CS_2(\gamma) = \begin{cases} CS_N & \text{if } \hat{\gamma} \leq \gamma \\ CS_R & \text{if } \hat{\gamma} > \gamma \end{cases} \tag{2}$$

We refer to $\hat{\gamma}$ as the distortion cutoff. If a reader prefers a coverage distortion greater than this cutoff $\gamma \geq \hat{\gamma}$, we have $CS_P(\gamma) \subseteq CS_p(\hat{\gamma})$. Since $CS_P(\hat{\gamma}) \subseteq CS_N$, this means that $CS_P(\gamma) \subseteq CS_N$ as well and the reader should focus on CS_N. If a reader instead prefers a coverage distortion less than this cutoff $\gamma < \hat{\gamma}$, we have $CS_P(\hat{\gamma}) \subseteq CS_P(\gamma)$. Since $\hat{\gamma}$ is the smallest coverage distortion $\tilde{\gamma}$ such that $CS_P(\tilde{\gamma}) \subseteq CS_N$, this means that $CS_P(\gamma) \nsubseteq CS_N$. The decision rule in (2) is thus the same as the definition in (1), and the resulting $CS_2(\gamma)$ has asymptotic coverage at least $1-\alpha-\gamma$ under both weak and strong identification. Thus, by reporting $(CS_N, CS_R, \hat{\gamma})$ we provide all the ingredients to construct two-step confidence sets with any level of maximal coverage distortion γ that the reader prefers.

To implement this decision rule, given γ_{min}, we first construct CS_R and $CS_p(\gamma_{min})$

based on the LC test such that $CS_P(\gamma)$ decreases in γ for $\gamma \geq \gamma_{min}$ and $CS_p(\gamma_{min}) \subseteq CS_R$. Then we can calculate $\hat{\gamma}$ by solving $\min_{\gamma \geq \gamma_{min}} CS_P(\gamma) \subseteq CS_N$. So the only choice researchers need to make is γ_{min}.

The choice of γ_{min} affects how we construct CS_R. In particular, for $\gamma_{min} = 0$, CS_R degenerates to a robust confidence set based on the K test, which has some undesirable properties that we discuss in Section 3. Thus, we impose that $\gamma_{min} > 0$. Since $\hat{\gamma} \geq \gamma_{min}$, readers who prefer a distortion smaller than γ_{min} would always focus on CS_R. Thus, a larger γ_{min} would point readers more often to CS_R. `twostepweakiv` pre-tabulates several γ_{min} and sets the default γ_{min} to be 5%. However, our focus on these γ_{min} is not an endorsement of these choices, but rather to facilitate faster computing as explained in the next section.

Below we repeat the simulation exercise introduced at the beginning of this section and study the performance of $CS_2(\gamma)$ with $\gamma = 10\%$ based on (2), where CS_N is a nominal 95% Wald confidence set and CS_R is a nominal 95% LC confidence set. We set γ_{min} to be 5%. In Figure 2, we plot the coverage of these confidence sets against the mean of the F statistic as we vary the first-stage coefficients. Contrary to the simulation results shown in Figure 1, using the same simulation design, $CS_2(10\%)$ has coverage probability at least 85% regardless of the value of F statistic.

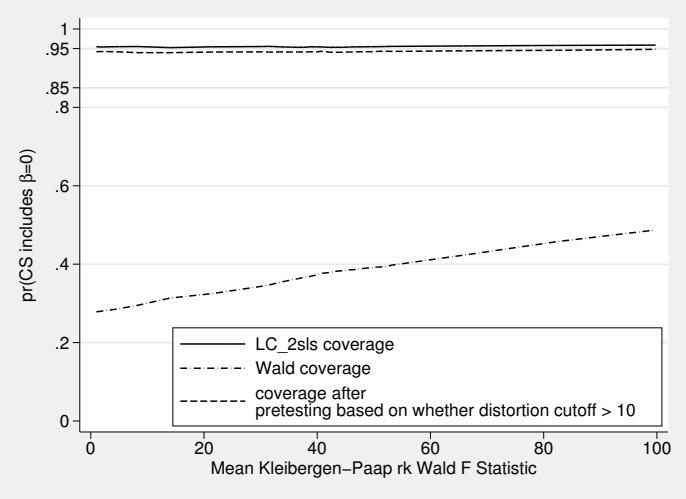

Figure 2: Coverage of two-step confidence set based on whether distortion cutoff $\hat{\gamma} > 10\%$

3 Constructing $(CS_N, CS_R, \hat{\gamma})$ for linear IV

To form a two-step confidence set with bounded coverage distortion as discussed in the previous section, we just need to construct $(CS_N, CS_R, \hat{\gamma})$. This section describes how we construct them for linear IV models based on minimum distance (MD) approach.

Specifically, we provide more details on constructing CS_R based on the LC test and calculating $\hat{\gamma}$. Readers can proceed directly to Section 4 to learn how to implement the two-step confidence sets using `twostepweakiv`.

We start by considering a linear IV model

$$\begin{aligned} Y &= X\theta_0 + \epsilon \\ X &= Z\pi_0 + V, \end{aligned}$$

which can equivalently be written as a reduced-form model

$$Y = Z\delta_0 + U \tag{3}$$
$$X = Z\pi_0 + V \tag{4}$$

for Z a $(N \times k)$ matrix of instruments, X a $(N \times m)$ vector of endogenous regressors, Y a $(N \times 1)$ vector of outcome variables. Assume that $E[Z_i U_i] = E[Z_i V_i^{'}] = 0$. Note that $\delta_0 = \pi_0 \theta_0$. We also assume that any exogenous regressors have already been partialled out.[6]

Suppose we are interested a p-dimensional parameter ($p \leq m$) $\beta = f(\theta)$. Here f is a continuously differentiable function. If we are interested in θ, then let f be the identity matrix. Denote $\frac{\partial}{\partial \theta'} f(\theta) = F(\theta)$. Assume $F(\theta_0)$ has full rank. For example we may be interested in constructing a confidence set for the j-th coordinate of the parameter vector e.g. $f(\theta) = \theta_j$. We focus on constructing a confidence set for a single coordinate of the parameter vector in Section 3.1. Below we discuss the general case where β is p-dimensional and first define its MD estimators.

The reduced-form parameter vector is $\tau = (\delta, \pi)$ and its estimator is $\hat{\tau} = (\hat{\delta}, \hat{\pi})$. The structural parameter vector is θ. Instead of deriving GMM estimators based on the moment function $g(\theta) = \frac{1}{N} \sum_i Z_i (Y_i - X_i \theta_0)$, the MD approach focuses on the structural function, which is $r(\tau, \theta) = \delta - \pi\theta$. Note that $r(\hat{\tau}, \theta_0) = \hat{\delta} - \hat{\pi}\theta_0$ and $r(\tau_0, \theta_0) = \delta_0 - \pi_0 \theta_0 = 0$. If $\frac{\partial r(\tau_0, \theta_0)}{\partial \theta}$ does not have full rank, then the model is under-identified. If $\frac{\partial r(\tau_0, \theta_0)}{\partial \theta}$ has full rank, then the model is identified. Weak identification arises when $\frac{\partial r(\tau_0, \theta_0)}{\partial \theta}$ is very close to a reduced rank matrix. Otherwise the model is well (strongly) identified.

Assume we have some estimator $\tilde{\theta}$ for θ which under strong identification is first-order equivalent to $\hat{\theta}$, which solves $\min_{\theta \in \Theta} r(\hat{\tau}, \theta)^{'} \hat{\Omega}(\theta) r(\hat{\tau}, \theta)$ for $\hat{\Omega}(\theta)$ a symmetric positive definite weighting matrix that converges uniformly to $\Omega(\theta)$ under strong identification.[7] Examples of such estimators include one-step MD, efficiently and inefficiently weighted two-step MD, and continuously updating MD.

Next we list assumptions for the asymptotic normality of $f\left(\tilde{\theta}\right)$, the estimator for β.

6. For exogenous regressors W and initial data $\left(\tilde{Y}, \tilde{X}, \tilde{Z}, W\right)$, we partial out W by letting $Y = M_W \tilde{Y}$,$X = M_W \tilde{X}$, $Z = M_W \tilde{Z}$ where $M_W = I - W\left(W'W\right)^{-1} W'$.

7. By first-order asymptotic equivalence, we mean that $\sqrt{N}(\hat{\theta} - \tilde{\theta}) \to_p 0$ under strong identification.

Assumption 2. *Under both strong and weak identification, the estimators of reduced-form parameter vector have the following asymptotic distribution,*

$$\sqrt{N}\begin{pmatrix} \hat{\delta}-\delta_0 \\ \hat{\pi}-\pi_0 \end{pmatrix} \to_d N\left(0, \begin{pmatrix} \Sigma_\delta & \Sigma_{\delta\pi} \\ \Sigma_{\pi\delta} & \Sigma_\pi \end{pmatrix}\right)$$

and we have consistent estimators $\hat{\Sigma}_\delta$, $\hat{\Sigma}_\pi$, *and* $\hat{\Sigma}_{\delta\pi}$ *for the asymptotic variance-covariance matrix.*

Then by the delta method, we have $\sqrt{N}\left(r\left(\hat{\tau},\theta_0\right)-r\left(\tau_0,\theta_0\right)\right) \to_d N\left(0,\Sigma_r\right)$. Let $\hat{\Sigma}_r\left(\theta_0\right)$ be a consistent estimator for Σ_r.

Additionally, under strong identification and regularity conditions, $\tilde{\theta}$ is consistent and asymptotic normal, and the delta method implies

$$\sqrt{N}\left(f\left(\tilde{\theta}\right)-\beta_0\right) \to_d N\left(0,V_\beta\right)$$

where $\beta_0 = f\left(\theta_0\right)$ and $V_\beta = F(\theta_0)V_\theta F(\theta_0)'$. The asymptotic normality of $f\left(\tilde{\theta}\right)$ under strong identification means that the Wald statistic

$$W\left(\beta\right) = N\cdot\left(f\left(\tilde{\theta}\right)-\beta\right)'\hat{V}_\beta^{-1}\left(f\left(\tilde{\theta}\right)-\beta\right)$$

for $\hat{V}_\beta$ a consistent estimator of V_β has an asymptotic distribution

$$W\left(\beta_0\right) \to_d \chi^2_p. \tag{5}$$

and under $H_0 : \beta = \beta_0$ and strong identification. Collecting the set of values β such that the Wald test does not reject gives us a non-robust Wald confidence sets for $\beta = f\left(\theta\right)$,

$$CS_N = \left\{\beta : W\left(\beta\right) \leq \chi^2_{p,1-\alpha}\right\}.$$

In the online appendix we prove that we can construct CS_R based on the the LC test, which builds on the K and S test. We define the K statistic as

$$K_{\Omega,f}\left(\theta\right) = N\cdot\left(\beta^*\left(\theta\right)-f\left(\theta\right)\right)'\left(M\left(\theta\right)'\hat{\Sigma}_r\left(\theta\right)M\left(\theta\right)\right)^{-1}\left(\beta^*\left(\theta\right)-f\left(\theta\right)\right).$$

For more details on $M\left(\theta\right)$, see Andrews (2018) and the online appendix. Also define the S statistic as

$$S\left(\theta\right) = N\cdot r\left(\hat{\tau},\theta\right)'\hat{\Sigma}_r\left(\theta\right)^{-1}r\left(\hat{\tau},\theta\right).$$

The LC statistic is defined as

$$LC_{\Omega,f,a}\left(\theta\right) = K_{\Omega,f}\left(\theta\right)+a\cdot S\left(\theta\right). \tag{6}$$

For $a = 0$, the LC statistic degenerates to the K statistic. However, the K statistic has some undesirable properties. Kleibergen (2005) shows that the K statistic can

be interpreted as a score statistic based on the continuously updating GMM objective function. This means that the confidence set based on the K statistic always contains all local minima and maxima of the continuously updating objective function. Thus, the K test can have low power if the continuously updating GMM objective has multiple local minima or maxima. By combining the K statistic with the S statistic, the LC statistic improves the power of the K test. In the online appendix, we show that under both strong and weak identification, under $H_0 : \theta = \theta_0$, $LC_{\Omega,f}(\theta)$ has the following asymptotic distribution

$$LC_{\Omega,f,a}(\theta_0) \to_d (1+a) \cdot \chi^2_p + a \cdot \chi^2_{k-p}.$$

To form CS_R and $CS_p(\gamma)$ based on $LC_{\Omega,f}(\theta)$, we let a to be a function of γ. Specifically, for a given value of γ, define $a(\gamma)$ by

$$H^{-1}(1-\alpha-\gamma; a(\gamma), k, p) = \chi^2_{p,1-\alpha} \tag{7}$$

where $H^{-1}(1-\alpha-\gamma; a(\gamma), k, p)$ is the $1-\alpha-\gamma$ quantile of a $(1+a(\gamma))\cdot\chi^2_p + a(\gamma)\cdot\chi^2_{k-p}$ distribution, and $\chi^2_{p,1-\alpha}$ is the $1-\alpha$ quantile of a χ^2_p distribution.[8] We can now construct robust confidence sets for $\beta = f(\theta)$. In particular, define

$$CS_{R,\theta} = \left\{\theta : K_{\Omega,f}(\theta) + a(\gamma) \cdot S(\theta) \leq H^{-1}(1-\alpha; a(\gamma), k, p)\right\}$$

$$CS_R = \{f(\theta) : \theta \in CS_{R,\theta}\} = \left\{\beta : \min_{\theta:\beta=f(\theta)} K_{\Omega,f}(\theta) + a(\gamma) \cdot S(\theta) \leq H^{-1}(1-\alpha; a(\gamma), k, p)\right\} \tag{9}$$

where $H^{-1}(1-\alpha; a(\gamma), k, p)$ is the $1-\alpha$ quantile of a $(1+a(\gamma)) \cdot \chi^2_p + a(\gamma) \cdot \chi^2_{k-p}$ distribution. The initial confidence set $CS_{R,\theta}$ collects the set of values θ where the linear combination statistic falls below the critical value, and so will cover θ_0 with probability tending to $1-\alpha$ under both strong and weak identification. Similar to the conventional projection method, CS_R takes the image of the initial confidence set under $f(\cdot)$ to construct a confidence set for $\beta = f(\theta)$. Unlike the conventional projection method, we alter test statistic (specifically the K statistic in the linear combination) based on $f(\cdot)$. Thus we refer to this method the refined projection method. We discuss this refinement more in Section 3.1.

We can similarly define $CS_P(\gamma)$ where the critical value in (9) is replaced with $\chi^2_{p,1-\alpha}$. We can show that the above CS_R and $CS_p(\gamma)$, together with the Wald CS_N,

8. By definition $a(\gamma)$ solves

$$Pr\left\{(1+a(\gamma)) \cdot \chi^2_p + a(\gamma) \cdot \chi^2_{k-p} \leq \chi^2_{p,1-\alpha}\right\} = 1-\alpha-\gamma \tag{8}$$

To find this value in practice, we take one hundred thousand independent simulation draws from χ^2_p and χ^2_{k-p} distributions and solve numerically for the value a which sets the $1-\alpha-\gamma$ quantile of the corresponding linear combination of these draws to $\chi^2_{p,1-\alpha}$. The simulation can be computationally intensive and we pre-tabulate $a(\gamma)$ for a few levels of γ in `twostepweakiv`. For levels of γ not pre-tabulated, `twostepweakiv` solves for $a(\gamma)$ based on ten thousand simulations before it calculates confidence sets.

satisfy Assumption 1 and thus can be used to construct $CS_2(\gamma)$ with coverage probability at least $1-\alpha-\gamma$.

We demonstrated in Section 2 that by reporting $(CS_N, CS_R, \hat{\gamma})$, we provide all the ingredients needed to construct $CS_2(\gamma)$ with any level of maximal coverage distortion γ that the reader prefers. Here we describe how to calculate the distortion cutoff $\hat{\gamma}$. Andrews (2018) shows that to find $\hat{\gamma}$ that solves $\min_{\gamma \geq \gamma_{min}} CS_P(\gamma) \subseteq CS_N$, we first need to find

$$\tilde{a} = \sup_{\Theta} \frac{\chi^2_{p,1-\alpha} - K_{\Omega,f}(\theta)}{S(\theta)} \cdot 1\left\{W_{\Omega,f}(\theta) > \chi^2_{p,1-\alpha}\right\} \tag{10}$$

and then calculate

$$\tilde{\gamma} = 1 - \alpha - Pr\left\{(1+\tilde{a}) \cdot \chi^2_p + \tilde{a} \cdot \chi^2_{k-p} \leq \chi^2_{p,1-\alpha}\right\}. \tag{11}$$

In practice $Pr\left\{(1+\tilde{a}) \cdot \chi^2_p + \tilde{a} \cdot \chi^2_{k-p} \leq \chi^2_{p,1-\alpha}\right\}$ is again calculated based on one hundred thousand independent simulation draws. The distortion cutoff is then $\hat{\gamma} = \max\{\tilde{\gamma}, \gamma\}$.

3.1 Refined robust confidence sets for a single parameter

Let $\theta = (\beta_1, \ldots, \beta_m)$ be the full parameter vector. In this subsection, we discuss how to construct a two-step confidence set for the j-th coordinate β_j, as we are only interested in the causal effect of its corresponding regressor. The remaining parameters are called nuisance parameters since they are not of immediate interest. This can be achieved by setting $f(\theta) = e_j' \cdot \theta$ where e_j is the j-th coordinate vector. Then we have $f(\theta) = e_j' \cdot \theta = \beta_j$ and $F(\theta) = \frac{\partial}{\partial \theta} f(\theta) = e_j'$. Below we derive test statistics for a single element of the parameter vector.

It is straightforward to construct non-robust Wald confidence set for β_j - simply its estimator $\tilde{\beta}_j$ plus and minus $c_{\alpha/2}$ times its standard error where $c_{\alpha/2}$ is the $1-\alpha/2$ quantile of a standard normal distribution. By contrast, most robust confidence sets are based on test inversion on a fine grid of parameters. Namely, we test all potential hypotheses $H_0 : \theta = \theta_0$ and define our confidence set as the set of all θ_0 for which the hypothesis is accepted. As a result of such an inversion, we would end up with a joint (multi-dimensional) confidence set for β and nuisance parameters. If we assume the nuisance parameters are well-identified, then by the results of Kleibergen (2004), we can plug in the continuously updating estimator for the nuisance parameters, and construct robust confidence set for β_j based on test inversion on a grid of values for the parameter of interest only. This option is supported by `twostepweakiv` - see `strong_options`. If we don't want to assume the nuisance parameters are well-identified, the conventional method, projection method, projects the original confidence set for the full parameter vector onto the j-th dimension. The resulting confidence set for β_j is typically conservative, particularly when the nuisance parameters are actually well-identified.

Originally proposed by Chaudhuri and Zivot (2011), the refined projection method alters the K statistic to construct the joint confidence set, and then project it onto the

j-th dimension to reduce the degree of conservativeness. This method is more powerful than the conventional projection method in the well-identified case. Intuitively, one reason why the projection method has low power is that it first tests all parameters simultaneously and then projects the joint confidence sets onto β_j. By refining the K statistic, we can focus power on testing the one dimensional hypothesis $H_0 : \beta_j = \beta_{j0}$. Andrews (2018) generalizes this method to allow for inefficient weighting matrices, which accommodates the inefficiently weighted two-step estimators. Specifically, when constructing a robust confidence set for β_j based on the K test, we refine the K statistic to be

$$K_{\Omega,e_j}(\theta) = N \cdot \frac{\left(\beta_j^*(\theta) - \beta_j\right)' \left(\beta_j^*(\theta) - \beta_j\right)}{M(\theta)' \hat{\Sigma}_r(\theta) M(\theta)}.$$

If we further construct a robust confidence set for β_j based on the LC test, we define the LC statistic as

$$LC_{\Omega,e_j,a}(\theta) = K_{\Omega,e_j}(\theta) + a \cdot S(\theta). \tag{12}$$

where the S statistic remains the same as in the full parameter case. This statistic is used to form CS_R in (9). In Section 5, we present a simulation study that compares this refined projection method with the conventional projection method.

4 Command description

The command syntax for `twostepweakiv` is

`twostepweakiv` *estimator depvar* [*varlist1*] *(varlist2=varlist_iv)* [*if*] [*in*] [*weight*]
[`, project(varlist) test_options grid_options size_options strong_options`]

`estimator` all estimators are formulated as minimum distance estimators following `weakiv` and Finlay and Magnusson (2009).

estimator	Description
`2sls`	two-stage least squares estimator
`liml`	limited-information maximum likelihood
`md2s`	two-step minimum distance estimator
`cue`	continuous updating estimator

Notes: We drop the `robust` option because the choice of estimator implies the choice of weight matrix (`2sls` and `liml` for inefficient weight matrix and `md2s` and `cue` for efficient weight matrix). The choice of VCE estimator is not necessarily the same as the choice of the weight matrix. By default, we calculate VCE estimators robust to heteroskedasticity regardless of choice of estimator, so that test statistics are robust to heteroskedasticity in all cases. Other types of VCE estimators can be specified in `cluster` for clustered VCE estimator and, `kernel` for kernel-based VCE estimator. More details can be found in help file for `ivreg2`.

`varlist1` is the list of exogenous variables.

`varlist2` is the list of endogenous variables. Users can specify up to five endogenous variables without further calculation - see `gammalevel(#)`.

`varlist_iv` is the list of instruments for *varlist2*. Users can specify up to 50 instruments without further calculation - see `gammalevel(#)`.

4.1 Options

Below we describe the options for the command. For each option, we describe when it should be used or adjusted. We also describe the default values for options when applicable.

`test_options`

Users can specify which tests to use in constructing the confidence intervals. Note that even though `twostepweakiv` can report robust confidence sets based on several test statistics, it calculates the coverage distortion for the LC test only (i.e. readers should form two-step confidence sets based on the LC confidence interval.) When there are more than one endogenous variable, users can also specify for which parameter to construct confidence interval using the refined projection method.

`citestlist(testlist)` For the full parameter vector, several tests are inherited from original `weakiv`. When the model is exactly identified, only the AR test is performed.

	test	addition to `weakiv`
Wald	non-robust Wald test	No
AR	AR test	No
K	K test with the efficient weight matrix	No
K_2SLS	K test with 2sls inefficient weight matrix	Yes
LC	LC test with the efficient weight matrix	Yes
LC_2SLS	LC test with 2sls inefficient weight matrix	Yes

Table 1: Tests available for full parameter vector

Notes: default tests are Wald, AR, K_2SLS and LC_2SLS when the estimator is `2sls` or `liml`; default tests are Wald, AR, K and LC when the estimator is `md2s` or `cue`. When the model is exactly identified, only Wald and AR tests are performed.

`project(varlist)` When there are more than one weak endogenous variable, and we are interested in inference for one endogenous regressor β only, we calculate confidence sets using the "refined" projection method for β while treating the other endogenous regressors η as free unknown parameters (nuisance parameters.) See option `strong` if we are willing to make additional assumptions about η.

`ptestlist(project_testlist)` `twostepweakiv` allows the following tests for constructing confidence sets for β specified in `project` using the "refined" projection method.

	test
Wald	non-robust Wald test
K	K test with the efficient weight matrix
K_2SLS	K test with 2sls inefficient weight matrix
LC	LC test with the efficient weight matrix
LC_2SLS	LC test with 2sls inefficient weight matrix

Table 2: Tests available for scalar parameters using "refined" projection method
Notes: default tests are Wald and LC_2SLS when the estimator is `2sls` or `liml`; default tests are Wald and LC when the estimator is `md2s` or `cue`.

`grid_options`

Since we construct confidence sets based on test inversion, we need to specify grid points over which to calculate the test statistics. If left unspecified, the default grid is centered around the point estimate with a width equal to five times the Wald confidence interval. The default number of grid points is 100 for one endogenous variable; 25×25 points for two endogenous variables. With weak/strong instruments, this may often be too small/large a grid to estimate the confidence sets. We recommend users specify the grid using the `gridmin/gridmax(numlist)` option.

`cuepoint` adds the CUE estimator to the grid for calculating confidence sets. If confidence sets are non-empty, they should contain the CUE estimator.

`gridmin/gridmax(numlist)` sets the lower/upper bounds for grid search in order corresponding to endogenous regressors.

`gridpoints(numlist)` specifies the number of equally spaced values over the grid.

`gridmult(#)` specifies the multiplier of Wald confidence interval for grid. The default is `gridmult(5)`.

`size_options`

Recall that to calculate the distortion cutoff $\hat{\gamma}$ and to construct two-step confidence interval, we need to specify the nominal size of test α and an initial coverage distortion γ_{min} as explained in Section 2.1. If left unspecified, the default values are $\alpha = 95$ and $\gamma_{min} = 5$.

`level(#)` specifies the nominal size of tests. Calculation is faster for $\alpha = 90,\ 95,\ 99$. For other values, additional calculation is needed - see `gammalevel(#)`.

`gammalevel(#)` specifies γ_{min}. Calculation is faster for $\gamma_{min} = 1,\ 2,\ 5,\ 10,\ 15,\ 20$. For other values, additional calculation is needed. As explained in Section 3, the LC test requires that we use simulations to calculate the weights and critical values. Since simulations are computationally intensive, we tabulate the weights $a(\gamma)$ and critical values $H^{-1}(1-\alpha; a(\gamma), k, p)$ for $p = 1, \ldots, 5$, $k = 1, \ldots, 50$, $\alpha = 90,\ 95,\ 99$ and $\gamma = 1,\ 2,\ 5,\ 10,\ 15,\ 20$ in advance. For values not pre-tabulated, we include simulation code to calculate the corresponding weights and critical values, which can

be slow.

`strong_options`

When there is more than one endogenous variable, we can specify `project(varlist)` to construct confidence set for one endogenous regressor with coefficient β only. If we assume the coefficients η on the other regressors are strongly-identified as specified in `strong(varlist)`, then we can plug in their estimators and construct robust confidence set for β based on test inversion on a grid of values for β only. The choice of the estimators is specified by `cuestrong`.

`strong(varlist)` specifies strongly-identified endogenous regressors with coefficients η when there is more than one weak endogenous variable, and treats the coefficients β on the rest of the endogenous regressors as potentially weakly-identified. If unspecified, then all endogenous regressors are assumed to be potentially weakly-identified. If `strong` is specified and `cuestrong` is not evoked, then at each grid point β_0 for the potentially weakly-identified parameters β, we calculate `md2s` estimates for strongly-identified parameters η under the null hypothesis $H_0 : \beta = \beta_0$. We then evaluate test statistics for β, plugging the `md2s` estimates $\widehat{\eta}$.

`cuestrong` uses CUE point estimates for strongly-identified endogenous regressors (specified in `strong(varlist)`) and include these point estimates in grid. That is, at each grid point β_0 for the potentially weakly-identified parameter β, we calculate CUE for strongly-identified parameters η under the null hypothesis $H_0 : \beta = \beta_0$. We then evaluate test statistics for β, plugging the CUE estimates $\widehat{\eta}$ in the strongly-identified parameters. Note that this option may be computationally intensive.

4.2 Saved results

Macros

`e(ar_cset)`	confidence set based on AR test	`e(k_cset)`	confidence set based on K or K_2SLS test
`e(wald_cset)`	confidence set based on Wald test	`e(lc_cset)`	confidence set based on LC or LC_2SLS test

Matrices

`e(citable)`	test statistics and rejection indicators for every grid point	`e(pxxcitable)`	test statistics and rejection indicators for the endogenous regressors specified in `project`
`e(F)`	heteroskedastic-robust F statistic	`e(wbeta)`	weakly-identified coefficients from IV model used for Wald tests
`e(wbeta_var)`	VCE from IV model used for Wald tests	`e(sbeta)`	if `strong(.)` is used, estimated strongly-identified coefficients at null
`e(ebeta)`	Wald point estimates for full set of endogenous regressors	`e(cuebeta)`	if `cuepoint` is used, CUE point estimates for full set of endogenous regressors

4.3 Example: Married female labor market participation

We demonstrate the use of this command with the data from Mroz (1987), available from the Stata website at http://www.stata.com/data/jwooldridge/eacsap/mroz.dta. In this example, we are interested in the effect of wages on married female labor supply. We regress working hours on the log wage. As instruments for the wage, we use labor market experience and its square, father's and mother's years of education. We consider the subsample of women who are participating in the labor market and have strictly positive wages. This example is identical to the example considered in Finlay and Magnusson (2009), implemented using `ivreg2` and `weakiv`.

Here we use `twostepweakiv` to construct an identification-robust confidence set for the effect of wages on labor supply based on the 2SLS estimator. In the Stata command, we specify the estimator to be `2sls`, followed by the regression specification. We let `twostepweakiv` construct robust confidence sets based on its default `test_options` for a 2SLS estimator, which are K_2SLS, LC_2SLS and AR. Instead of the K and LC test, we consider the K_2SLS and LC_2SLS tests with inefficient weight matrix, since it is used by the 2SLS estimator. While only the LC_2SLS confidence set (together with the Wald confidence set) is needed to construct a two-step identification robust confidence set, we report the other two confidence sets for reference.

Recall that these robust confidence sets are based on test inversion. Instead of letting `twostepweakiv` use the default `grid_options`, we specify the grid to be $[-1000, 8000]$. While a fine grid takes longer to compute, to calculate the robust confidence sets and the distortion cutoff $\hat{\gamma}$, we need a rather fine and wide grid. We use the default `size_options`, which set the test size $\alpha = 5\%$ and the initial coverage distortion $\gamma_{min} = 5\%$. Since there is only one endogenous variable, there is no `strong_options` to specify.

```
. use http://www.stata.com/data/jwooldridge/eacsap/mroz.dta

. twostepweakiv 2sls hours nwifeinc educ age kidslt6 kidsge6 (lwage = exper expe
> rsq fatheduc motheduc) if inlf==1, gridmin(-1000) gridmax(8000) gridpoints(901
> )
Estimating model using 2sls estimator ...
Obtaining 2SLS point estimates...
Estimating confidence sets over 901 grid points
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100
..................................................   150
..................................................   200
..................................................   250
..................................................   300
..................................................   350
..................................................   400
..................................................   450
..................................................   500
..................................................   550
..................................................   600
..................................................   650
..................................................   700
```

```
.................................................   750
.................................................   800
.................................................   850
.................................................   900
.
Weak instrument robust tests and confidence sets for linear IV
Confidence sets based on k_2sls lc_2sls ar tests are shown below.
------------------------------------------------------------------------------
   Test |   Conf. level  Conf. Set
--------+---------------------------------------------------------------------
 K_2sls |       95%      [    -840,    -680] U [     710,    4070]
LC_2sls |       95%      [     750,    4100]
     AR |       95%      [     770,    6930]
--------+---------------------------------------------------------------------
   Wald |       95%      [ 350.552,  2180.1]
------------------------------------------------------------------------------
Confidence sets (if calculated) based on 901 points in [ -1000, 8000].
LC test gamma_min is 5%; distortion cutoff is 33% based on the given grid,
obtained by 10^6 simulation draws.
Number of obs N = 428.
Method = minimum distance/Wald.
Tests robust to heteroskedasticity.
Wald confidence set is based on 2sls estimates and is not robust to weak
instruments.
```

In the output, `twostepweakiv` first reports the robust confidence sets. The negative values of the K_2SLS confidence set are discarded in the LC_2SLS confidence interval, indicating the spurious behavior of the K_2SLS test in that part of the parameter space. `twostepweakiv` also reports the non-robust Wald confidence set. Compared to the results shown in Finlay and Magnusson (2009), our Wald confidence set is slightly different. Even though `2SLS` point estimates are the same under MD (returned by `twostepweakiv` in `e(ebeta)`, not shown here) and under GMM (returned by `ivreg2`, not shown here), their standard errors are slightly different due to the MD approach we take.

Lastly, we explain how to form a two-step identification robust confidence set based the above results. Recall that we need three ingredients: the distortion cutoff $\hat{\gamma}$ that indicates identification strength, the robust LC_2SLS confidence set, and the non-robust Wald confidence set. The latter two are shown in the table, and the distortion cutoff is shown in the caption.

The distortion cutoff (33%) is rather large, indicating the instrument may be weak. Suppose a reader is willing to tolerate a coverage distortion of up to 10%, which is less than the distortion cutoff, then to form a two-step identification-robust confidence set with coverage distortion bounded by 10%, he should focus on the LC_2SLS confidence interval. If a reader is willing to tolerate a coverage distortion of up to 50%, which is larger than the distortion cutoff, then he should focus on the Wald confidence set.

The rest of the caption reminds the readers that we calculate MD versions of robust tests as well as the non-robust Wald test, all test statistics use heteroskedasticity-robust VCE estimators, and the Wald confidence set is not robust to weak identification.

The help file for `twostepweakiv` contains more examples using data from Mroz

(1987), including inference for a single regressor using the refined projection method when there are more than one endogenous regressors.

5 Linear IV simulation design

For inference on subsets of parameters, `twostepweakiv` also implements the refined projection method discussed in Section 3.1, which alters the test statistics before projecting the joint confidence set onto the subset. In this section, we illustrate that the refined projection method is more powerful than the conventional projection method using Monte Carlo simulations.

We simulate the following linear IV model under heteroskedasticity, where any exogenous regressors have already been partialled out as in the main text. As before the model is

$$\begin{aligned} Y &= X\theta_0 + \epsilon \\ X &= Z\pi_0 + V, \end{aligned}$$

which can equivalently be written as a reduced-form model

$$\begin{aligned} Y &= Z\delta_0 + U \\ X &= Z\pi_0 + V \end{aligned}$$

for Z a $N \times k$ matrix of instruments, X a $N \times m$ vector of endogenous regressors, Y a $N \times 1$ vector of outcome variables. We set $k = 5$. We take $Z_i \sim N(0, I)$. Since we are interested in the performance of projection tests, we set $m = 2$. Let $\theta_0 = (\eta_0, \beta_0)$. We are interested in constructing confidence sets for the scalar coefficient β, treating η as a nuisance parameter. We set $N = 500$.

To simulate heteroskedastic data, we draw $\begin{pmatrix} \epsilon_i \\ V_i \end{pmatrix} | Z_i \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_V \cdot \exp\left(\frac{1}{2} Z_i^1\right)\right)$ where Z_i^1 is the first instrument and $\Sigma_V = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$.

Since we are interested in the performance of projection tests in this Section, we set $m = 2$. Let $\theta_0 = (\eta_0, \beta_0)$. We are interested in constructing confidence sets for the scalar coefficient β, treating η as a nuisance parameter. For weak identification, we set $\pi_0' = \begin{pmatrix} 0.1 & 0.2 & 0.3 & 0.4 & 0.5 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix}$. For strong identification, we set $\pi_0' = \begin{pmatrix} 0.1 & 0.2 & 0.3 & 0.4 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \end{pmatrix}$. Note that in the weak identification case we consider, the local rank reduction in the matrix of the first-stage coefficients π_0 is just from one first stage, π_0^2, being close to zero, not from π_0^1 and π_0^2 being almost the same. Thus, we report the mean heteroskedasticity-robust F statistic for each of the first-

stage regressions, $F_j = \frac{N}{k}\widehat{\pi}_j'\widehat{\Sigma}_{\widehat{\pi}_j}^{-1}\widehat{\pi}_j$ for $j = 1, 2$, $\widehat{\pi}_j = \left(Z'Z\right)^{-1} Z'X^j$ the OLS estimator for π_0^j, and $\widehat{\Sigma}_{\widehat{\pi}_j}$ a heteroskedasticity-robust estimator for the variance of $\sqrt{N}\left(\widehat{\pi}_j - \pi_0^j\right)$.

We compare power of the conventional projection method with the refined projection method. We calculate the probability of rejecting $\beta = 0$ when $\beta_0 = b$ using both methods based on 500 simulations of the linear IV model, where we set $\frac{b}{\|\pi_2\|}$ to be 51 equidistant points over $[-3, 3]$. In all cases $\eta_0 = 0$. The grid consists of 101×1 equally spaced values: for the nuisance parameter η, the grid is centered around 0 with width equal to twelve times $\|\pi_2\|$; for β, we only include one grid point at 0. We use 2SLS estimators for point estimates $\hat{\beta}$ and $\hat{\eta}$.

We simulate the performance of conventional and refined projection methods for LC_2SLS test. Recall that LC_2SLS test is LC test with inefficient weights. Since the point estimates are estimated by 2sls estimators, we use LC_2SLS test because it also uses inefficient weight matrices. We include the performance of K_2SLS test to illustrate its spurious power losses.[9]

We scale b inversely by $\|\pi_2\|$ for the power curves to be comparable.

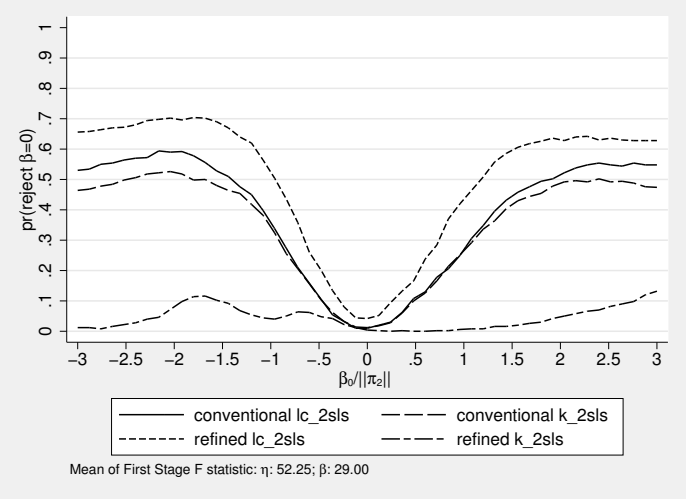

Figure 3: Comparison between conventional and refined projection methods under weak identification

Figure 3 and 4 report power of the LC_2SLS test and K_2SLS test coupled with either conventional and refined projection methods under weak and strong identification respectively. For LC_2SLS test, the refined projection method yields higher power than

9. Even in strong identified models confidence sets based on K test are not necessarily consistent for θ_0. This issue is more pronounced for refined projection method than for conventional projection method. See Section 3.2 of Andrews (2018) for more details.

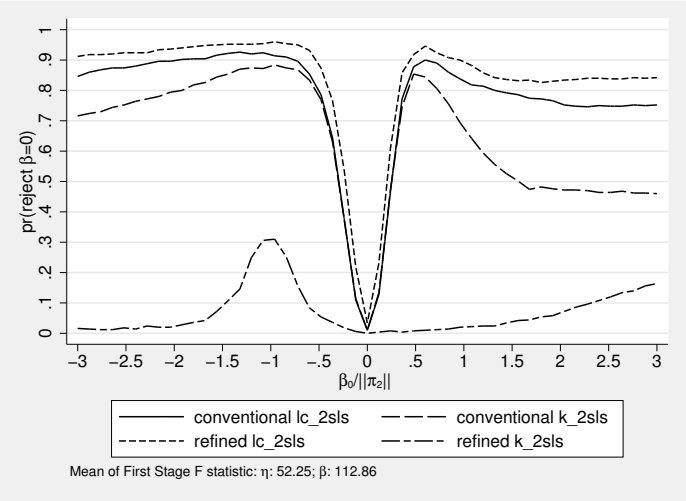

Figure 4: Comparison between conventional and refined projection methods under strong identification

the conventional projection method in all cases. The K_2SLS test is in general not consistent, even under strong identification.

6 Acknowledgments

I am grateful to Isaiah Andrews for his guidance and support. I thank Kevin Kainan Li and an anonymous referee for helpful comments.

7 References

Andrews, I. 2016. Conditional Linear Combination Tests for Weakly Identified Models. *Econometrica* 84(6): 2155–2182.

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

Baum, C. F., M. E. Schaffer, and S. Stillman. 2007. Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata Journal* 7(4): 465–506.

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

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

Kleibergen, F. 2004. Testing Subsets of Structural Parameters in the Instrumental Variables Regression Model. *The Review of Economics and Statistics* 86(1): 418–423.

———. 2005. Testing Parameters in GMM Without Assuming that They Are Identified. *Econometrica* 73(4): 1103–1123.

Kleibergen, F., and R. Paap. 2006. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics* 133(1): 97–126.

Mroz, T. A. 1987. The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions. *Econometrica* 55(4): 765–799.

Stock, J., and M. Yogo. 2005. Testing for Weak Instruments in Linear IV Regression. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. D. W. Andrews, 80–108. New York: Cambridge University Press.

About the authors

Liyang Sun is a graduate student in the Department of Economics at MIT.