The Stata Journal (2018)
18, Number 4, pp. 803–825

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

Liyang Sun
MIT
Cambridge, MA
lsun20@mit.edu

Abstract. In this article, we consider inference in the linear instrumental-variables models with one or more endogenous variables and potentially weak instruments. I developed a command, `twostepweakiv`, to implement the two-step identification-robust confidence sets proposed by Andrews (2018, *Review of Economics and Statistics* 100: 337–348) based on Wald tests and linear combination tests (Andrews, 2016, *Econometrica* 84: 2155–2182). Unlike popular procedures based on first-stage F statistics (Stock and Yogo, 2005, Testing for weak instruments in linear IV regression, in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*), the two-step identification-robust confidence sets control coverage distortion without assuming the data are homoskedastic. I 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, *Journal of Econometrics* 164: 239–251). I illustrate that this method is more powerful than the conventional projection method using Monte Carlo simulations.

Keywords: st0541, twostepweakiv, coverage, first-stage F statistic, pretesting, weak instruments

1 Introduction

Linear instrumental-variables (IV) regression is commonly used to estimate the causal effect of a potentially endogenous regressor, $\mathbf{X}$, on an outcome variable, $\mathbf{Y}$. For linear IV estimators to be consistent, the instruments, $\mathbf{Z}$, should be relevant for $\mathbf{X}$ 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 a true size larger than α. Thus, the usual (Wald) confidence sets (CSs), 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 one intends. To conduct reliable inference, one can always report robust CSs 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, one commonly conducts a pretest based on the first-stage F statistic. Researchers compare their F statistic with some critical value to gauge the degree of weak identification. Stock and Yogo (2005) tabulate critical values for a 5%-level pretest that ensure the actual coverage probability of a nominal 95% Wald CS 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 CS. Otherwise, researchers report a robust CS. We can view this practice as constructing a CS in two steps: the first step conducts a pretest, and the second step reports a CS based on the result in the first step. To obtain the coverage probability of the resulting two-step CS, we need to account for the type I error due to the pretest itself. Accounting for this error via a Bonferroni correction gives an 85% lower bound on the coverage probability of the two-step CS; 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 the F statistic with critical values from Stock and Yogo (2005). Two-step CSs based on such pretests can have severe coverage distortions. We illustrate this coverage distortion using simulations in section 2.

To address this gap between empirical practice and the theoretical econometrics literature, Andrews (2018) developed an alternative method to gauge identification strength. I developed a command, `twostepweakiv`, to implement such a 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 $\widehat{\gamma}$—which we call the distortion cutoff—returned by `twostepweakiv` to gauge identification strength. Based on such a comparison, researchers can form two-step CSs with coverage level at least $1-\alpha-\gamma$, even under heteroskedastic, clustered, or serially correlated data. The idea behind this method is as 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 CS with coverage level $1-\alpha-\gamma$ should be contained in the Wald (nonrobust) CS with coverage level $1-\alpha$ when instruments are strong. We can thus assess the strength of identification by checking how much larger γ needs to be for

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 two-stage least-squares (2SLS) estimator not to exceed 10% of the bias of the ordinary least-squares 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 generalized method of moments (GMM) model, I implement this valid two-step procedure using a classical minimum distance (MD) model in Stata following prior work by Finlay and Magnusson (2009). I establish validity of the two-step procedure using the MD approach in the online appendix.

this containment to hold, and we can bound coverage distortions accordingly. I explain this two-step procedure in more detail in section 2.

Building on the existing command `weakiv` (Finlay, Magnusson, and Schaffer 2014), the new command `twostepweakiv` adds the following features: It first adds to `weakiv` the linear combination (LC) test, which is used to form the robust CS mentioned in the previous paragraph. The LC statistic is an LC of the S (that is, Anderson–Rubin [AR]) and K statistics that yields more powerful tests in some cases when identification is weak. I discuss more of its properties in section 3.

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

The next section illustrates the two-step procedures using simulations and shows 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 CSs using `twostepweakiv`, readers can skip section 3 and proceed directly to section 4, which describes the syntax of `twostepweakiv` and demonstrates its usage with an example. Section 5 details the simulation design and presents additional simulation results to illustrate the improved performance of `twostepweakiv`. I refer interested readers to the online appendix for more details and relevant proofs.[3]

2 Constructing valid two-step CSs

Validity of two-step CSs 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 are nonhomoskedastic based on a simulation exercise.

We focus on constructing a two-stage least squares (2SLS) CSs for a linear IV model with a single endogenous regressor and 10 instruments under heteroskedasticity.[4] In each simulation, we vary the identification strength as measured by $\|\boldsymbol{\pi}_0\|$ in the following linear IV model:

3. The relevant ado-file, proofs, and simulation replication do-files are available on https://github.com/lsun20/TwoStep.
4. We replicate the simulation design for moderate endogeneity under heteroskedasticity with 1 endogenous variable and 10 instruments presented in the supplementary appendix to Andrews (2018). The results are based on 2,500 simulations, each with 10,000 observations.

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\beta_0 + \epsilon \\ \mathbf{X} &= \mathbf{Z}\boldsymbol{\pi}_0 + V \end{aligned}$$

Specifically, we fix $\boldsymbol{\pi}_0/\|\boldsymbol{\pi}_0\|$ and $\beta_0 = 0$ while varying $\|\boldsymbol{\pi}_0\| \in [0, 0.11]$. For each value of $\|\boldsymbol{\pi}_0\|$, we construct a nominal 95% Wald CS and a nominal 95% AR CS. The AR CS is robust to weak identification, whereas the Wald CS is not. We also construct a nominal 95% two-step CS based on the F statistic with the critical value $c = 38.54$. For the F statistic, we use the Kleibergen and Paap (2006) *rk* Wald F statistic returned by `ivreg2` (Baum, Schaffer, and Stillman 2002).[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 the results of Stock and Yogo (2005) imply that we cannot reject the null of weak identification or, more accurately, that the Wald CS has coverage less than 90% under homoskedasticity. In this case, we use the AR CS for the two-step CS. If the first-stage F statistic is larger than c, then we reject the null and use the Wald CS. As a result, under homoskedasticity, after we account for 5% errors due to the pretest itself, this choice of critical value ensures coverage no less than 85% for a two-step CS with nominal coverage 95%.

Under nonhomoskedasticity, however, it is unclear whether the two-step CS 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, Schaffer, and Stillman (2007) suggest that users cautiously apply the critical values from Stock and Yogo (2005), which are intended for homoskedastic data.

We study the coverage probability for each of the three CSs under heteroskedasticity at different values of $\|\boldsymbol{\pi}_0\|$ by calculating the probability that the CS includes the true parameter value $\beta = 0$ based on simulations. We expect the nonrobust Wald CS to have poor coverage for small values of $\|\boldsymbol{\pi}_0\|$ and the robust AR CS to have coverage at least 95% for all values of $\|\boldsymbol{\pi}_0\|$.

In figure 1, we plot the coverage of these CSs at each $\|\boldsymbol{\pi}_0\|$ against the mean of the F statistic at the same $\|\boldsymbol{\pi}_0\|$ as we vary $\|\boldsymbol{\pi}_0\| \in [0, 0.11]$. As expected, the nominal 95% Wald CSs have actual coverage probability much smaller than 95% when identification is weak, which corresponds to small values of $\|\boldsymbol{\pi}_0\|$ and a small mean F statistic. In contrast, the nominal 95% AR CS has coverage at 95% for all values of $\|\boldsymbol{\pi}_0\|$.

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 $H_0 : \boldsymbol{\pi}_0 = 0$.

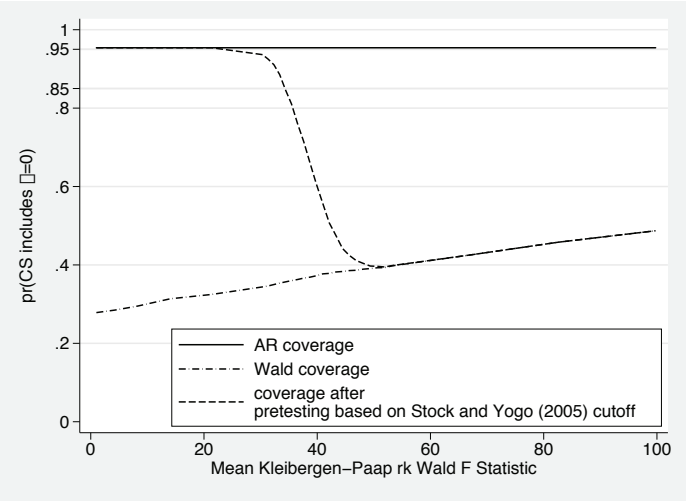

Figure 1. Coverage of two-step CS based on the F statistic

The coverage of the two-step CS (dashed line) drops quickly as $\|\boldsymbol{\pi}_0\|$ 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 CS for the two-step CS. Therefore, the coverage of the Wald CS and the coverage of the two-step CS gradually coincide. However, coverage of the two-step CS is far below 85% above this value of $\|\boldsymbol{\pi}_0\|$.

The issue is that an 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 $c = 38.54$, for example, at 90, many nominal 95% Wald CSs exhibit coverage distortions exceeding 10%. The two-step CS thus has coverage distortion larger than 15%. Our simulation results suggest that the critical value $c = 38.54$ fails to bound its coverage distortion at 15% with heteroskedastic data.

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

This means that to construct a valid two-step CS, we need three inputs: a level $1-\alpha$ nonrobust (Wald) CS CS_N, a level $1-\alpha$ robust CS CS_R, and a preliminary robust CS $\text{CS}_P(\gamma)$, where γ is a specified maximal coverage distortion.

Assumption 1 *The preliminary robust CS $CS_P(\gamma)$ 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 1 regardless of identification strength; and

$CS_P(\gamma)$ is contained in CS_N with probability tending to 1 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 generalized method of moments (GMM) models, I extend his results to minimum distance (MD) models; section 3 details the construction, and I provide proofs in the online appendix.

Given such CS_N, CS_R, and $\text{CS}_p(\gamma)$, the indicator $\mathbf{1}\{\text{CS}_p(\gamma) \not\subseteq \text{CS}_N\}$ is an indicator of weak identification because when the identification is strong, the nonrobust CS_N should contain the robust $\text{CS}_P(\gamma)$. We can construct a two-step CS $\text{CS}_2(\gamma)$ accordingly by

$$\text{CS}_2(\gamma) = \begin{cases} \text{CS}_N & \text{if } \text{CS}_P(\gamma) \subseteq \text{CS}_N \\ \text{CS}_R & \text{if } \text{CS}_P(\gamma) \not\subseteq \text{CS}_N \end{cases} \tag{1}$$

The two-step CS $\text{CS}_2(\gamma)$ is CS_N when $\text{CS}_P(\gamma)$ is contained in CS_N, and CS_R otherwise. Furthermore, because $\text{CS}_P(\gamma) \subseteq \text{CS}_2(\gamma)$ by construction, $\text{CS}_2(\gamma)$ has coverage of 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 $\text{CS}_2(\gamma)$, is at most γ.

2.1 Specify coverage distortion γ

The approach outlined above involves the maximal coverage distortion γ. While researchers can specify a γ in forming the $\text{CS}_2(\gamma)$, different readers may prefer different choices for the maximal coverage distortion. To avoid making a choice on γ, I report both CS_R and CS_N with some indication of the reliability of CS_N, and readers can decide which CSs to focus on based on their preference for coverage distortion.

Start with some $\gamma_{\min}$. `twostepweakiv` allows several levels of $\gamma_{\min}$, and I will discuss the choice of $\gamma_{\min}$ later in this section. Note that we can construct robust CSs $\text{CS}_P(\gamma)$ that decrease in γ in the sense that $\text{CS}_P(\widetilde{\gamma}) \subseteq \text{CS}_p(\gamma)$ for $\widetilde{\gamma} \geq \gamma \geq \gamma_{\min}$. We can also construct CS_R such that $\text{CS}_P(\gamma_{\min}) \subseteq \text{CS}_R$ and thus $\text{CS}_P(\gamma) \subset \text{CS}_R$ for $\gamma \geq \gamma_{\min}$. Define $\widehat{\gamma}$ to be the smallest coverage distortion, $\gamma \geq \gamma_{\min}$, such that $\text{CS}_P(\gamma) \subseteq \text{CS}_N$. Intuitively, $\widehat{\gamma}$ is the smallest distortion γ such that we would report CS_N for $\text{CS}_2(\gamma)$, and we can use $\widehat{\gamma}$ to guide the choice of CS to focus on.

If you prefer a maximal coverage distortion γ, then you can adopt the following decision rule on which CS to focus on:

$$\text{CS}_2(\gamma) = \begin{cases} \text{CS}_N & \text{if } \widehat{\gamma} \leq \gamma \\ \text{CS}_R & \text{if } \widehat{\gamma} > \gamma \end{cases} \tag{2}$$

We refer to $\widehat{\gamma}$ as the distortion cutoff. If you prefer a coverage distortion greater than this cutoff $\gamma \geq \widehat{\gamma}$, we have $\text{CS}_P(\gamma) \subseteq \text{CS}_p(\widehat{\gamma})$. Because $\text{CS}_P(\widehat{\gamma}) \subseteq \text{CS}_N$, this means that $\text{CS}_P(\gamma) \subseteq \text{CS}_N$ as well, and you should focus on CS_N. If you instead prefer a coverage distortion less than this cutoff $\gamma < \widehat{\gamma}$, we have $\text{CS}_P(\widehat{\gamma}) \subseteq \text{CS}_P(\gamma)$. Because $\widehat{\gamma}$ is the smallest coverage distortion $\widetilde{\gamma}$ such that $\text{CS}_P(\widetilde{\gamma}) \subseteq \text{CS}_N$, this means that $\text{CS}_P(\gamma) \nsubseteq \text{CS}_N$. The decision rule in (2) is thus the same as the definition in (1), and the resulting $\text{CS}_2(\gamma)$ has asymptotic coverage of at least $1-\alpha-\gamma$ under both weak and strong identification. Thus, by reporting $(\text{CS}_N, \text{CS}_R, \widehat{\gamma})$, we provide all the ingredients to construct two-step CSs with any level of maximal coverage distortion γ that the reader prefers.

To implement this decision rule, given $\gamma_{\min}$, we first construct CS_R and $\text{CS}_p(\gamma_{\min})$ based on the LC test such that $\text{CS}_P(\gamma)$ decreases in γ for $\gamma \geq \gamma_{\min}$ and $\text{CS}_p(\gamma_{\min}) \subseteq \text{CS}_R$. Then, we can calculate $\widehat{\gamma}$ by solving $\min_{\gamma \geq \gamma_{\min}} \text{CS}_P(\gamma) \subseteq \text{CS}_N$. So the only choice researchers need to make is $\gamma_{\min}$.

The choice of $\gamma_{\min}$ affects how we construct CS_R. In particular, for $\gamma_{\min} = 0$, CS_R degenerates to a robust CS based on the K test, which has some undesirable properties that we discuss in section 3. Thus, we impose that $\gamma_{\min} > 0$. Because $\widehat{\gamma} \geq \gamma_{\min}$, readers who prefer a distortion smaller than $\gamma_{\min}$ would always focus on CS_R. Thus, a larger $\gamma_{\min}$ would point readers more often to CS_R. `twostepweakiv` pretabulates several $\gamma_{\min}$ and sets the default $\gamma_{\min}$ to be 5%. However, we focus on these $\gamma_{\min}$ not to endorse 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 $\text{CS}_2(\gamma)$ with $\gamma = 10\%$ based on (2), where CS_N is a nominal 95% Wald CS and CS_R is a nominal 95% LC CS. We set $\gamma_{\min}$ to be 5%. In figure 2, we plot the coverage of these CSs 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, $\text{CS}_2(10\%)$ has coverage probability of at least 85% regardless of the value of the F statistic.

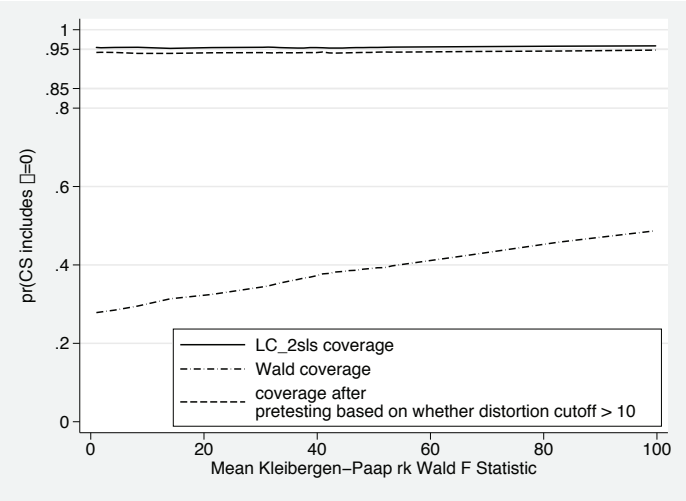

Figure 2. Coverage of two-step CS based on whether distortion cutoff $\widehat{\gamma} > 10\%$

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

To form a two-step CS with bounded coverage distortion as discussed in the previous section, we just need to construct $(\mathrm{CS}_N, \mathrm{CS}_R, \widehat{\gamma})$. This section describes how we construct them for linear IV models based on the MD approach. Specifically, we provide more details on constructing CS_R based on the LC test and calculating $\widehat{\gamma}$. Readers can proceed directly to section 4 to learn how to implement the two-step CSs using `twostepweakiv`.

We start by considering a linear IV model

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\theta}_0 + \boldsymbol{\epsilon} \\ \mathbf{X} &= \mathbf{Z}\boldsymbol{\pi}_0 + \mathbf{V} \end{aligned}$$

which can equivalently be written as a reduced-form model

$$\begin{aligned} \mathbf{Y} &= \mathbf{Z}\boldsymbol{\delta}_0 + \mathbf{U} \\ \mathbf{X} &= \mathbf{Z}\boldsymbol{\pi}_0 + \mathbf{V} \end{aligned}$$

for $\mathbf{Z}$, an $N \times k$ matrix of instruments; $\mathbf{X}$, an $N \times m$ matrix of endogenous regressors; and $\mathbf{Y}$, an $N \times 1$ vector of outcome variables. Assume that $E(\mathbf{Z}_i U_i) = E(\mathbf{Z}_i \mathbf{V}_i^{'}) = \mathbf{0}$. Note that $\boldsymbol{\delta}_0 = \boldsymbol{\pi}_0 \boldsymbol{\theta}_0$. We also assume that any exogenous regressors have already been partialled out.[6]

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

Suppose we are interested in a p-dimensional parameter ($p \leq m$) $\boldsymbol{\beta} = f(\boldsymbol{\theta})$. Here f is a continuously differentiable function. If we are interested in $\boldsymbol{\theta}$, then let f be the identity matrix. Denote $\partial/(\partial \boldsymbol{\theta}^{'}) f(\boldsymbol{\theta}) = F(\boldsymbol{\theta})$. Assume $F(\boldsymbol{\theta}_0)$ has full rank. For example, we may be interested in constructing a CS for the jth coordinate of the parameter vector and $f(\boldsymbol{\theta}) = \theta_j$. We focus on constructing a CS for a single coordinate of the parameter vector in section 3.1. Below, we discuss the general case where $\boldsymbol{\beta}$ is p-dimensional and first define its MD estimators.

The reduced-form parameter vector is $\boldsymbol{\tau} = (\boldsymbol{\delta}, \boldsymbol{\pi})$, and its estimator is $\widehat{\boldsymbol{\tau}} = (\widehat{\boldsymbol{\delta}}, \widehat{\boldsymbol{\pi}})$. The structural parameter vector is $\boldsymbol{\theta}$. Instead of deriving GMM estimators based on the moment function $g(\boldsymbol{\theta}) = (1/N)\sum_i \mathbf{Z}_i(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\theta}_0)$, the MD approach focuses on the structural function, which is $r(\boldsymbol{\tau}, \boldsymbol{\theta}) = \boldsymbol{\delta} - \boldsymbol{\pi}\boldsymbol{\theta}$. Note that $r(\widehat{\boldsymbol{\tau}}, \boldsymbol{\theta}_0) = \widehat{\boldsymbol{\delta}} - \widehat{\boldsymbol{\pi}}\boldsymbol{\theta}_0$ and $r(\boldsymbol{\tau}_0, \boldsymbol{\theta}_0) = \boldsymbol{\delta}_0 - \boldsymbol{\pi}_0\boldsymbol{\theta}_0 = \mathbf{0}$. If $\{\partial r(\boldsymbol{\tau}_0, \boldsymbol{\theta}_0)\}/(\partial\boldsymbol{\theta})$ does not have full rank, then the model is underidentified. If $\{\partial r(\boldsymbol{\tau}_0, \boldsymbol{\theta}_0)\}/(\partial\boldsymbol{\theta})$ has full rank, then the model is identified. Weak identification arises when $\{\partial r(\boldsymbol{\tau}_0, \boldsymbol{\theta}_0)\}/(\partial\boldsymbol{\theta})$ is very close to a reduced rank matrix. Otherwise, the model is well (strongly) identified.

Assume we have some estimator $\widetilde{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ that under strong identification is first-order equivalent to $\widehat{\boldsymbol{\theta}}$, which solves $\min_{\boldsymbol{\theta}\in\Theta} r(\widehat{\boldsymbol{\tau}}, \boldsymbol{\theta})^{'} \widehat{\Omega}(\boldsymbol{\theta}) r(\widehat{\boldsymbol{\tau}}, \boldsymbol{\theta})$ for $\widehat{\Omega}(\boldsymbol{\theta})$, a symmetric positive definite weighting matrix that converges uniformly to $\Omega(\boldsymbol{\theta})$ 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 $f(\widetilde{\boldsymbol{\theta}})$, the estimator for $\boldsymbol{\beta}$.

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

$$\sqrt{N}\begin{pmatrix} \widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0 \\ \widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0 \end{pmatrix} \to_d N\left\{\mathbf{0}, \begin{pmatrix} \boldsymbol{\Sigma}_\delta & \boldsymbol{\Sigma}_{\delta\pi} \\ \boldsymbol{\Sigma}_{\pi\delta} & \boldsymbol{\Sigma}_\pi \end{pmatrix}\right\}$$

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

Then, by the delta method, we have $\sqrt{N}\{r(\widehat{\boldsymbol{\tau}}, \boldsymbol{\theta}_0) - r(\boldsymbol{\tau}_0, \boldsymbol{\theta}_0)\} \to_d N(\mathbf{0}, \boldsymbol{\Sigma}_r)$. Let $\widehat{\boldsymbol{\Sigma}}_r(\boldsymbol{\theta}_0)$ be a consistent estimator for $\boldsymbol{\Sigma}_r$.

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

$$\sqrt{N}\left\{f\left(\widetilde{\boldsymbol{\theta}}\right) - \boldsymbol{\beta}_0\right\} \to_d N(\mathbf{0}, \mathbf{V}_\beta)$$

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

where $\boldsymbol{\beta}_0 = f(\boldsymbol{\theta}_0)$ and $\mathbf{V}_\beta = F(\boldsymbol{\theta}_0) V_{\boldsymbol{\theta}} F(\boldsymbol{\theta}_0)'$. The asymptotic normality of $f(\widetilde{\boldsymbol{\theta}})$ under strong identification means that the Wald statistic

$$W(\boldsymbol{\beta}) = N \times \left\{ f\left(\widetilde{\boldsymbol{\theta}}\right) - \boldsymbol{\beta} \right\}' \widehat{\mathbf{V}}_\beta^{-1} \left\{ f\left(\widetilde{\boldsymbol{\theta}}\right) - \boldsymbol{\beta} \right\}$$

for $\widehat{\mathbf{V}}_\beta$, a consistent estimator of $\mathbf{V}_\beta$, has an asymptotic distribution

$$W(\boldsymbol{\beta}_0) \to_d \chi^2_p$$

under H_0: $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ and strong identification. Collecting the set of values $\boldsymbol{\beta}$ such that the Wald test does not reject it gives us a nonrobust Wald CSs for $\boldsymbol{\beta} = f(\boldsymbol{\theta})$,

$$\text{CS}_N = \left\{ \boldsymbol{\beta} \colon W(\boldsymbol{\beta}) \leq \chi^2_{p,1-\alpha} \right\}$$

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

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

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

$$S(\boldsymbol{\theta}) = N \times r(\widehat{\boldsymbol{\tau}}, \boldsymbol{\theta})' \widehat{\boldsymbol{\Sigma}}_r(\boldsymbol{\theta})^{-1} r(\widehat{\boldsymbol{\tau}}, \boldsymbol{\theta})$$

The LC statistic is defined as

$$\text{LC}_{\Omega,f,a}(\boldsymbol{\theta}) = K_{\Omega,f}(\boldsymbol{\theta}) + a \times S(\boldsymbol{\theta})$$

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 CS based on the K statistic always contains all local minimums and maximums of the continuously updating objective function. Thus, the K test can have low power if the continuously updating GMM objective has multiple local minimums or maximums. By combining the K statistic with the S statistic, the LC statistic improves the power of the K test. In the online appendix, I show that under both strong and weak identification, under $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, $\text{LC}_{\Omega,f}(\boldsymbol{\theta})$ has the following asymptotic distribution:

$$\text{LC}_{\Omega,f,a}(\boldsymbol{\theta}_0) \to_d (1+a) \times \chi^2_p + a \times \chi^2_{k-p}$$

To form CS_R and $\text{CS}_p(\gamma)$ based on $\text{LC}_{\Omega,f}(\boldsymbol{\theta})$, we let a be a function of γ. Specifically, for a given value of γ, define $a(\gamma)$ by

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

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

$$\begin{aligned} \text{CS}_{R,\boldsymbol{\theta}} &= \left[\boldsymbol{\theta} : K_{\Omega,f}(\boldsymbol{\theta}) + a(\gamma)\times S(\boldsymbol{\theta}) \leq H^{-1}\{1-\alpha; a(\gamma), k, p\}\right] \\ \text{CS}_R &= \{f(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \text{CS}_{R,\boldsymbol{\theta}}\} \\ &= \left[\boldsymbol{\beta} : \min_{\boldsymbol{\theta}:\boldsymbol{\beta}=f(\boldsymbol{\theta})} K_{\Omega,f}(\boldsymbol{\theta}) + a(\gamma)\times S(\boldsymbol{\theta}) \leq H^{-1}\{1-\alpha; a(\gamma), k, p\}\right] \end{aligned} \tag{3}$$

where $H^{-1}\{1-\alpha; a(\gamma), k, p\}$ is the $1-\alpha$ quantile of a $\{1+a(\gamma)\}\times\chi^2_p + a(\gamma)\times\chi^2_{k-p}$ distribution. The initial CS, $\text{CS}_{R,\boldsymbol{\theta}}$, collects the set of values $\boldsymbol{\theta}$ where the LC statistic falls below the critical value and so will cover $\boldsymbol{\theta}_0$ with probability tending to $1-\alpha$ under both strong and weak identification. Similarly to the conventional projection method, CS_R takes the image of the initial CS under $f(\cdot)$ to construct a CS for $\boldsymbol{\beta} = f(\boldsymbol{\theta})$. Unlike the conventional projection method, we alter the test statistic (specifically the K statistic in the LC) based on $f(\cdot)$. Thus, we refer to this method as the refined projection method. I discuss this refinement more in section 3.1.

We can similarly define $\text{CS}_P(\gamma)$, where the critical value in (3) is replaced with $\chi^2_{p,1-\alpha}$. We can show that the above CS_R and $\text{CS}_p(\gamma)$, with the Wald CS_N, satisfy assumption 1 and thus can be used to construct $\text{CS}_2(\gamma)$ with coverage probability of at least $1-\alpha-\gamma$.

I demonstrated in section 2 that by reporting $(\text{CS}_N, \text{CS}_R, \widehat{\gamma})$, we provide all the ingredients needed to construct $\text{CS}_2(\gamma)$ with any level of maximal coverage distortion γ that the reader prefers. Here we describe how to calculate the distortion cutoff $\widehat{\gamma}$. Andrews (2018) shows that to find $\widehat{\gamma}$ that solves $\min_{\gamma\geq\gamma_{\min}} \text{CS}_P(\gamma) \subseteq \text{CS}_N$, we first need to find

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

and then calculate

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

In practice, $\Pr\left\{(1+\widetilde{a})\times\chi^2_p + \widetilde{a}\times\chi^2_{k-p} \leq \chi^2_{p,1-\alpha}\right\}$ is again calculated based on 100,000 independent simulation draws. The distortion cutoff is then $\widehat{\gamma} = \max\{\widetilde{\gamma}, \gamma\}$.

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

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

To find this value in practice, we take 100,000 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 LC of these draws to $\chi^2_{p,1-\alpha}$. The simulation can be computationally intensive, and we pretabulate $a(\gamma)$ for a few levels of γ in `twostepweakiv`. For levels of γ that are not pretabulated, `twostepweakiv` solves for $a(\gamma)$ based on 10,000 simulations before it calculates CSs.

3.1 Refined robust CSs for a single parameter

Let $\boldsymbol{\theta} = (\beta_1, \ldots, \beta_m)$ be the full-parameter vector. In this subsection, I discuss how to construct a two-step CS for the jth coordinate β_j because we are interested only in the causal effect of its corresponding regressor. The remaining parameters are called nuisance parameters because they are not of immediate interest. This can be achieved by setting $f(\boldsymbol{\theta}) = \mathbf{e}_j' \times \boldsymbol{\theta}$, where $\mathbf{e}_j$ is the jth coordinate vector. Then, we have $f(\boldsymbol{\theta}) = \mathbf{e}_j' \times \boldsymbol{\theta} = \beta_j$ and $F(\boldsymbol{\theta}) = \partial/(\partial\boldsymbol{\theta}) f(\boldsymbol{\theta}) = \mathbf{e}_j'$. Below, we derive test statistics for a single element of the parameter vector.

It is straightforward to construct nonrobust Wald CS for β_j—simply its estimator $\widetilde{\beta}_j$ 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 CSs are based on test inversion on a fine grid of parameters. Namely, we test all potential hypotheses $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ and define our CS as the set of all $\boldsymbol{\theta}_0$ for which the hypothesis is accepted. As a result of such an inversion, we would end up with a joint (multidimensional) CS for $\boldsymbol{\beta}$ 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 CS 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 do not want to assume the nuisance parameters are well identified, the conventional method (the projection method) projects the original CS for the full-parameter vector onto the jth dimension. The resulting CS 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 CS and then project it onto the jth 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 CSs 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 CS for β_j based on the K test, we refine the K statistic to be

$$K_{\Omega,\mathbf{e}_j}(\boldsymbol{\theta}) = N \times \frac{\left\{\beta_j^*(\boldsymbol{\theta}) - \beta_j\right\}' \left\{\beta_j^*(\boldsymbol{\theta}) - \beta_j\right\}}{M(\boldsymbol{\theta})' \widehat{\boldsymbol{\Sigma}}_r(\boldsymbol{\theta}) M(\boldsymbol{\theta})}$$

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

$$\mathrm{LC}_{\Omega,\mathbf{e}_j,a}(\boldsymbol{\theta}) = K_{\Omega,\mathbf{e}_j}(\boldsymbol{\theta}) + a \times S(\boldsymbol{\theta})$$

where the S statistic remains the same as in the full-parameter case. This statistic is used to form CS_R in (3). In section 5, I present a simulation study that compares this refined projection method with the conventional projection method.

4 The twostepweakiv command

4.1 Syntax

The command syntax for `twostepweakiv` is

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

With regard to *estimator*, all estimators are formulated as MD estimators following `weakiv` and Finlay and Magnusson (2009).

estimator	Description
`2sls`	2SLS estimator
`liml`	limited-information maximum likelihood
`md2s`	two-step MD estimator
`cue`	continuous updating estimator (CUE)

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

4.2 Options

Below, I describe the options for the command. For each option, I describe when it should be used or adjusted. I 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 CSs based on several test statistics, it calculates the coverage distortion for the LC test only (that is, readers should form two-step CSs based on the LC confidence interval). When there is more than one endogenous variable, users can also specify for which parameter to construct the confidence interval using the refined projection method.

`citestlist(`*testlist*`)` constructs CSs for the full-parameter vector; several tests are inherited from the original `weakiv` command.

Table 1. Tests available for full-parameter vector

testlist	Test	Addition to `weakiv`
`Wald`	nonrobust 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

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`. Because `Wald` is always provided, you do not need to specify this test.

`project(`*varlist*`)`, when there is more than one weak endogenous variable and we are interested in inference for one endogenous regressor β only, calculates CSs using the "refined" projection method for β while treating the other endogenous regressors $\boldsymbol{\eta}$ as free unknown parameters (nuisance parameters). See option `strong()` if you are willing to make additional assumptions about $\boldsymbol{\eta}$.

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

Table 2. Tests available for scalar parameters using refined projection method

project_testlist	Test
`Wald`	nonrobust 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

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`. Because `Wald` is always provided, you do not need to specify this test.

grid_options

Because we construct CSs 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 and 25×25 points for two endogenous variables. With weak or strong instruments, this may often be too small or large a grid to estimate the CSs. I recommend that users specify the grid using the `gridmin()` and `gridmax()` options.

`cuepoint` adds the CUE estimator to the grid for calculating CSs. If CSs are nonempty, they should contain the CUE estimator.

`gridmin(`*numlist*`)` sets the lower limit or limits for grid search (in dimensions corresponding to endogenous regressors).

`gridmax(`*numlist*`)` sets the upper limit or limits for grid search (in dimensions corresponding to endogenous regressors).

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

`gridpoints(`*numlist*`)` specifies the numbers of equally spaced grid points (in dimensions corresponding to endogenous regressors) over which to calculate the CSs. The default is `gridpoints(100 25 11 7 5)` for the cases of one, two, three, four, or five endogenous regressors, respectively. For testing a point null hypothesis, for example, 0, set the grid point to 1 and `gridmin()`=`gridmax()`=0.

size_options

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

`level(#)` specifies the confidence level as a percentage (same for all tests performed). The default is `level(95)` for size 5% tests; if specified, `level(#)` with values 99, 95, or 90 allows for faster computation because weights used in `LC` and `LC_2sls` tests are pretabulated only for these values. For other values, see `gammalevel(#)`.

`gammalevel(#)` specifies $\gamma_{\min}$. Calculation is faster for $\gamma_{\min} = 1,\ 2,\ 5,\ 10,\ 15$, or 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. Because simulations are computationally intensive, we tabulate the weights $a(\gamma)$ and critical values $H^{-1}\{1-\alpha; a(\gamma), k, p\}$ for $p = 1, \ldots, 5$, $k = 1, \ldots, 50$, $\alpha = 90,\ 95$, or 99 and $\gamma = 1,\ 2,\ 5,\ 10,\ 15$, or 20 in advance. For values not pretabulated, 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 CSs for one endogenous regressor with coefficient β only. If we assume the coefficients $\boldsymbol{\eta}$ on the other regressors are strongly identified as specified in `strong(`*varlist*`)`, then we can plug in their estimates and construct robust CSs for β based on a 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 $\boldsymbol{\eta}$ when there is more than one weak endogenous variable and treats the coefficients $\boldsymbol{\beta}$ on the rest of the endogenous regressors as potentially weakly identified. By default, all endogenous regressors are assumed to be potentially weakly identified. If `strong()` is specified and `cuestrong` is not evoked, then at each grid point $\boldsymbol{\beta}_0$ for the potentially weakly identified parameters $\boldsymbol{\beta}$, we calculate `md2s` estimates for strongly identified parameters $\boldsymbol{\eta}$ under the null hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. We then evaluate test statistics for $\boldsymbol{\beta}$, plugging in the `md2s` estimates $\widehat{\boldsymbol{\eta}}$.

`cuestrong` uses CUE point estimates for strongly identified endogenous regressors (specified in `strong(`*varlist*`)`) and includes these point estimates in the grid. That is, at each grid point $\boldsymbol{\beta}_0$ for the potentially weakly identified parameter $\boldsymbol{\beta}$, we calculate CUE for strongly identified parameters $\boldsymbol{\eta}$ under the null hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. We then evaluate test statistics for $\boldsymbol{\beta}$, plugging the CUE estimates $\widehat{\boldsymbol{\eta}}$ in the strongly identified parameters. Note that this option may be computationally intensive.

4.3 Stored results

`twostepweakiv` stores the following in `e()`:

Scalars

`e(N)`	sample size
`e(N_clust)`	number of clusters (if cluster–robust VCE used)
`e(kwt)`	weight on K in $K - J$ test
`e(endo_ct)`	number of endogenous regressors
`e(wendo_ct)`	number of weakly identified endogenous regressors
`e(sendo_ct)`	number of strongly identified endogenous regressors
`e(overid)`	degree of overidentification
`e(small)`	`1` if small-sample adjustments used, `0` otherwise
`e(ar_level)`	level in percent used for `AR` confidence interval
`e(k_level)`	level in percent used for `K` confidence interval
`e(wald_level)`	level in percent used for `Wald` confidence interval
`e(gamma_level)`	gamma_min used for `LC_2sls` confidence interval
`e(gamma_hat)`	distortion cutoff for `LC_2sls` confidence interval
`e(points)`	number of points in grid used to estimate confidence sets
`e(clrsims)`	number of draws used in simulations to obtain p-values for conditional likelihood-ratio test

Macros

`e(cmd)`	`twostepweakiv`
`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
`e(p`*xx*`_`*yy*`_cset)`	as above, projection-based confidence set for variable *xx*, test *yy*
`e(inexog)`	list of exogenous regressors (excluding any included in the tests)
`e(exexog)`	list of excluded instruments
`e(depvar)`	dependent variable
`e(endo)`	endogenous variable or variables
`e(wendo)`	weakly identified endogenous variable or variables
`e(sendo)`	strongly identified endogenous variable or variables
`e(pwendo)`	endogenous variables with projection-based confidence sets
`e(pwendo_nlist)`	corresponding numbers for `e(pwendo)`; used to identify projection-based confidence sets in `e()` (see above)
`e(gridpoints)`	list of grid points in each dimension
`e(model)`	linear IV
`e(waldcmd)`	*iv_command* estimator used to fit standard IV model: `2sls`, `md2s`, `liml`, or `cue`
`e(level)`	default confidence level in percent used for tests of null $= 100 \times (1 - \alpha)$
`e(method)`	`md`

Matrices

`e(citable)`	table with test statistics, p-values, and rejection indicators for every grid point over which hypotheses are tested; if `strong()` is used, the estimated coefficients for the strongly identified regressors are also recorded
`e(p`*xx*`citable)`	grid table with rejection indicators for projection-based inference; *xx* will be one or two numbers corresponding to the endogenous regressor or regressors
`e(F)`	first-stage F statistics for all endogenous regressors
`e(wbeta)`	weakly identified coefficients from IV model used for `Wald` tests
`e(var_wbeta)`	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

4.4 Example: Married female labor market participation

We demonstrate the use of this command with data from Mroz (1987). 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 and 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 CS for the effect of wages on labor supply based on the `2sls` estimator. In the command, we specify the estimator to be `2sls`, followed by the regression specification. We let `twostepweakiv` construct robust CSs 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 an inefficient weight matrix because it is used by the `2sls` estimator. While only the `LC_2sls` CS (with the Wald CS) is needed to construct a two-step identification robust CS, we report the other two CSs for reference.

Recall that these robust CSs are based on test inversion. Instead of letting the command `twostepweakiv` use the default *grid_options*, we specify the grid as $[-1000, 8000]$. While a fine grid takes longer to compute, to calculate the robust CSs and the distortion cutoff $\widehat{\gamma}$, we need a rather fine and wide grid. We use the default *size_options*, which sets the test size $\alpha = 5\%$ and the initial coverage distortion $\gamma_{\min} = 5\%$. Because there is only one endogenous variable, there are 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 expersq 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 CSs. The negative values of the `K_2sls` CS 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 nonrobust Wald CS. Compared with the results shown in Finlay and Magnusson (2009), our Wald CS 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 because of the MD approach we take.

Finally, we explain how to form a two-step identification robust CS based on the above results. Recall that we need three ingredients: the distortion cutoff $\widehat{\gamma}$ that indicates identification strength, the robust `LC_2sls` CS, and the nonrobust Wald CS. The last 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. If 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 CS with coverage distortion bounded by 10%, the reader 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 the reader should focus on the Wald CS.

The rest of the caption reminds the readers that we calculate MD versions of robust tests as well as the nonrobust Wald test, all test statistics use heteroskedasticity-robust VCE estimators, and the Wald CS 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 is more than one endogenous regressor.

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 CS onto the subset. In this section, I 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

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\theta}_0 + \boldsymbol{\epsilon} \\ \mathbf{X} &= \mathbf{Z}\boldsymbol{\pi}_0 + \mathbf{V} \end{aligned}$$

which can equivalently be written as a reduced-form model

$$\begin{aligned} \mathbf{Y} &= \mathbf{Z}\boldsymbol{\delta}_0 + \mathbf{U} \\ \mathbf{X} &= \mathbf{Z}\boldsymbol{\pi}_0 + \mathbf{V} \end{aligned}$$

for $\mathbf{Z}$, an $N \times k$ matrix of instruments; $\mathbf{X}$, an $N \times m$ vector of endogenous regressors; $\mathbf{Y}$, an $N \times 1$ vector of outcome variables. We set $k = 5$. We take $\mathbf{Z}_i \sim N(0, I)$. Because we are interested in the performance of projection tests, we set $m = 2$. Let $\boldsymbol{\theta}_0 = (\eta_0, \beta_0)$. We are interested in constructing CSs for the scalar coefficient β, treating η as a nuisance parameter. We set $N = 500$.

To simulate heteroskedastic data, we draw

$$\left(\begin{array}{c} \epsilon_i V_i \end{array} \right) | \mathbf{Z}_i \sim N \left[\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \boldsymbol{\Sigma}_V \times \exp\{(1/2)\mathbf{Z}_i^1\} \right]$$

where $\mathbf{Z}_i^1$ is the first instrument and $\boldsymbol{\Sigma}_V = \left(\begin{array}{cc} 1 & 0.5 \\ 0.5 & 1 \end{array} \right)$.

Because we are interested in the performance of projection tests in this section, we set $m = 2$. Let $\boldsymbol{\theta}_0 = (\eta_0, \beta_0)$. We are interested in constructing CSs for the scalar coefficient β, treating η as a nuisance parameter. For weak identification, we set $\boldsymbol{\pi}_0' = \left(\begin{array}{ccccc} 0.1 & 0.2 & 0.3 & 0.4 & 0.5 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \end{array} \right)$. For strong identification, we set $\boldsymbol{\pi}_0' = \left(\begin{array}{ccccc} 0.1 & 0.2 & 0.3 & 0.4 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \end{array} \right)$. Note that in the weak identification case that we consider, the local rank reduction in the matrix of the first-stage coefficients $\boldsymbol{\pi}_0$ is just from one first stage, $\boldsymbol{\pi}_0^2$, being close to zero, not from $\boldsymbol{\pi}_0^1$ and $\boldsymbol{\pi}_0^2$ being almost the same. Thus, we report the mean heteroskedasticity-robust F statistic for each of the first-stage regressions, $F_j = (N/k)\widehat{\boldsymbol{\pi}}_j' \widehat{\boldsymbol{\Sigma}}_{\widehat{\pi}_j}^{-1} \widehat{\boldsymbol{\pi}}_j$ for $j = 1, 2$, $\widehat{\boldsymbol{\pi}}_j = \left(\mathbf{Z}'\mathbf{Z} \right)^{-1} \mathbf{Z}'\mathbf{X}^j$, the ordinary least-squares estimator, for $\boldsymbol{\pi}_0^j$, and $\widehat{\boldsymbol{\Sigma}}_{\widehat{\pi}_j}$, a heteroskedasticity-robust estimator, for the variance of $\sqrt{N}\left(\widehat{\boldsymbol{\pi}}_j - \boldsymbol{\pi}_0^j \right)$.

We compare the 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 $b/\|\boldsymbol{\pi}_2\|$ 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 12 times $\|\boldsymbol{\pi}_2\|$; for β, we include only one grid point at 0. We use `2sls` estimators for point estimates $\widehat{\beta}$ and $\widehat{\eta}$.

We simulate the performance of conventional and refined projection methods for the `LC_2sls` test. Recall that the `LC_2sls` test is an LC test with inefficient weights. Because the point estimates are estimated by `2sls` estimators, we use the `LC_2sls` test because it also uses inefficient weight matrices. We include the performance of the `K_2sls` test to illustrate its spurious power losses.[9]

We scale b inversely by $\|\boldsymbol{\pi}_2\|$ for the power curves to be comparable.

Figures 3 and 4 report the powers of the `LC_2sls` test and `K_2sls` test coupled with either conventional or refined projection methods under weak and strong identification, respectively. For the `LC_2sls` test, the refined projection method yields higher power than the conventional projection method in all cases. The `K_2sls` test is in general not consistent, even under strong identification.

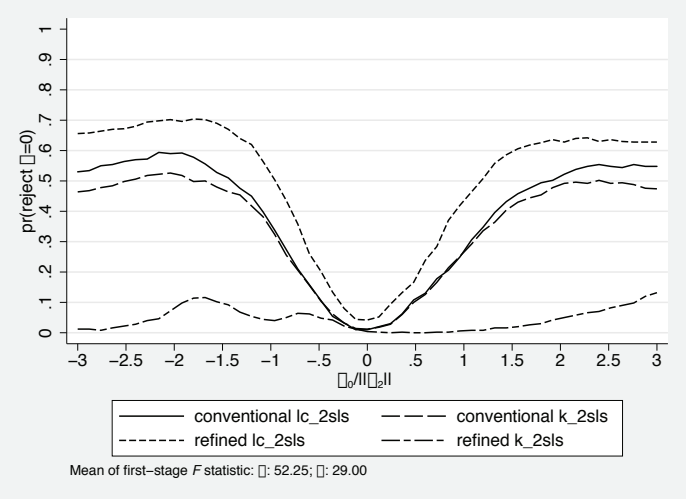

Figure 3. Comparison between conventional and refined projection methods under weak identification

9. Even in strong identified models, CSs based on a K test are not necessarily consistent for $\boldsymbol{\theta}_0$. This issue is more pronounced for the refined projection method than for the conventional projection method. See section 3.2 of Andrews (2018) for more details.

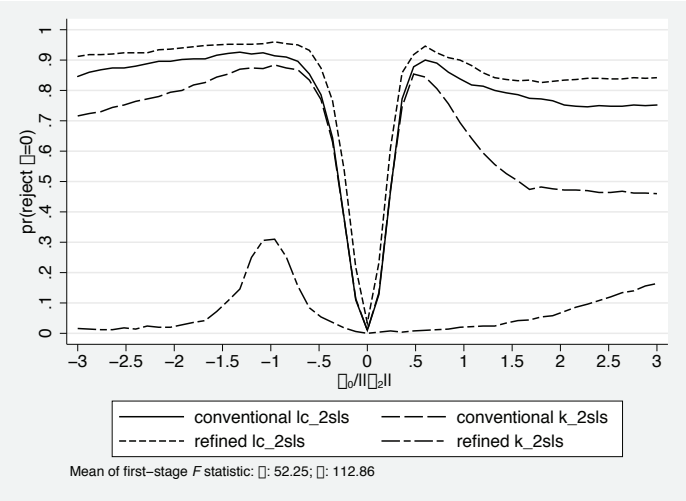

Figure 4. Comparison between conventional and refined projection methods under strong identification

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: 2155–2182.

———. 2018. Valid two-step identification-robust confidence sets for GMM. *Review of Economics and Statistics* 100: 337–348.

Baum, C. F., M. E. Schaffer, and S. Stillman. 2002. ivreg2: Stata module for extended instrumental variables/2SLS and GMM estimation. Statistical Software Components S425401, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s425401.html.

———. 2007. Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata Journal* 7: 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: 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: 398–421.

Finlay, K., L. M. Magnusson, and M. E. Schaffer. 2014. weakiv: Stata module to perform weak-instrument-robust tests and confidencce intervals for instrumental-variable (IV) estimation of linear, probit and tobit models. Statistical Software Components S457684, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s457684.html.

Kleibergen, F. 2004. Testing subsets of structural parameters in the instrumental variables regression model. *Review of Economics and Statistics* 86: 418–423.

———. 2005. Testing parameters in GMM without assuming that they are identified. *Econometrica* 73: 1103–1123.

Kleibergen, F., and R. Paap. 2006. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics* 133: 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: 765–799.

Stock, J. H., 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. K. Andrews and J. H. Stock, 80–108. New York: Cambridge University Press.

About the author

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