Robust Two Step Confidence Sets, and the Trouble with the First Stage F Statistic

Isaiah Andrews
Discussion by Bruce Hansen

September 27, 2014
Classic Weak IV

- $Y = X\theta + e$, $\theta$ scalar
- $X = Z\pi + \nu$
- $k$ instruments
- Goal: 95% confidence set (CS) for $\theta$
- Three contributions:
  1. Numerically demonstrates that Stock-Yogo two-step method fails miserably under heteroskedasticity
  2. Proposes valid two-step CS with bounded size distortion
  3. Introduces diagnostic for degree of non-robust size distortion
Stock-Yogo Two-Step

- $CS_{NR} = \hat{\theta} \pm 2\sigma_{\hat{\theta}}$
- $CS_{R} = \{ \theta : S(\theta) \leq \chi_{k,1-\alpha}^2 \}$ where $S(\theta)$ is AR-like statistic
- $F = T \hat{\pi}' \hat{\Sigma}_{\pi}^{-1} \hat{\pi} / k$, first-stage $F$
- $\gamma = \text{maximum allowed size distortion}$
- $CS_{SY} = \begin{cases} 
CS_{NR} & \text{if } F > c(\gamma) \\
CS_{R} & \text{if } F < c(\gamma) 
\end{cases}$
  where $c(\gamma)$ is selected so that the maximum size distortion is bounded below $\gamma$
- Assumes homoskedasticity
Suppose we construct $CS_{SY}$ using hetero-robust variances, but use Stock-Yogo critical values $c(\gamma)$.

What is the worst-case coverage?

Numerical finding: coverage is 0%

Conclusion: Stock-Yogo method fails miserably under heteroskedasticity
Consider 2SLS under weak IV and heteroskedastic. Suppose

\[
\frac{1}{T} Z' Z \quad \rightarrow \quad I_k \\
\frac{1}{\sigma_e \sqrt{T}} Z' e \quad \rightarrow \quad \zeta_2 \\
\frac{1}{\sigma_v \sqrt{T}} Z' X \quad \rightarrow \quad \lambda + \zeta_1
\]

with \((\zeta_1, \zeta_2)\) jointly normal

Then as in Stock-Staiger, for 2SLS

\[
\hat{\beta} - \beta \quad \rightarrow \quad \frac{\sigma_u \nu_2}{\sigma_v \nu_1} \\
\nu_2 = (\lambda + \zeta_1)' \zeta_2 \\
\nu_1 = (\lambda + \zeta_1)' (\lambda + \zeta_1)
\]
In homoskedastic case the asymptotic distribution $v_2 / v_1$ only depends on $k$ and the concentration parameter $\lambda'\lambda$.

This is because the covariance matrix of $(\zeta_1, \zeta_2)$ takes a simple form.

Under heteroskedasticity, the distribution is much more complicated, due to the covariance matrix.

Still, “weak identification” appears to be a problem due to small values of the denominator $v_1 = (\lambda + \zeta_1)'(\lambda + \zeta_1)$.

An estimator of $E(v_1)$ under homo- or heteroskedasticity is the classic first-stage F statistic.

- Not the hetero-robust F statistic.

Thus to detect weak instruments under heteroskedasticity, it may still be appropriate to examine the classic F.

However the Stock-Yogo critical values will not provide size control.

Recommendation: The Stock-Yogo approach may still work, but developing appropriate cut-offs would be challenging.
**New Two-Step CS**

- Pick $\gamma$, the maximum size distortion (as in Stock-Yogo)
- Set $K_\gamma(\theta) = K(\theta) + a(\gamma)S(\theta)$ where $K(\theta)$ is a Kleibergen-like statistic and $S(\theta)$ is the AR-like statistic
- Preliminary robust: $CS_P(\gamma) = \{ \theta : K_\gamma(\theta) \leq \chi^2_{k,1-\alpha} \}$
  - Has coverage exceeding $1 - \alpha - \gamma$
- Robust: $CS_R(\gamma) = \{ \theta : K_\gamma(\theta) \leq H_{k,1-\alpha} \}$, where $H_{k,1-\alpha}$ is $1 - \alpha$ quantile of $(1 + a(\gamma))\chi^2_1 + a(\gamma)\chi^2_{k-1}$
  - Has coverage exceeding $1 - \alpha$
- Two-Step: $CS_2(\gamma) = \begin{cases} CS_{NR} & \text{if } CS_P(\gamma) \subseteq CS_{NR} \\ CS_R(\gamma) & \text{if } CS_P(\gamma) \not\subseteq CS_{NR} \end{cases}$
- Property: $CS_P(\gamma) \subseteq CS_2(\gamma)$
  - So $CS_2(\gamma)$ has coverage exceeding $CS_P(\gamma)$ which exceeds $1 - \alpha - \gamma$
- Under strong identification, $CS_P(\gamma) \subseteq CS_{NR}$, so $CS_2(\gamma) = CS_{NR}$ asymptotically
Define $\hat{\gamma}$ as the smallest $\gamma$ such that $CS_p(\hat{\gamma}) \subseteq CS_{NR}$

Recommends reporting $CS_{NR}$, $CS_R(\hat{\gamma})$, and $\hat{\gamma}$

If $\hat{\gamma}$ is small (e.g. less than 10%), we focus on $CS_{NR}$

If $\hat{\gamma}$ is large (e.g. more than 10%) we focus on $CS_R(\hat{\gamma})$
<table>
<thead>
<tr>
<th>Specification</th>
<th>I</th>
<th></th>
<th>II</th>
<th></th>
<th>III</th>
<th></th>
<th>IV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}$</td>
<td>s.e.</td>
<td>$\hat{\gamma}$</td>
<td>$\hat{\beta}$</td>
<td>s.e.</td>
<td>$\hat{\gamma}$</td>
<td>$\hat{\beta}$</td>
<td>s.e.</td>
</tr>
<tr>
<td>2SLS</td>
<td>0.0990</td>
<td>0.0207</td>
<td>2.5%</td>
<td>0.0806</td>
<td>0.0165</td>
<td>18%</td>
<td>0.0600</td>
<td>0.0292</td>
</tr>
<tr>
<td>CUGMM</td>
<td>0.1000</td>
<td>0.0208</td>
<td>1.9%</td>
<td>0.0857</td>
<td>0.0165</td>
<td>9.5%</td>
<td>0.0613</td>
<td>0.0292</td>
</tr>
<tr>
<td>CUGMM, NW</td>
<td>0.1000</td>
<td>0.0213</td>
<td>1.3%</td>
<td>0.0857</td>
<td>0.0196</td>
<td>1.6%</td>
<td>0.0613</td>
<td>0.0504</td>
</tr>
<tr>
<td>HLIM</td>
<td>0.0999</td>
<td>0.0213</td>
<td>1.3%</td>
<td>0.0838</td>
<td>0.0196</td>
<td>1.5%</td>
<td>0.0574</td>
<td>0.0515</td>
</tr>
<tr>
<td>HFUL</td>
<td>0.0995</td>
<td>0.0210</td>
<td>1.8%</td>
<td>0.0836</td>
<td>0.0194</td>
<td>1.9%</td>
<td>0.0576</td>
<td>0.0496</td>
</tr>
<tr>
<td>First Stage F</td>
<td>30.5822</td>
<td></td>
<td>4.6245</td>
<td></td>
<td>1.5788</td>
<td></td>
<td></td>
<td>1.8232</td>
</tr>
<tr>
<td>Efficient K CS</td>
<td>[0.059,0.144]</td>
<td></td>
<td>[0.0474,0.1261]</td>
<td></td>
<td>[-0.5,0.5]</td>
<td></td>
<td></td>
<td>[0.0581,0.1475]</td>
</tr>
<tr>
<td>2SLS K CS</td>
<td>[0.059,0.144]</td>
<td></td>
<td>[0.0456,0.1239]</td>
<td></td>
<td>[-0.5,0.5]</td>
<td></td>
<td></td>
<td>[-0.5,0.146]</td>
</tr>
<tr>
<td>SC CS</td>
<td>[0.052,0.153]</td>
<td></td>
<td>[-0.0023,0.1851]</td>
<td></td>
<td>[-0.4855,0.5]</td>
<td></td>
<td></td>
<td>[-0.0175,0.2511]</td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base Controls</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Age, Age$^2$</td>
<td>No</td>
<td></td>
<td>No</td>
<td></td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>SOB</td>
<td>No</td>
<td></td>
<td>No</td>
<td></td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Instruments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QOB</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>QOB*YOB</td>
<td>No</td>
<td></td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>QOB*SOB</td>
<td>No</td>
<td></td>
<td>No</td>
<td></td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td># instruments</td>
<td>3</td>
<td></td>
<td>30</td>
<td></td>
<td>28</td>
<td></td>
<td>178</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>329,509</td>
<td></td>
<td>329,509</td>
<td></td>
<td>329,509</td>
<td></td>
<td></td>
<td>329,509</td>
</tr>
</tbody>
</table>

Table 5: Results for Angrist and Krueger (1991) data. Specifications as in Staiger and Stock (1997): $Y =$ log weekly wages, $X=$ years of schooling, instruments $Z$ and exogenous controls as indicated. Base controls are Race, MSA, Married, Region, and Year of Birth. QOB, YOB, and SOB are quarter, year, and state of birth dummies, respectively. CUGMM, NW is CUGMM together with Newey and Windermijer (2009) standard errors, while HLIM and HFUL are estimators and standard errors proposed in Hausman et al. (2012). The maximal distortion cutoffs $\hat{\gamma}$ are based on comparing nominal 95% Wald confidence sets to the appropriate robust confidence sets $CS_{R,P}$ as described in Section 3.5. Efficiently weighted K statistic confidence sets with $\hat{\Omega} = \hat{\Sigma}_g^{-1}$ are used to calculate $\hat{\gamma}$ for CUGMM with usual and NW standard errors, while 2SLS-weighted K statistic with $\hat{\Omega} = \left( \frac{1}{T}Z'\hat{Z} \right)^{-1}$ is used for other estimators. For non-convex confidence sets we report the convex hull.
Questions for Future Consideration

- **CS** are inverse tests. Good **CS** are constructed from good (powerful) tests
  - Can we think of \( CS_2 \) as a good test statistic for \( \theta \)?
- Stock-Yogo motivated **CS** selection by testing the hypothesis of weak instruments
  - Is the test “\( CS_p \subseteq CS_{NR} \)” a good “test” of this hypothesis?
- The paper examines **CS** coverage, e.g. test size
  - What about power? Does \( CS_2 \) have good power?
- The theory concerns \( CS_R(\gamma) \) for fixed \( \gamma \). Does \( CS_R(\hat{\gamma}) \) have similar properties?
- The diagnostic \( \hat{\gamma} \) is very clever. What are its properties? Is it an estimate, or more like a p-value?
- When we have multiple coefficients, does it make sense to have \( \hat{\gamma} \) vary across coefficients? (Would it make sense to pick the robust **CS** for one coefficient and the non-robust **CS** for another?)
Uniformity Versus Bounded Distortion

- When test statistics are (asymptotically) non-pivotal, there is not a unique method to conduct tests or construct confidence sets.
- One strong (Berkeley) tradition is to focus on uniform tests:
  - Uniform tests are currently quite fashionable.
  - They can have the pitfall of being quite conservative.
- This paper focuses on tests with bounded size distortion:
  - Similar to Stock and Yogo (2005).
  - The advantage is that the tests have good size (and power) in the leading case of strong identification.
  - There is bounded distortion under weak identification.
- This seems to be a useful and practical compromise.
- We should not be overly concerned with uniformity (it does not correspond to optimal decision-making).