Outlier Robust Inference in the (Weak) IV Model

Jens Klooster, Mikhail Zhelonkin

Erasmus University Rotterdam

August 30, 2023

Motivation

- Outliers are widespread in empirical IV research (Young 2022).
- Frequently only one or two outlying observations.
- Several ways how outliers can harm inference.
- Inference in IV is typically done in two steps.
    1. First determine instrument's strength by means of F-test.
    2. Based on first stage result: use 2SLS estimator or weak instrument robust test.
- One outlier in any stage can break down the whole procedure.
- In particular, an outlier can cause an instrument to be "seemingly" strong.
- How can we solve this problem?

Motivation

- Robust estimation: Cohen Freue *et al.* (2013), Sølvsten (2020), Jiao (2022).
- What can we do when the instrument is weak?
- We show how to construct outlier robust AR, K and CLR tests.
- These tests are robust to outliers and weak instruments.
- Benefits of weak instrument robust tests:
    1. CLR test is known to have good power properties in the (homoskedastic) linear IV model irrespective of the strength of the instrument (Andrews, Stock, Moreira, 2006).
    2. Not necessary to rely on a pre-test.

Outline

1. The IV model
2. Classical tests: AR, K and CLR
3. Robustness properties
4. The robust AR, K and CLR tests
5. Simulation study
6. Empirical example

The IV model

- We assume data is generated according to the following model F_θ:

$$\text{Structural equation: } y = x\beta + u,$$
$$\text{First-stage equation: } x = z^\top \pi + v.$$

- We are interested in testing the hypothesis $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$.
- We assume we observe data $d_i = (y_i, x_i, z_i^\top)$ from Huber (1964) gross-error model:

$$F_t = (1-t)F_\theta + tG,$$

where G is assumed to be completely unknown.

The IV model

- We assume data is generated according to the following model F_θ:

$$\text{Reduced form equation: } y = z^\top \delta + \epsilon,$$
$$\text{First-stage equation: } x = z^\top \pi + v,$$

  where $\delta = \pi\beta$.
- We are interested in testing the hypothesis $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$.
- We assume we observe data $d_i = (y_i, x_i, z_i^\top)$ from Huber (1964) gross-error model:

$$F_t = (1 - t)F_\theta + tG,$$

  where G is assumed to be completely unknown.

Classical tests: AR, K and CLR

- Under the null hypothesis $\beta = \beta_0$,

$$\delta - \pi\beta_0 = \pi\beta - \pi\beta_0 = 0.$$

- Following Andrews, Stock and Sun (2019), we construct the AR, K and CLR statistics based on two statistics:

$$\begin{aligned} g &= \hat{\delta} - \hat{\pi}\beta_0 \\ D &= \hat{\pi} - \left(\Sigma_{\pi\delta} - \Sigma_{\pi\pi}\beta_0\right)\Omega^{-1}g, \end{aligned}$$

where $\hat{\delta}$ and $\hat{\pi}$ are LS estimators of $\delta = \pi\beta$ and π.

- g and D are asymptotically normal and uncorrelated.
- We can then introduce the AR, K and Wald statistic:

$$AR = g^\top \Omega^{-1} g, \quad W = D^\top \Psi^{-1} D, \quad K = g^\top D (D^\top \Omega D)^{-1} D^\top g,$$
$$CLR = \frac{1}{2}\left(AR - W + \sqrt{(AR - W)^2 + 4W \cdot K}\right).$$

Robustness properties

- Let T denote a statistical functional, then the influence function is defined as

$$\mathrm{IF}(d; T, F) = \lim_{t \downarrow 0} \frac{T\left((1-t)F + t\Delta_d\right) - T(F)}{t},$$

  where Δ_d denotes a point mass at the point d.
- Maximum bias over the neighborhood described by the pertubations $F_t = (1-t)F + tG$, where G is an arbitrary distribution, is approximately

$$\sup_G ||T(F_t) - T(F)|| \approx t \sup_d ||\mathrm{IF}(d; T, F)||.$$

- Condition for (local) robustness is a bounded influence function.

Robustness properties

Proposition

Under the null hypothesis $\beta = \beta_0$ the influence function of the CLR statistic, conditional on $D = \tilde{D}$, is

$$\mathrm{IF}(d; \sqrt{CLR}, F_\theta) = \begin{cases} \mathrm{IF}(d; \sqrt{AR}, F_\theta), & \text{if } \tilde{D} = 0, \\ \mathrm{IF}(d; \sqrt{K}, F_\theta), & \text{if } |\tilde{D}| > 0. \end{cases}$$

- The IFs of the AR and K statistic depend on the IF of g:

$$\mathrm{IF}(d; g, F_\theta) = \mathrm{IF}(d; \hat{\delta}, F_\theta) - \beta_0 \mathrm{IF}(d; \hat{\pi}, F_\theta).$$

- The IF of the LS estimators $\hat{\delta}$ and $\hat{\pi}$ are not bounded!
- One outlying observations can bias the estimators and this will affect the test statistics.

Robust AR, K and CLR statistics

- We construct the robust AR, K and Wald statistic based on two statistics:

$$
\begin{aligned}
g &= \hat{\delta}_M - \hat{\pi}_M \beta_0, \\
D &= \hat{\pi}_M - \left(\Sigma_{\pi\delta} - \Sigma_{\pi\pi}\beta_0\right)\Omega^{-1} g,
\end{aligned}
$$

  where $\hat{\delta}_M$ and $\hat{\pi}_M$ are M-estimators with a bounded IF.
- We can then introduce the robust AR, K and Wald statistics:

$$
\begin{aligned}
RAR &= g^\top \Omega^{-1} g, \\
RW &= D^\top \Psi^{-1} D, \\
RK &= g^\top D (D^\top \Omega D)^{-1} D^\top g.
\end{aligned}
$$

- We can write the robust CLR statistic as follows:

$$
RCLR = \frac{1}{2}\left(RAR - RW + \sqrt{(RAR - RW)^2 + 4RW \cdot RK}\right).
$$

Robust CLR test

Proposition

Under the null hypothesis $\beta = \beta_0$ the influence function of the CLR statistic, conditional on $D = \tilde{D}$, is

$$\mathrm{IF}(d; \sqrt{RCLR}, F_\theta) = \begin{cases} \mathrm{IF}(d; \sqrt{RAR}, F_\theta), & \text{if } \tilde{D} = 0, \\ \mathrm{IF}(d; \sqrt{RK}, F_\theta), & \text{if } |\tilde{D}| > 0, \end{cases}$$

and is bounded.

Proposition

Under the null hypothesis $\beta = \beta_0$ it holds that conditional on $D = \tilde{D}$

$$nRCLR \xrightarrow{d} \frac{1}{2}\left(\chi^2_{k-1} + \chi^2_1 - \tilde{W} + \sqrt{(\chi^2_{k-1} + \chi^2_1 + \tilde{W})^2 - 4\tilde{W}\chi^2_{k-1}}\right),$$

where $\tilde{W} = \tilde{D}^\top \Psi^{-1} \tilde{D}$, χ^2_{k-1} and χ^2_1 are independent chi-squared distributed random variables with $k-1$ and 1 degrees of freedom.

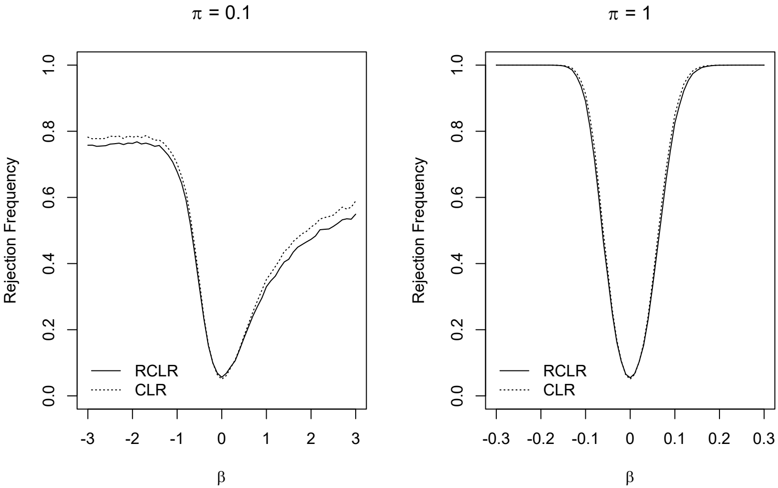
Simulation Study

- We sample data from the model:

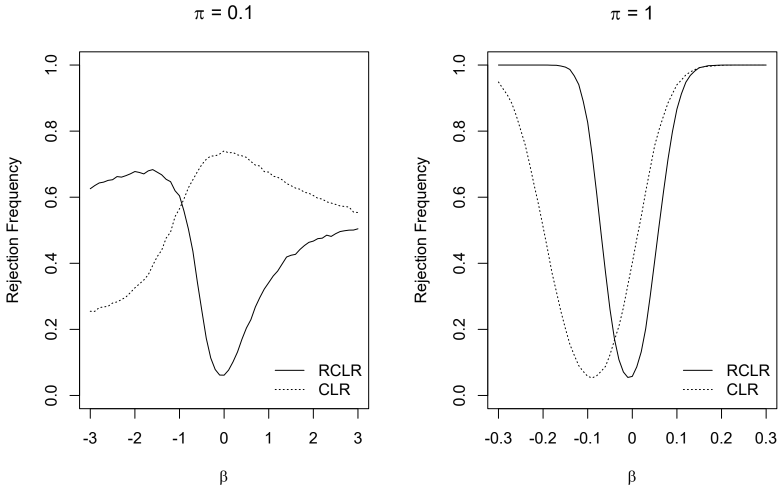
$$\text{Second stage: } y = x\beta + u,$$
$$\text{First stage: } x = z^\top \pi + v.$$

- We set $n = 200$, $k = 5$ and $\pi \in \{0.1, 1\}$.
- Each instrument is sampled from independent standard normal.
- We sample (u, v) from a bivariate normal with variances 1 and covariances 0.5.
- Test $H_0 \colon \beta = 0$ at a 5% significance level.
- Consider three different settings:
  1. Setting without outliers.
  2. Setting where we set $y_1 = 20$ and $z_{11} = 5$.
  3. Setting where 20% of the errors are generated by a $t(3)$-distribution.

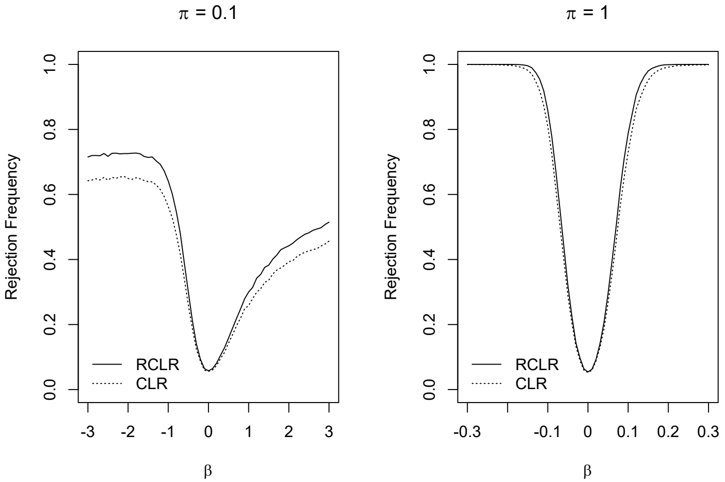
Simulation Study: no outlier



Simulation Study: large outlier



Simulation Study: “distributional” contamination



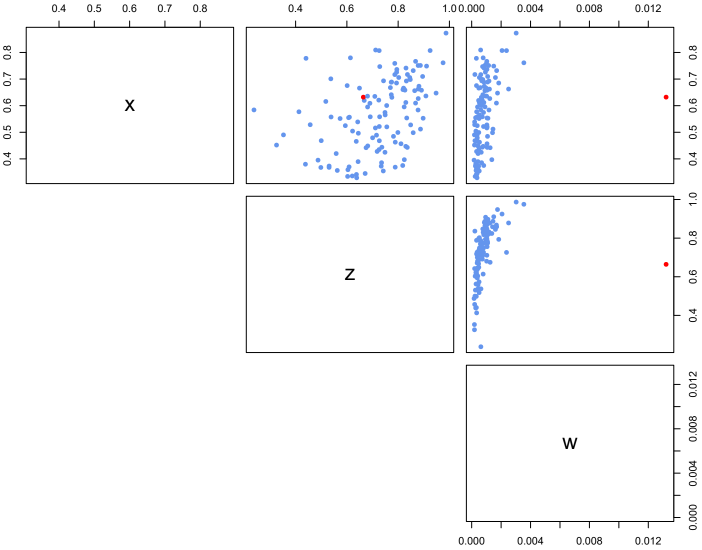
Empirical Example: Ananat (2011)

- Revisit Ananat (2011): "The wrong side(s) of the tracks: The causal effects of racial segregation on urban poverty and inequality"
- Following IV model is used:

$$\text{Structural equation: } y = x\beta + w\gamma_1 + u,$$
$$\text{First-stage equation: } x = z\pi + w\gamma_2 + v,$$

  where
  - y: different poverty and inequality measures
  - x: segregation
  - z: railroad division index (instrument)
  - w: length of the railroad track (control variable)
- Scatter plot of the data shows a large outlier in the control variable.

Empirical Example: Ananat (2011)



Empirical Example: Ananat (2011)

Specification	I Gini index whites	II Gini index blacks	III Poverty rate whites	IV Poverty rate blacks
95% AR confidence set	$(-0.64, -0.18)$	$(0.22, 2.15)$	$(-0.38, -0.09)$	$(0.00, 0.48)$
95% RAR confidence set	$(-\infty, -0.12)$ $\cup(1.62, \infty)$	$(-\infty, -3.79)$ $\cup(0.19, \infty)$	$(-\infty, -0.08)$ $\cup(0.90, \infty)$	$(-\infty, \infty)$
First-stage F	19.32	19.32	19.32	19.32
No. of observations	121	121	121	121

- Large difference between AR and RAR confidence sets indicate the AR confidence set might not be reliable.

Empirical Example: Angrist and Krueger (1991)

- Revisit Angrist and Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?"
- Replicate the Staiger and Stock (1997) specifications.
- Following IV model is used:

$$\text{Structural equation: } y = x\beta + w^\top \gamma_1 + u,$$
$$\text{First-stage equation: } x = z^\top \pi + w^\top \gamma_2 + v,$$

  where
  - y is the wage
  - x is the education level
  - z are quarter of birth (QOB) instruments
  - w are (base) control variables (age, age^2, SOB)
- Recently Sølvsten (2020) shows distribution of residuals fit better with $t(3)$ than normal. However, reasonably normal in the center.
- Reminiscent of the "distributional" contamination.

Empirical Example: Angrist and Krueger (1991)

Specification	I	II	III*	IV
95% CLR confidence set	$[0.042, 0.136]$	$[0.026, 0.116]$	$[-0.069, 0.274]$	$[-0.070, 0.261]$
95% RCLR confidence set	$[0.047, 0.122]$	$[0.032, 0.100]$	$[-0.044, 0.185]$	$[-0.053, 0.172]$
First-stage F	30.53	4.74	2.43	1.87
controls (w)				
Base controls	Yes	Yes	Yes	Yes
SOB	No	No	Yes	Yes
Age, Age^2	No	No	No	Yes
Instruments (z)				
QOB	Yes	Yes	Yes	Yes
QOB*YOB	No	Yes	Yes	Yes
QOB*SOB	No	No	Yes	Yes
No. of instruments	3	30	180	178
Observations	329,509	329,509	329,509	329,509

Summary

- Outliers are widespread in empirical IV research.
- Showed how to robustify the AR, K and CLR tests.
- The robust tests are robust against outliers (and weak instruments).
- Can be used as a robustness check or stand-alone method!

The end

Thank you for your time!