ivcheck

Tests of IV validity in R

Charles Coverdale

London, United Kingdom

April 2026

The question

Is this instrumental variable doing the job?

Empirical IV papers routinely estimate a LATE and move on. The identifying assumptions are testable, yet the tests are almost never reported.

Why it matters

- **Applied econometricians**: IV designs are used by every top-5 journal, yet fewer than 5 percent of published papers report a falsification test
- **Referees**: limited tooling makes the test hard to ask for, even when the design cries out for it
- **Methodologists**: three landmark tests exist (2015, 2017, 2023) but live only as author-hosted Stata and Matlab scripts
- **Replication**: published falsification claims cannot be reproduced without porting the tests by hand

What is already out there

- **`fixest`, `ivreg`**: estimate IV models, but do not test validity[1]
- **`ivDiag`** (Lal, Lockhart, Xu, Zu 2024): first-stage F, weak-IV, LATE bounds, partial robustness; does not implement the LATE-validity tests
- **Stata `testjfe`** (Frandsen, BYU 2020): FLL test only; no R equivalent
- **Author replication archives**: per-paper Matlab, Gauss, Stata; not packaged

The gap: **no single R package runs the published falsification tests on a fitted IV model.**

[1] Bergé (2018), *fixest*; Kleiber & Zeileis (2020), *ivreg*.

What ivcheck offers

1. **Coverage**: three published tests plus a one-shot wrapper that inspects a fitted IV model and runs every applicable test
2. **Dispatch**: S3 methods on `fixest::feols` and `ivreg::ivreg` objects, so the tests drop in after an existing IV estimation
3. **Fidelity**: variance-weighted Kitagawa, full CLR with adaptive moment selection for Mourifi'e-Wan, chi-squared asymptotic plus bootstrap for FLL

Source on GitHub, CRAN submission in progress[2].

[2] Coverdale (2026), *ivcheck: Tests of IV Validity in R*, available at github.com/charlescoverdale/ivcheck.

Three testable implications

Kitagawa (2015) for discrete Z, binary D:

$$\sup_{y \le y'} \left[P(Y \in [y, y'], D = d \mid Z = z_L) - P(Y \in [y, y'], D = d \mid Z = z_H)\right]^+ \tag{1}$$

Mourifi'e-Wan (2017) same null, conditional on covariates:

$$E[m(W;\ \theta) \mid X] \le 0 \text{ for all } X \tag{2}$$

Frandsen-Lefgren-Leslie (2023) judge-IV designs, K judges:

$$T_n = \sum_j n_j(\hat{\mu}_j - \hat{\alpha} - \hat{\beta}\hat{p}_j)^2/\hat{\sigma}^2 \sim \chi^2_{K-2} \tag{3}$$

Package architecture

Core exports:

- **`iv_kitagawa()`**: binary D, discrete Z
- **`iv_mw()`**: same, plus covariate path
- **`iv_testjfe()`**: judge-IV design
- **`iv_check()`**: wrapper, runs all applicable
- **`iv_power()`**: Monte Carlo power curves

Dispatch: `fixest`, `ivreg`. Deps: `cli`, `stats`, `parallel`.

Uniform output

Every test returns an `iv_test` object:

`statistic` | `p_value` | `binding`

Print, plot, summary methods; composes with `modelsummary`.

Minimal working example

```r
library(ivcheck)
library(fixest)

data(card1995)
m <- feols(
  lwage ~ age + black + south | college ~ near_college,
  data = card1995
)

iv_check(m, n_boot = 500)
```

One line after the IV estimation. `iv_check()` inspects the model, detects a binary treatment and a binary instrument, runs Kitagawa and Mourifi'e-Wan, and returns a tidy verdict table.

Case study: Card (1995) proximity to college

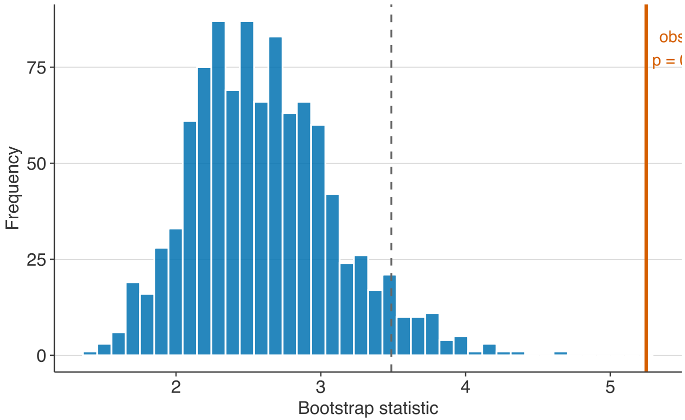
Data. 3{,}003 observations from the National Longitudinal Survey of Young Men (Card 1995), distributed on CRAN via the `wooldridge` package.

Question. *Does the standard college / near-college IV satisfy the LATE validity conditions on the discretised treatment?*

Why this case. Canonical applied IV example. Tested by Kitagawa (2015) and replicated by Mourifi'e-Wan (2017), both rejecting on the discretised form[3]. Our replication should reproduce the rejection.

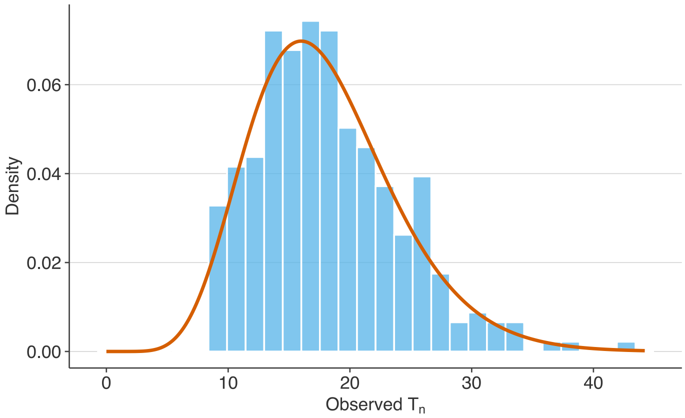
[3] Kitagawa (2015), Table 2; Mourifié-Wan (2017), Table 1.

Card rejects: Kitagawa statistic 5.25, p < 0.01



*Bootstrap null distribution of the Kitagawa statistic on Card (1995), N = 3,003. Observed statistic (red line) sits far in the upper tail: p-value below 0.01 at 500 bootstrap draws. Binding interval **[6.25, 7.78]** on log wage, d = 0 direction.*

FLL asymptotic distribution matches Monte Carlo



Empirical null distribution of `iv_testjfe` at K = 20 judges and N = 3,000 over 200 replications. Histogram is the empirical T_n density; red curve is the asymptotic χ^2_{18}. Test is mildly conservative (empirical size 1.5 percent at

What ivcheck does not yet do

- **Continuous instruments**: discretise into quantile bins for now; Andrews-Shi continuous-Z extension in v0.2.0
- **Fuzzy regression discontinuity**: has its own test (Arai et al. 2022); `iv_frd()` deferred to v0.2.0
- **Fixed-effects IV models**: aborts with a clear message; pre-demean within FE cells and call the default method
- **Stata cross-validation**: numerical agreement with Frandsen's `testjfe` and Chernozhukov et al. `clrtest` pending

v0.2.0: `iv_frd()`, continuous-Z path, multivariate-x in `iv_mw`, tensor-product basis, FLL bounded-slope moment inequalities.

Contact, code, paper

Charles Coverdale

`charles.f.coverdale@gmail.com`
`github.com/charlescoverdale/ivcheck`
`charlescoverdale.github.io/publications`

Acknowledgements. Thanks to Toru Kitagawa, Ismael Mourifié, Yuanyuan Wan, Brigham Frandsen, Lars Lefgren, and Emily Leslie for the methodology and publicly-available replication archives. Card (1995) data via the CRAN `wooldridge` package.

Scan for paper PDF