Data Availability and Provenance Statements

Data on the 1980 census extract from Angrist and Krueger (1991) were downloaded from Angrist Data Archive (https://economics.mit.edu/faculty/angrist/data1/data/angkru1991). We use replication code for Table V in Angrist and Krueger (1991) to clean the raw data. A copy of the cleaned data is provided as part of this archive. The data are in the public domain.

Datafile: `ak91.dta, AK_3039_analysis_dataset.mat` (Matlab version of `ak91.dta`)

Computational requirements

Software Requirements

- Stata (code was last run with version 13)
- Matlab (code was run with Matlab Release 2017b)
  - The loop runs in parallel with Parallel Computing Toolbox (PCT), but can be executed without PCT by replacing "parfor" with "for".

Memory and Runtime Requirements

The simulation code was last run on a **4-core Intel-based laptop with MacOS version 10.12.6**.

The code for estimating the variance for AK91 was last run on a **72-core Intel server with 11.5 TB of RAM**. Computation took 4 hours.

Description of programs

- Programs in `jackknife_AK_orig.m` implements the pre-test, Wald-JIV test and the jackknife AR test on AK91 with 180 and 1530 instruments. Solves for the jackknife AR confidence interval both via grid search and analytically.
- Programs in `qobsims_reverse_engineer.do` generates the dataset calibrated to AK91 with additional OVB following Angrist and Frandsen (2019).
- Programs in `jackknife_AK_AR_wP_compare.m` generates simulated data based on calibrated AK91 datasets and calculates bias, sizes, and power curves.
- The program `jackknife_AR_power_compare.m` includes a simple simulation design to compare the power implication of cross-fit vs. naive variance estimation.
- The program `lAR_test.m` is a Matlab function for jackknife AR test, can specify the variance estimate.

- The program `uncond_Wald_size.m` simulates critical value table similar to that of Stock and Yogo (2005) for various combinations of nominal sizes of JIVE-Wald and jackknife AR tests.

Instructions to Replicators

- To replicate analysis of the original AK91, download `AK_3039_analysis_dataset.mat` (Matlab version of `ak91.dta`) and run `jackknife_AK_orig.m`. With this sample size, we cannot vectorize the estimation for the quadratic-form covariance matrix because it involves the projection matrix. We cannot hold the NxN projection matrix in memory due to memory constraint. Instead, we increment the estimation and runtime is about four hours on a server. For faster replication we save the covariance estimates in 'Sigma1_180_YOB_SOB.mat' (for 180 instruments, cross-fit estimator), 'Sigma1_1530_YOB_SOB.mat' (for 1530 instruments, cross-fit estimator), in 'Sigma0_180_YOB_SOB.mat' (for 180 instruments, naive estimator), 'Sigma0_1530_YOB_SOB.mat' (for 1530 instruments, naive estimator).
  - To replicate Table 1 and inline results in Section 4.2, vary the number of instruments (K) and the type of variance covariance estimator ('crossfit' or 'naïve') in line 108.
- To replicate simulation of the calibrated AK91, download `ak91.dta` and run `qobsims_reverse_engineer.do` and save the output dataset in the same directory as `jackknife_AK_AR_wP_compare.m`. Then run `jackknife_AK_AR_wP_compare.m` The runtime is about five minutes on a laptop.
  - To replicate Table 3-Table6 and Figure 2 in the main text, vary sh in line 32 to choose different sample size, and execute the script.
- To replicate the comparison of variance estimators, run `jackknife_AR_power_compare.m` The runtime is about two minutes on a laptop.
  - To replicate the Figure 1 in the main text and Figure 1-3 in the Supplementary Appendix, vary pi and r for different specifications as specified in line 43-73, and execute the script.
  - This script calls the function `lAR_test` in the same directory.
- To calculate cut-offs and critical values for the two-step procedure, run `uncond_Wald_size.m` The runtime is about two minutes on a laptop.
  - For desired Wald test sizes other than 1%, 2%, 5%, vary nominal_size_grid in line 11 to choose desired sizes.

List of tables and programs

Figure/ Table #	Program	Line Number in the Program for Output	Output file	Note
Table 1	jackknife_ AK_orig.m	108; set vce = 'crossfit'	tables/AK91crossfit.tex	Vary the block of #(instruments) in line 108; Use the block of cross-fit variance estimator, or load Sigma1_180_YOB_SOB.mat or Sigma1_1530_YOB_SOB.mat
Inline results in Section 4.2 on naïve variance estimator for AK91	jackknife_ AK_orig.m	108; set vce = 'naïve'	tables/AK91naive.tex	Vary the block of #(instruments) in line 108 ; Use the block of implied error variance estimator, or load Sigma0_180_YOB_SOB.mat or Sigma0_1530_YOB_SOB.mat
Table 2 and the cutoff of 4.14 mentioned throughout the article	uncond_Wald_size.m		tables/table2.tex	
Table 3	jackknife_ AK_AR_wP_compar	451	tables/table3.tex	Vary sh in line 32 to choose different sample

	e.m			size and append to output
Table 4	jackknife_AK_AR_wP_compare.m	461	tables/table4.tex	Vary sh in line 32 to choose different sample size and append to output
Table 5	jackknife_AK_AR_wP_compare.m	471	tables/table5.tex	Vary sh in line 32 to choose different sample size and append to output
Table 6	jackknife_AK_AR_wP_compare.m	480	tables/table6a.tex tables/table6b.tex	Vary sh in line 32 to choose different sample size and append to output
Figure 2	jackknife_AK_AR_wP_compare.m	497	figures/jackknife_power_AK91_005.eps figures/jackknife_power_AK91_0025.eps	Vary sh in line 32 to choose different sample size and append to output
Figure 1(a) and SA Figure 1(a)	jackknife_AR_power_compare.m	45-47	figures/jackknifepowerdense1r02.eps	Uncomment and execute the script
Figure 1(a) and SA Figure 1(a)	jackknife_AR_power_compare.m	50-52	figures/jackknifepowersparse1r02.eps	Uncomment and execute the script
SA Figure 2(a)	jackknife_AR_power_compare.m	55-57	figures/jackknifepowersparse1r02pi3.eps	Uncomment and execute the script
SA Figure 2(b)	jackknife_AR_power_compare.m	60-62	figures/jackknifepowersparse1r02pi36.eps	Uncomment and execute the script

SA Figure 3(a)	jackknife_AR_power_compare.m	65-67	figures/jackknifepowersparse1r05.eps	Uncomment and execute the script
SA Figure 3(b)	jackknife_AR_power_compare.m	70-72	figures/jackknifepowersparse1r09.eps	Uncomment and execute the script

References

Angrist, J.D., and Frandsen, B. (2019). Machine Labor (National Bureau of Economic Research).

Angrist, J.D., and Krueger, A.B. (1991). Does Compulsory School Attendance Affect Schooling and Earnings? The Quarterly Journal of Economics *106*, 979–1014.

Acknowledgements

We are grateful to Brigham Frandsen for sharing the code underlying our simulation design.