

# Package ‘ManyIV’

February 22, 2021

**Title** Inference in Instrumental Variables Models with Many Instruments

**Version** 0.0.2

**Description** Calculate estimators and standard errors in a linear instrumental variables model with possibly many instruments and many exogenous regressors under homoskedasticity of the reduced form. Also computes tests of overidentifying restrictions.

**Depends** R ( $\geq 4.0.0$ )

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**Imports** Matrix,  
stats,  
Formula,  
numDeriv

**RoxygenNote** 7.1.1

**Suggests** spelling,  
testthat,  
knitr,  
rmarkdown

**VignetteBuilder** knitr

**LazyLoad** yes

**Language** en-US

**BugReports** <https://github.com/kolesarm/ManyIV/issues>

**URL** <https://github.com/kolesarm/ManyIV>

## R topics documented:

ak80 . . . . .	2
IVoverid . . . . .	3
IVreg . . . . .	3

<b>Index</b>	<b>6</b>
--------------	----------

---

ak80*Angrist and Krueger (1991) Census data*

---

**Description**

Sample of males born in the US in 1930-39 from 5 percent sample of the 1980 US Census

**Usage**

ak80

**Format**

A data frame with 329,509 observations on 10 variables:

**age** Age, measured at quarterly precision

**education** Years of education

**lwage** Log of weekly wage

**married** Indicator for being married

**qob** Quarter of birth

**sob** State of birth

**black** Indicator for being black

**smsa** SMSA indicator

**yob** Year of birth

**division** Census division

**Source**

Josh Angrist's website, <http://economics.mit.edu/faculty/angrist/data1/data/angkru1991>

**References**

Angrist, Joshua D., and Alan B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106 (4): 979–1014. doi: [10.2307/2937954](https://doi.org/10.2307/2937954).

IVoverid

*Test of overidentifying restrictions***Description**

Report the Sargan and modified Cragg-Donald test statistics and  $p$ -values for testing of overidentifying restrictions, assuming homoskedasticity of the reduced form. The Sargan test is valid under few instruments. The Modified Cragg-Donald test (Modified-CD) corresponds to a test due to Cragg and Donald (1993), with a modified critical value. The modification was suggested in Kolesár (2018) to make it robust to many instruments and many exogenous regressors.

**Usage**

```
IVoverid(r)
```

**Arguments**

`r` An object of class `RDResults`

**References**

Kolesár, Michal. *Minimum Distance Approach to Inference with Many Instruments.* *Journal of Econometrics* 204 (1): 86–100. doi: [10.1016/j.jeconom.2018.01.004](https://doi.org/10.1016/j.jeconom.2018.01.004).

Cragg, John G., and Stephen G. Donald. 1993. "Testing Identifiability and Specification in Instrumental Variable Models." *Econometric Theory* 9 (2): 222–40. doi: [10.1017/S0266466600007519](https://doi.org/10.1017/S0266466600007519).

Sargan, John Denis. 1958. "The Estimation of Economic Relationships Using Instrumental Variables." *Econometrica* 26 (3): 393–415. doi: [10.2307/1907619](https://doi.org/10.2307/1907619).

**Examples**

```
r1 <- IVreg(lwage~education+black+married | as.factor(qob), data=ak80,
            inference="standard")
IVoverid(r1)
```

IVreg

*Instrumental Variables Regression***Description**

Fit instrumental variables regression by a number of methods, and compute associated standard errors, as specified by inference

**Usage**

```
IVreg(formula, data, subset, na.action, inference = "standard", approx = TRUE)
```

## Arguments

<code>formula</code>	Specification of the regression relationship and the instruments of the form $y \sim x + w_1 + w_2 \mid z_1 + z_2 + z_3$ , where $y$ is the outcome variable, $x$ is a scalar endogenous variable, $w_1, w_2$ are exogenous regressors, and $z_1, z_2$ , and $z_3$ are excluded instruments.
<code>data</code>	An optional data frame, list or environment (or object coercible by <code>as.data.frame</code> to a data frame) containing the variables in the model. If not found in <code>data</code> , the variables are taken from <code>environment(formula)</code> , typically the environment from which the function is called.
<code>subset</code>	An optional vector specifying a subset of observations to be used in the fitting process.
<code>na.action</code>	A function indicating what should happen when the data contain NAs. The default is set by the <code>na.action</code> setting of options (usually <code>na.omit</code> ).
<code>inference</code>	<p>Vector specifying inference method(s). The elements of the vector can consist of the following methods:</p> <ul style="list-style-type: none"> <li>"standard" Report inference based on TSLS, LIML, and MBTSLS, along with homoskedastic and heteroskedasticity-robust standard errors, standard errors that are valid under heteroskedasticity and treatment effect heterogeneity. All three standard errors are valid under standard asymptotics only.</li> <li>"re" Report standard errors for LIML based on Hessian of random effects likelihood</li> <li>"il" Report standard errors for LIML based on Hessian of invariant likelihood, evaluated numerically</li> <li>"lil" Report standard errors for LIML based on the information matrix of limited information likelihood</li> <li>"md" Compute the EMD, LIML, and MBTSLS estimators, and report standard errors for LIML, MBTSLS, and EMD based on the minimum distance objective function proposed in Kolesár (2018)</li> </ul> <p>See the vignette <code>vignette("ManyIV", package = "ManyIV")</code> for a detailed description of these methods.</p>
<code>approx</code>	if TRUE, then estimates of third and fourth moments used in inference based on the minimum distance objective function ( <code>inference="md"</code> ) are based on an approximation to speed up the calculations.

## Value

An object of class `"IVResults"`, which is a list with the following components:

**IVData** An object of class `"IVData"`, which is a list with at least the following components:

- Z** Matrix of instruments
- Y** Matrix with two columns collecting the endogenous variables
- W** Matrix of exogenous regressors
- n** Number of observations used, the number of rows of **Z**, **W**, or **Yp**
- l** Dimension of the exogenous regressors, the number of columns of **W**
- k** Dimension of the instruments, the number of columns of **Z**

**F** First-stage  $F$  statistic

**call** The matched call.

**estimate** A data frame containing the estimation results.

The print function can be used to print a summary of the results.

## References

*Kolesár, Michal. Minimum Distance Approach to Inference with Many Instruments.” Journal of Econometrics 204 (1): 86–100.*

## Examples

```
## Use quarter of birth as an instrument for education, controlling for
## marriage and black indicators
IVreg(lwage~education+black+married | as.factor(qob),
      data=ak80, inference=c("standard", "re", "il", "lil"))
```

# Index

\* **datasets**

ak80, [2](#)

ak80, [2](#)

IOverid, [3](#)

IVreg, [3](#)