

Package: ivDiag (via r-universe)

September 13, 2024

Title Estimation and Diagnostic Tools for Instrumental Variables Designs

Version 1.0.6

Date 2023-09-16

Maintainer Yiqing Xu <yiqingxu@stanford.edu>

Description Estimation and diagnostic tools for instrumental variables designs, which implements the guidelines proposed in Lal et al. (2023) <[arXiv:2303.11399](https://arxiv.org/abs/2303.11399)>, including bootstrapped confidence intervals, effective F-statistic, Anderson-Rubin test, valid-t ratio test, and local-to-zero tests.

URL <https://yiqingxu.org/packages/ivDiag/>

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

Depends R (>= 3.5.0)

Imports foreach, future, doParallel, lfe, fixest, ggplot2, ggfortify, wCorr, haven, glue, patchwork, testthat

NeedsCompilation no

Author Apoorva Lal [aut] (<<https://orcid.org/0000-0002-3697-614X>>), Yiqing Xu [aut, cre] (<<https://orcid.org/0000-0003-2041-6671>>)

Repository CRAN

Date/Publication 2023-09-17 06:00:02 UTC

Contents

ivDiag-package	2
AR_test	2
eff_F	4
gsz	5

gsz_south	5
ivDiag	6
ltz	8
plot_coef	9
plot_ltz	10
rueda	11
tF	11

Index	13
--------------	-----------

ivDiag-package	<i>IV Estimation and Diagnostics</i>
----------------	--------------------------------------

Description

Conducts various estimation and diagnostic procedure for instrumental variable designs in one shot.

Details

Provides estimation and diagnostic tools for instrumental variables designs, which implements the guidelines proposed in Lal et al. (2023) <arXiv:2303.11399>, including bootstrapped confidence intervals, effective F-statistic, Anderson-Rubin test, valid-t ratio test, and local-to-zero tests.

See [ivDiag](#) for details.

Author(s)

Apoorva Lal; Yiqing Xu

References

Lal, Apoorva, Mackenzie William Lockhart, Yiqing Xu, and Ziwen Zu. 2023. "How Much Should We Trust Instrumental Variable Estimates in Political Science? Practical Advice Based on 67 Replicated Studies." Available at: https://yiqingxu.org/papers/english/2021_iv/LLXZ.pdf

AR_test	<i>Anderson Rubin Test</i>
---------	----------------------------

Description

Performs the Anderson Rubin test, which is robust to weak instruments.

Usage

```
AR_test(data, Y, D, Z, controls, FE = NULL, cl = NULL, weights = NULL,
  prec = 4, CI = TRUE, alpha = 0.05, parallel = NULL, cores = NULL)
```

Arguments

<code>data</code>	name of a dataframe.
<code>Y</code>	a string indicating the outcome variable.
<code>D</code>	a string indicating the treatment variable.
<code>Z</code>	a vector of strings indicating the instrumental variables.
<code>controls</code>	a vector of strings indicating the control variables.
<code>FE</code>	a vector of strings indicating the fixed effects variables.
<code>cl</code>	a string indicating the clustering variable.
<code>weights</code>	a string indicating the variable that stores weights.
<code>CI</code>	a logical flag controlling whether to calculate the confidence interval using the inversion method.
<code>prec</code>	precision of results (4 by default).
<code>alpha</code>	level of statistical significance; the default is 0.05.
<code>parallel</code>	a logical flag controlling parallel computing.
<code>cores</code>	setting the number of cores.

Value

<code>Fstat</code>	F statistic, degrees of freedoms, and p-value.
<code>ci.print</code>	Confidence interval via inversion (printed version).
<code>ci</code>	Confidence interval via inversion (numeric version).
<code>bounded</code>	If the confidence interval is bounded.

References

Chernozhukov, Victor, and Christian Hansen. 2008. "The Reduced Form: A Simple Approach to Inference with Weak Instruments." *Economics Letters* 100 (1): 68–71.

See Also

[ivDiag](#)

Examples

```
data(ivDiag)
AR.out <- AR_test(data = rueda, Y = "e_vote_buying", D = "lm_pob_mesa",
  Z = "lz_pob_mesa_f", controls = c("lpopulation", "lpotencial"),
  cl = "muni_code", CI = FALSE)
library(testthat)
test_that("Check AR results", {
  expect_equal(as.numeric(AR.out$Fstat[1]), 48.4768)
})
```

eff_F	<i>Effective F</i>
-------	--------------------

Description

Computes the effective F statistic.

Usage

```
eff_F(data, Y, D, Z, controls = NULL, FE = NULL, cl = NULL,  
       weights = NULL, prec = 4)
```

Arguments

data	name of a dataframe.
Y	a string indicating the outcome variable.
D	a string indicating the treatment variable.
Z	a vector of strings indicating the instrumental variables.
controls	a vector of strings indicating the control variables.
FE	a vector of strings indicating the fixed effects variables.
cl	a string indicating the clustering variable.
weights	a string indicating the variable that stores weights.
prec	precision of results (4 by default).

Value

the effective F statistic.

References

Olea, José Luis Montiel, and Carolin Pflueger. 2013. "A Robust Test for Weak Instruments." *Journal of Business & Economic Statistics* 31 (3): 358–69.

See Also

[ivDiag](#)

Examples

```
effF <- eff_F(data = rueda, Y = "e_vote_buying", D = "lm_pob_mesa",  
             Z = "lz_pob_mesa_f", controls = c("lpopulation", "lpotencial"),  
             cl = "muni_code")  
library(testthat)  
test_that("Check effective F", {  
  expect_equal(floor(as.numeric(effF)), 8598)  
})
```

gsz

Data from GSZ (2016)

Description

Data from Guiso, Sapienza, and Zingales (2016)

Format

A data frame with 5357 rows and 11 columns.

Details

The authors revisit Putnam, Leonardi, and Nanetti (1992)'s celebrated conjecture that Italian cities that achieved self-government in the Middle Ages have higher modern-day levels of social capital. More specifically, they study the effects of free city-state status on social capital as measured by the number of nonprofit organizations and organ donations per capita, and a measure of whether students cheat in mathematics. We focus on the first outcome, the number of nonprofit organizations.

References

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2016. "Long-Term Persistence." *Journal of the European Economic Association* 14 (6): 1401–36.

gsz_south

Data from GSZ (2016): Subsample

Description

Data from Guiso, Sapienza, and Zingales (2016); southern Italian cities

Format

A data frame with 2175 rows and 11 columns.

Details

The authors revisit Putnam, Leonardi, and Nanetti (1992)'s celebrated conjecture that Italian cities that achieved self-government in the Middle Ages have higher modern-day levels of social capital. More specifically, they study the effects of free city-state status on social capital as measured by the number of nonprofit organizations and organ donations per capita, and a measure of whether students cheat in mathematics. We focus on the first outcome, the number of nonprofit organizations.

This dataset is a subsample of southern Italian cities, which is used as a zero-first-stage sample.

References

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2016. "Long-Term Persistence." *Journal of the European Economic Association* 14 (6): 1401–36.

 ivDiag

Omnibus Function for IV Estimation and Diagnostics

Description

Conducts various estimation and diagnostic procedure for instrumental variable designs in one shot.

Usage

```
ivDiag(data, Y, D, Z, controls = NULL, FE = NULL, cl = NULL, weights = NULL,
        bootstrap = TRUE, run.AR = TRUE,
        nboots = 1000, parallel = TRUE, cores = NULL,
        seed = 94305, prec = 4, debug = FALSE)
```

Arguments

data	name of a dataframe.
Y	a string indicating the outcome variable.
D	a string indicating the treatment variable.
Z	a vector of strings indicating the instrumental variables.
controls	a vector of strings indicating the control variables.
FE	a vector of strings indicating the fixed effects variables.
cl	a string indicating the clustering variable.
weights	a string indicating the variable that stores weights.
bootstrap	whether to turn on bootstrap (TRUE by default).
run.AR	whether to run AR test (TRUE by default).
nboots	a numeric value indicating the number of bootstrap runs.
parallel	a logical flag controlling parallel computing.
cores	setting the number of cores.
prec	precision of CI in string (4 by default).
seed	setting seed.
debug	for debugging purposes.

Value

est_ols	results from an OLS regression.
est_2sls	results from a 2SLS regression.
AR	results from an Anderson-Rubin test
F_stat	various F statistics.
rho	Pearson correlation coefficient between the treatment and predicted treatment from the first stage regression (all covariates are partialled out).
tF	results from the tF procedure based on Lee et al. (2022)
est_rf	results from the first stage regression.
est_fs	results from the reduced form regression.
p_iv	the number of instruments.
N	the number of observations.
N_cl	the number of clusters.
df	the degree of freedom left from the 2SLS regression
nvalues	the unique values the outcome Y, the treatment D, and each instrument in Z in the 2SLS regression.

Author(s)

Apoorva Lal; Yiqing Xu

References

Lal, Apoorva, Mackenzie William Lockhart, Yiqing Xu, and Ziwen Zu. 2023. "How Much Should We Trust Instrumental Variable Estimates in Political Science? Practical Advice Based on 67 Replicated Studies." Available at: https://yiqingxu.org/papers/english/2021_iv/LLXZ.pdf

Lee, David S, Justin McCrary, Marcelo J Moreira, and Jack Porter. 2022. "Valid t-Ratio Inference for IV." *American Economic Review* 112 (10): 3260–90.

See Also

[plot_coef](#) [eff_F](#) [AR_test](#) [tF](#)

Examples

```
data(ivDiag)
g <- ivDiag(data = rueda, Y = "e_vote_buying", D = "lm_pob_mesa",
  Z = "lz_pob_mesa_f", controls = c("lpopulation", "lpotencial"),
  cl = "muni_code", bootstrap = FALSE, run.AR = FALSE)
plot_coef(g)
library(testthat)
test_that("Check ivDiag output", {
  expect_equal(as.numeric(g$est_2sls[1,1]), -0.9835)
})
```

`ltz`*Local-to-Zero Test*

Description

Estimates Local-to-Zero IV coefficients and SEs for a single instrument.

Usage

```
ltz(data, Y, D, Z, controls, FE = NULL, cl = NULL, weights = NULL, prior, prec = 4)
```

Arguments

<code>data</code>	name of a dataframe.
<code>Y</code>	a string indicating the outcome variable.
<code>D</code>	a string indicating the treatment variable.
<code>Z</code>	a vector of strings indicating the instrumental variables.
<code>controls</code>	a vector of strings indicating the control variables.
<code>FE</code>	a vector of strings indicating the fixed effects variables.
<code>cl</code>	a string indicating the clustering variable.
<code>weights</code>	a string indicating the variable that stores weights.
<code>prior</code>	prior mean and standard deviation of the direct effect of instrument on outcome.
<code>prec</code>	precision of results (4 by default).

Value

<code>iv</code>	results from a 2SLS regression.
<code>ltz</code>	results after local-to-zero adjustment.
<code>prior</code>	prior mean and standard deviation

References

Conley, Timothy G, Christian B Hansen, and Peter E Rossi. 2012. "Plausibly Exogenous." *Review of Economics and Statistics* 94 (1): 260–72.

See Also

[plot_coef](#)

Examples

```

data(ivDiag)
controls <- c('altitudine', 'escursione', 'costal', 'nearsea', 'population',
             'pop2', 'gini_land', 'gini_income')
ltz_out <- ltz(data = gsz, Y = "totassoc_p", D = "libero_comune_allnord",
              Z = "bishopcity", controls = controls, weights = "population",
              prior = c(0.178, 0.137))
plot_ltz(ltz_out)

library(testthat)
test_that("Check local-to-zero adjustment", {
  expect_equal(as.numeric(ltz_out$ltz[1]), 3.6088)
})

```

plot_coef

*Plot OLS and IV Coefficients***Description**

Visualise point estimates and confidence intervals of OLS and IV estimates.

Usage

```

plot_coef(out,
          ols.methods = c("analy", "bootc", "boott"),
          iv.methods = c("analy", "bootc", "boott", "ar", "tf"),
          main = NULL, ylab = "Coefficient", grid = TRUE,
          stats = TRUE, ylim = NULL)

```

Arguments

out	output from ivDiag::ivDiag.
ols.methods	a vector specifying the inferential methods for OLS to be shown. The default is c("analy", "bootc", "boott").
iv.methods	a vector specifying inferential methods for 2SLS to be shown. The default is c("analy", "bootc", "boott", "ar", "tf").
main	a string specifying the title of the plot.
ylab	a string specifying ylab of the plot.
grid	a logical flag indicating whether to show the grids.
stats	a logical flag indicating whether to show the statistics, including the effective F, the number of observations, and the number of clusters (if applicable).
ylim	a two-element vector specifying the range of the y-axis.

Value

A base R plot object.

See Also[ivDiag](#)

`plot_ltz`*Visualizing Local-to-Zero Adjustment*

Description

Visualise approximate sampling distributions for scalar IV coefficient with local-to-zero adjustment.

Usage

```
plot_ltz(out = NULL, iv_est = NULL, ltz_est = NULL, prior = NULL, xlim = NULL)
```

Arguments

<code>out</code>	output from <code>ivDiag::ltz</code> .
<code>iv_est</code>	a two-element vector of IV estimate and standard error.
<code>ltz_est</code>	a two-element vector of local-to-zero estimate and standard error.
<code>prior</code>	a two-element vector of prior mean and standard deviation.
<code>xlim</code>	a two-element vector specifying the range of the x-axis.

Value

A `ggplot2` object.

References

Conley, Timothy G, Christian B Hansen, and Peter E Rossi. 2012. "Plausibly Exogenous." *Review of Economics and Statistics* 94 (1): 260–72.

See Also[ltz](#)

rueda	<i>Data from Rueda (2017)</i>
-------	-------------------------------

Description

Data from Rueda (2017) AJPS.

Format

A data frame with 4352 rows and 6 columns.

Details

Rueda (2017) studies the persistence of vote buying in developing democracies despite the use of secret ballots and argues that brokers condition future payments on published electoral results to enforce these transactions and that this is effective only when the results of small voting groups are available. The study examines the relationship between polling station size and vote buying using three different measures of the incidence of vote buying, two at the municipality level and one at the individual level.

The size of the polling station, predicted by the rules limiting the number of voters per polling station, is used as an instrument of the actual polling place size. The institutional rule predicts sharp reductions in the size of the average polling station of a municipality every time the number of registered voters reaches a multiple of the maximum number of voters allowed to vote in a polling station. Such sharp reductions are used as a source of exogenous variation in polling place size to estimate the causal effect of this variable on vote buying.

References

Rueda, Miguel R. 2017. "Small Aggregates, Big Manipulation: Vote Buying Enforcement and Collective Monitoring." *American Journal of Political Science* 61 (1): 163–177.

tF	<i>Valid t-Ratio Procedure</i>
----	--------------------------------

Description

Performs the valid t-ratio procedure.

Usage

tF(coef, se, Fstat, prec = 4)

Arguments

<code>coef</code>	a 2SLS coefficient.
<code>se</code>	a standard error estimate for the estimated 2SLS coefficient.
<code>Fstat</code>	a first-stage partial F statistic.
<code>prec</code>	precision of results (4 by default).

Value

Results from a valid t-ratio test given the first-stage F statistic.

References

Lee, David S, Justin McCrary, Marcelo J Moreira, and Jack Porter. 2022. "Valid t-Ratio Inference for IV." *American Economic Review* 112 (10): 3260–90.

Examples

```
tf.out <- tF(coef = -0.9835, se = 0.1540, Fstat = 8598)
library(testthat)
test_that("Check tF cF", {
  expect_equal(as.numeric(tf.out[2]), 1.96)
})
```

Index

* datasets

gsz, [5](#)

gsz_south, [5](#)

rueda, [11](#)

AR_test, [2](#), [7](#)

eff_F, [4](#), [7](#)

gsz, [5](#)

gsz_south, [5](#)

ivDiag, [2-4](#), [6](#), [10](#)

ivDiag-package, [2](#)

ltz, [8](#), [10](#)

plot_coef, [7](#), [8](#), [9](#)

plot_ltz, [10](#)

rueda, [11](#)

tF, [7](#), [11](#)