Package ‘ivpack’

February 20, 2015

Type Package
Title Instrumental Variable Estimation.
Version 1.2
Date 2014-10-24
Author Yang Jiang and Dylan Small
Maintainer Dylan Small <dsmall@wharton.upenn.edu>
Description This package contains functions for carrying out instrumental variable estimation of causal effects and power analyses for instrumental variable studies.
Depends AER, sandwich, lmtest
License GPL-2
NeedsCompilation no
Repository CRAN
Date/Publication 2014-10-25 08:07:54

R topics documented:

ivpack-package .......................................................... 2
anderson.rubin.ci ....................................................... 3
ARsensitivity.ci ......................................................... 4
ARsensitivity.power ..................................................... 5
ARsensitivity.size ....................................................... 7
card.data ............................................................ 9
cluster.robust.se ..................................................... 10
clx ................................................................. 11
power.iv ............................................................ 12
robust.se .......................................................... 13
sigmav.func .......................................................... 14

Index 15
ivpack-package

Instrumental Variable Analyses.

Description

The package implements several types of instrumental variable analyses for making causal inferences.

Details

Package: ivpack
Type: Package
Version: 1.0
Date: 2013-12-28
License: GPL-2

The functions robust.se and cluster.robust.se compute robust to heteroskedasticity and robust to clustering standard errors from an instrumental variable model fit using the ivreg command (from the AER package). The function anderson.rubin.ci computes the Anderson-Rubin confidence interval for an instrumental variable model, which is a confidence interval that is valid for both weak and strong instruments. The function power.iv computes the power for a planned instrumental variables analysis.

Author(s)

Dylan Small <dsmall@wharton.upenn.edu>

References


See Also

ivreg

Examples

### This is the IV model in panel A, column (5) of Table 3 from Card, 1995, "Using Geographic Variation in College Proximity to Estimate the Return from Schooling" data(card.data)
`ivmodel=ivreg(lwage ~ educ + exper + expersq + black + south + smsa + reg661 + reg662 + reg663 + reg664 + reg665 + reg666 + reg667 + reg668 + smsa66, ~ nearc4 + exper + expersq + black + south + smsa + reg661 + reg662 + reg663 + reg664 + reg665 + reg666 + reg667 + reg668 + smsa66, x=TRUE, data=card.data)`
# Anderson-Rubin confidence interval for effect of treatment
anderson.rubin.ci(ivmodel)
# Robust to heteroskedasticity standard errors
robust.se(ivmodel)

### Power for a study with in which the null hypothesis causal effect is 0,
### the true causal effect is 1, the sample size is 250, the instrument is
### binary with probability .5 (so variance = .25), the standard deviation
### of potential outcome under control is 1, the effect of the instrument
### is to increase the probability of a binary treatment being 1 from .25 to
### .75. The function sigmav.func computes the SD of v for a binary insrument,
### binary treatment. The correlation between u and v is assumed to be .5. The
### significance level for the study will be alpha = .05

```r
sigmav.func(prob.d1.given.z1=.75,prob.d1.given.z0=.25,prob.z1=.5)
# The sigmav.func finds sigmav=.4330127
power.iv(n=250, lambda=1, gamma=.5, var.z=.25, sigmav=1, sigmav=.4330127, rho=.5, alpha = .05)
```

Description

Calculates the Anderson-Rubin confidence interval for the effect of a treatment (endogenous) variable using an instrumental variable.

Usage

```r
anderson.rubin.ci(ivmodel, conflevel = 0.95)
```

Arguments

- **ivmodel**: Instrumental variable (IV) model fit using ivreg. Make sure to use the option x=TRUE when fitting the ivreg model.
- **conflevel**: Confidence level for confidence interval.

Value


Author(s)

Dylan Small

References

See Also

`ivreg`

Examples

```r
### This is the IV model in panel A, column (5) of Table 3 from Card, 1995, "Using
### Geographic Variation in College Proximity to Estimate the Return from Schooling"
data(card.data)
ivmodel=ivreg(lwage ~ educ + exper + expersq + black + south + smsa + reg661 + reg662 +
reg663 + reg664 + reg665 + reg666 + reg667 + reg668 + smsa66, - nearc4 + exper +
expersq + black + south + smsa + reg661 + reg662 + reg663 + reg664 + reg665 + reg666 +
reg667 + reg668 + smsa66, x=TRUE, data=card.data)
anderson.rubin.ci(ivmodel)
```

Description

Calculates the confidence interval for the effect of a treatment (endogenous) variable using an instrumental variable, which is based on an extension of Anderson-Rubin test and allows IV be possibly invalid within a certain range.

Usage

```r
ARsensitivity.ci(ivmodel, Delta=NULL, conflevel=.95)
```

Arguments

- `ivmodel`: Instrumental variable (IV) model fit using `ivreg`. Make sure to use the option `x=TRUE` when fitting the `ivreg` model.
- `Delta`: The allowance of sensitivity parameter for possibly invalid IV. If `Delta=NULL`, the `ARsensitivity.ci` function will calculate the confidence interval for a standard Anderson-Rubin test with valid IV.
- `conflevel`: Confidence level for confidence interval.

Value

- `confidence.interval`: Confidence interval for effect of treatment. If it’s a 2*2 matrix, the confidence interval is consisted of two disjoint intervals, each row of the matrix is one interval.
- `printinfo`: Report the confidence interval in one printing sentence.
- `ci.type`: If `ci.type=1`, the confidence interval is finite. If `ci.type=2`, the confidence interval is infinite. If `ci.type=3`, the confidence interval is an empty set.
**Description**

Computes the power of sensitivity analysis, which is based on an extension of Anderson-Rubin test and allows IV be possibly invalid within a certain range.

**Usage**

\[
\text{ARsensitivity.power}(n, k, \lambda, \gamma, \text{var.z, sigma1, sigma2, rho, alpha=0.05, Delta=NULL, delta=NULL})
\]

**Arguments**

- **n**: Sample size.
- **k**: Number of exogenous variables, intercept is also counted. If there’s no observed variable, k=1.
- **lambda**: True causal effect minus null hypothesis causal effect.
**gamma**
Regression coefficient for effect of instrument on treatment.

**var.z**
Variance of instrument.

**sigma1**
Standard deviation of potential outcome under control (structural error for y).

**sigma2**
Standard deviation of error from regressing treatment on instrument.

**rho**
Correlation between u (potential outcome under control) and v (error from regressing on instrument).

**alpha**
Significance level of test.

**Delta**
Range of sensitivity allowance. Usually a numeric vector of length 2. If Delta=NULL, the ARsensitivity.power function will calculate the power for standard Anderson-Rubin test with valid IV.

**delta**
True value of sensitivity parameter when calculating the power. Usually take delta=0 for the favourable situation or delta=NULL for unknown delta.

**Details**
The structural equations model assumed is: \( D = \gamma_0 + \gamma z + v \), \( Y = \beta_0 + \beta_1 D + \delta Z + u \).
This model can also be obtained by assuming the potential outcomes model. See Jiang, Zhang and Small (2013) for details.

\( \lambda \) is equal to the true \( \beta_1 \) minus the null hypothesis \( \beta_1 \).

**Value**
Power of sensitivity analysis for the proposed study, which extends the Anderson-Rubin (1949) test with possibly invalid IV. The power formula is derived in Jiang, Small and Zhang (2013).

**Author(s)**
Dylan Small

**References**

**Examples**

```r
### Power for a study with in which the null hypothesis causal effect is 0,
### the true causal effect is 1, the sample size is 250, the instrument is
### binary with probability .5 (so variance = .25), the standard deviation
### of potential outcome under control is 1, the effect of the instrument
### is to increase the probability of a binary treatment being 1 from .25 to
### .75. The function sigmav.func computes the SD of v for a binary instrument,
### binary treatment. The correlation between u and v is assumed to be .5. The
### significance level for the study will be alpha = .05

# sigmav.func finds sigmav = .4330127
```
### ARsensitivity.size

#### Description

Computes the minimum sample size required for achieving certain power of sensitivity analysis, which is based on an extension of Anderson-Rubin test and allows IV be possibly invalid within a certain range.

#### Usage

```
ARsensitivity.size(power, k, lambda, gamma, var.z, sigma1, sigma2, rho, alpha = 0.05, Delta=NULL, delta=NULL)
```

#### Arguments

- `power`: The goal of power achieving over a constant.
- `k`: Number of exogenous variables, intercept is also counted. If there’s no observed variable, k = 1.
- `lambda`: True causal effect minus null hypothesis causal effect.
- `gamma`: Regression coefficient for effect of instrument on treatment.
- `var.z`: Variance of instrument.
- `sigma1`: Standard deviation of potential outcome under control (structural error for y).
- `sigma2`: Standard deviation of error from regressing treatment on instrument.
- `rho`: Correlation between u (potential outcome under control) and v (error from regressing treatment on instrument).
- `alpha`: Significance level of test.
- `Delta`: Range of sensitivity allowance. Usually a numeric vector of length 2. If Delta=NULL, the ARsensitivity.power function will calculate the power for standard Anderson-Rubin test with valid IV.
- `delta`: True value of sensitivity parameter when calculating the power. Usual take delta=0 for the favourable situation or delta=NULL for unknown delta.
Details

The structural equations model assumed is: \( D = \gamma_0 + \gamma z + v \), \( Y = \beta_0 + \beta_1 D + \delta Z + u \). This model can also be obtained by assuming the potential outcomes model. See Jiang, Zhang and Small (2013) for details.

\( \lambda \) is equal to the true \( \beta_1 \) minus the null hypothesis \( \beta_1 \).

Value

Minimum sample size required for achieving certain power of sensitivity analysis for the proposed study, which extends the Anderson-Rubin (1949) test with possibly invalid IV. The power formula is derived in Jiang, Small and Zhang (2013).

Author(s)

Dylan Small

References


Examples

```R
### Minimum sample size needed for power of sensitivity analysis over 0.8.
### In a study where the null hypothesis causal effect is 0,
### the true causal effect is 1, the sample size is 250, the instrument is
### binary with probability .5 (so variance = .25), the standard deviation
### of potential outcome under control is 1, the effect of the instrument
### is to increase the probability of a binary treatment being 1 from .25 to
### .75. The function sigmav.func computes the SD of v for a binary instrument,
### binary treatment. The correlation between u and v is assumed to be .5. The
### significance level for the study will be alpha = .05
sigmaR = 0.4330127 # Minimum sample size for Anderson-Rubin test
sigma = 0.4330127
rho = 0.5
alpha = 0.05
pars = ARsensitivity.size(power=0.8, k=1, lambda=1, gamma=0.5, var.z=0.25, sigma=0.5, sigma2=0.4330127, rho=0.5, alpha = 0.05)
```

```R
### minimum sample size for sensitivity analysis under the favourable situation,
### assuming the range of sensitivity allowance is (-0.1, 0.1)
ARsensitivity.size(power=0.8, k=1, lambda=1, gamma=0.5, var.z=0.25, sigma=0.5, sigma2=0.4330127, rho=0.5, alpha = 0.05, Delta=c(-0.1, 0.1), delta=0)
```

```R
### minimum sample size for sensitivity analysis with unknown delta, assuming
### the range of sensitivity allowance is (-0.1, 0.1)
ARsensitivity.size(power=0.8, k=1, lambda=1, gamma=0.5, var.z=0.25, sigma=0.5, sigma2=0.4330127, rho=0.5, alpha = 0.05, Delta=c(-0.1, 0.1))
```
Description

Data from the National Longitudinal Survey of Young Men (NLSYM) that was used by Card (1995).

Usage

data(card.data)

Format

A data frame with 3010 observations on the following 35 variables.

id  subject id
nearc2  indicator for whether a subject grew up near a two-year college
nearc4  indicator for whether a subject grew up near a four-year college
educ  subject’s years of education
age  subject’s age at the time of the survey in 1976
fatheduc  subject’s father’s years of education
motheduc  subject’s mother’s years of education
weight  sampling weight
momdad14  indicator for whether subject lived with both mother and father at age 14
sinmom14  indicator for whether subject lived with single mom at age 14
step14  indicator for whether subject lived with step-parent at age 14
reg661  indicator for whether subject lived in region 1 (New England) in 1966
reg662  indicator for whether subject lived in region 2 (Middle Atlantic) in 1966
reg663  indicator for whether subject lived in region 3 (East North Central) in 1966
reg664  indicator for whether subject lived in region 4 (West North Central) in 1966
reg665  indicator for whether subject lived in region 5 (South Atlantic) in 1966
reg666  indicator for whether subject lived in region 6 (East South Central) in 1966
reg667  indicator for whether subject lived in region 7 (West South Central) in 1966
reg668  indicator for whether subject lived in region 8 (Mountain) in 1966
reg669  indicator for whether subject lived in region 9 (Pacific) in 1966
south66  indicator for whether subject lived in South in 1966
black  indicator for whether subject’s race is black
smsa  indicator for whether subject lived in SMSA in 1976
south  indicator for whether subject lived in the South in 1976
smsa66  indicator for whether subject lived in SMSA in 1966
wage subject's wage in cents per hour in 1976
enroll indicator for whether subject is enrolled in college in 1976
kww subject’s score on the Knowledge of the World of Work (KWW) test in 1966
IQ IQ-type test score collected from the high school of the subject.
marrried indicator for whether the subject was married in 1976.
libcrd14 indicator for whether subject had library card at age 14.
exper subject’s years of labor force experience in 1976
lwage subject’s log wage in 1976
expersq square of subject’s years of labor force experience in 1976
region region in which subject lived in 1976

Source

Examples
data(card.data)

cluster.robust.se

Description
Computes cluster robust standard errors for a two-stage least squares instrumental variable analysis.

Usage
cluster.robust.se(ivmodel, clusterid)

Arguments
ivmodel A model object fit using the ivreg command from the AER package.
clusterid A vector that contains an identifier for the cluster of each subject.

Details
The standard errors are computed using the method of White (1982) that assumes observations within a cluster may be dependent but the clusters are independent.

Value
Coefficient estimates, cluster robust standard errors and p-values using cluster robust standard errors.
Author(s)
Dylan Small

References

See Also
ivreg

Examples
# For Card's data, fit an IV model of log wage on the treatment variable (education)
# using the IV nearc4, with measured covariates (included exogenous variables)
# exper, expersq, black, south, smsa, smsa66
data(card.data)
ivmodel=ivreg(lwage ~ educ + exper + expersq + black + south + smsa + smsa66,
~ nearc4 + exper + expersq + black + south + smsa + smsa66, x=TRUE, data=card.data)
# Compute cluster robust standard errors when the clustering is by region
cluster.robust.se(ivmodel, card.data$region)

description
This is an internal function for computing cluster robust standard errors.

Usage
clx(fm, cluster)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>fm</td>
<td>Model fit.</td>
</tr>
<tr>
<td>cluster</td>
<td>Cluster identifier.</td>
</tr>
</tbody>
</table>

Author(s)
This function was created by Mahmood Arai and adapted by Dylan Small for use in the ivpack package.
Description
Computes the power for an instrumental variables analysis to be done using the Anderson-Rubin test.

Usage
```
power.iv(n, lambda, gamma, var.z, sigmav, rho, alpha = 0.05)
```

Arguments
- `n` Sample size.
- `lambda` True causal effect minus null hypothesis causal effect
- `gamma` Regression coefficient for effect of instrument on treatment.
- `var.z` Variance of instrument.
- `sigmav` Standard deviation of potential outcome under control (structural error for y)
- `sigmau` Standard deviation of error from regressing treatment on instrument
- `rho` Correlation between u(potential outcome under control) and v (error from regressing treatment on instrument)
- `alpha` Significance level of test.

Details
The structural equations model assumed is: \( D = \gamma_0 + \gamma z + v \), \( Y = \beta_0 + \beta_1 D + u \). This model can also be obtained by assuming the potential outcomes model \( Y^{d=0} = \beta_0 + u \), \( Y^d = Y^{d=0} + \beta_1 \). See Jiang, Small and Zhang (2013) for details.

lambda is equal to the true \( \beta_1 \) minus the null hypothesis \( \beta_1 \).

Value
Power for the proposed study, assuming that the Anderson-Rubin (1949) test will be used. The power formula is derived in Jiang, Small and Zhang (2013).

Author(s)
Dylan Small

References
Examples

### Power for a study with in which the null hypothesis causal effect is 0, the true causal effect is 1, the sample size is 250, the instrument is binary with probability .5 (so variance = .25), the standard deviation of potential outcome under control is 1, the effect of the instrument is to increase the probability of a binary treatment being 1 from .25 to .75. The function sigmav.func computes the SD of \( v \) for a binary instrument, binary treatment. The correlation between \( u \) and \( v \) is assumed to be .5. The significance level for the study will be alpha = .05.

```r
corruc.func(prob.d1.given.z1=.75, prob.d1.given.z0=.25, prob.z1=.5)
# The sigmav.func finds sigmav=.4330127
power.iv(n=250, lambda=1, gamma=.5, var.z=.25, sigmav=1, sigmav=.4330127, rho=.5, alpha = .05)
```

Description

Compute robust to heteroskedasticity standard errors for an instrumental variables analysis. These are the Huber-White standard errors for an instrumental variable analysis as described in White (1982).

Usage

```r
robust.se(ivmodel)
```

Arguments

- `ivmodel`: Model object fit by ivreg.

Value

Coefficient estimates, robust standard errors and t-tests based on the robust standard errors.

Author(s)

Dylan Small

References


See Also

`ivreg`
Examples

```r
## This is the IV model in panel A, column (5) of Table 3 from Card, 1995, "Using
## Geographic Variation in College Proximity to Estimate the Return from Schooling"
data(card.data)
ivmodel=ivreg(lwage ~ educ + exper + expersq + black + south + smsa + reg661 + reg662 +
reg663 + reg664 + reg665 + reg666 + reg667 + reg668 + smsa66, ~ nearc4 + exper +
expersq + black + south + smsa + reg661 + reg662 + reg663 + reg664 + reg665 + reg666 +
reg667 + reg668 + smsa66, x=TRUE, data=card.data)
robust.se(ivmodel)
```

Description

Calculates the standard deviation of the error when a linear probability model is fit to predict a
binary treatment based on a binary instrument.

Usage

```r
sigmav(func(prob.d1.given.z1, prob.d1.given.z0, prob.z1)
```

Arguments

- `prob.d1.given.z1`: Probability that the treatment D equals 1 given that the instrumental variable Z
  equals 1.
- `prob.d1.given.z0`: Probability that the treatment D equals 1 given that the instrumental variable Z
  equals 0.
- `prob.z1`: Probability that the instrumental variable Z equals 1.

Value

Standard deviation of the error v from $D=E(D|Z)+v$.

Author(s)

Dylan Small

Examples

```r
## sigmav when P(D=1|Z=1)=.75, P(D=1|Z=0)=.25, P(Z=1)=.5
sigmav(func(prob.d1.given.z1=.75,prob.d1.given.z0=.25,prob.z1=.5)
```
Index

*Topic \textasciitilde kwd1
  anderson.rubin.ci, 3
  ARsensitivity.ci, 4
  ARsensitivity.power, 5
  ARsensitivity.size, 7
  cluster.robust.se, 10
  power.iv, 12
  robust.se, 13
  sigmav.func, 14
*Topic \textasciitilde kwd2
  anderson.rubin.ci, 3
  ARsensitivity.ci, 4
  ARsensitivity.power, 5
  ARsensitivity.size, 7
  cluster.robust.se, 10
  power.iv, 12
  robust.se, 13
  sigmav.func, 14
*Topic \textasciitilde datasets
  card.data, 9
*Topic \textasciitilde package
  ivpack-package, 2
  anderson.rubin.ci, 3
  ARsensitivity.ci, 4
  ARsensitivity.power, 5
  ARsensitivity.size, 7
  card.data, 9
  cluster.robust.se, 10
  clx, 11
  ivpack (ivpack-package), 2
  ivpack-package, 2
  ivreg, 2, 4, 5, 11, 13
  power.iv, 12
  robust.se, 13
  sigmav.func, 14