Package ‘DBfit’

April 30, 2021

Type Package

Title A Double Bootstrap Method for Analyzing Linear Models with Autoregressive Errors

Version 2.0

Date 2021-04-30

Author Joseph W. McKean and Shaofeng Zhang

Maintainer Shaofeng Zhang <shaofeng.zhang@wmich.edu>

Description Computes the double bootstrap as discussed in McKnight, McKean, and Huitema (2000) <doi:10.1037/1082-989X.5.1.87>.

The double bootstrap method provides a better fit for a linear model with autoregressive errors than ARIMA when the sample size is small.

License GPL (>= 2)

Depends Rfit

NeedsCompilation no

Repository CRAN

Date/Publication 2021-04-30 20:30:02 UTC

R topics documented:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBfit-package</td>
<td>2</td>
</tr>
<tr>
<td>boot1</td>
<td>3</td>
</tr>
<tr>
<td>boot2</td>
<td>4</td>
</tr>
<tr>
<td>dbfit</td>
<td>5</td>
</tr>
<tr>
<td>durbin1fit</td>
<td>7</td>
</tr>
<tr>
<td>durbin1xy</td>
<td>8</td>
</tr>
<tr>
<td>durbin2fit</td>
<td>8</td>
</tr>
<tr>
<td>fullr</td>
<td>9</td>
</tr>
<tr>
<td>hmdesign2</td>
<td>10</td>
</tr>
<tr>
<td>hmat</td>
<td>10</td>
</tr>
<tr>
<td>hypothmat</td>
<td>11</td>
</tr>
<tr>
<td>lagx</td>
<td>12</td>
</tr>
<tr>
<td>nurho</td>
<td>13</td>
</tr>
</tbody>
</table>
DBfit-package

A Double Bootstrap Method for Analyzing Linear Models With Autoregressive Errors

Description

Computes the double bootstrap as discussed in McKnight, McKeen, and Huitema (2000) <doi:10.1037/1082-989X.5.1.87>. The double bootstrap method provides a better fit for a linear model with autoregressive errors than ARIMA when the sample size is small.

Details

The DESCRIPTION file:

Package: DBfit
Type: Package
Title: A Double Bootstrap Method for Analyzing Linear Models with Autoregressive Errors
Version: 2.0
Date: 2021-04-30
Author: Joseph W. McKean and Shaofeng Zhang
Maintainer: Shaofeng Zhang <shaofeng.zhang@wmich.edu>
Description: Computes the double bootstrap as discussed in McKnight, McKeen, and Huitema (2000) <doi:10.1037/1082-989X.5.1.87>
License: GPL (>= 2)
Depends: Rfit

Index of help topics:

DBfit-package A Double Bootstrap Method for Analyzing Linear Models With Autoregressive Errors
boot1 First Bootstrap Procedure For parameter estimations
boot2 First Bootstrap Procedure For parameter estimations
dbfit The main function for the double bootstrap method
durbin1fit Durbin stage 1 fit
Author(s)

Joseph W. McKean and Shaofeng Zhang

Maintainer: Shaofeng Zhang <shaofeng.zhang@wmich.edu>

References


Description

Function performing the first bootstrap procedure to yield the parameter estimates

Usage

boot1(y, phi1, arp, nbs, x, allb, method, scores)
Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>the response variable</td>
</tr>
<tr>
<td>phi1</td>
<td>the Durbin two-stage estimate of the autoregressive parameter rho</td>
</tr>
<tr>
<td>arp</td>
<td>the order of autoregressive errors</td>
</tr>
<tr>
<td>nbs</td>
<td>the bootstrap size</td>
</tr>
<tr>
<td>x</td>
<td>the original design matrix (including intercept), without centering</td>
</tr>
<tr>
<td>allb</td>
<td>all the Durbin two-stage estimates of the regression coefficients</td>
</tr>
<tr>
<td>method</td>
<td>If &quot;OLS&quot;, uses the ordinary least square; If &quot;RANK&quot;, uses the rank-based fit</td>
</tr>
<tr>
<td>scores</td>
<td>Default is Wilcoxon scores</td>
</tr>
</tbody>
</table>

Value

An estimate of the bias is returned

Note

This function is for internal use. The main function for users is dbfit.

boot2

First Bootstrap Procedure For parameter estimations

Description

Function performing the second bootstrap procedure to yield the inference of the regression coefficients

Usage

boot2(y, xcopy, phi1, beta, nbs, method, scores)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>the response variable</td>
</tr>
<tr>
<td>xcopy</td>
<td>the original design matrix (including intercept), without centering</td>
</tr>
<tr>
<td>phi1</td>
<td>the estimate of the autoregressive parameter rho from the first bootstrap procedure</td>
</tr>
<tr>
<td>beta</td>
<td>the estimates of the regression coefficients from the first bootstrap procedure</td>
</tr>
<tr>
<td>nbs</td>
<td>the bootstrap size</td>
</tr>
<tr>
<td>method</td>
<td>If &quot;OLS&quot;, uses the ordinary least square; If &quot;RANK&quot;, uses rank-based fit</td>
</tr>
<tr>
<td>scores</td>
<td>Default is Wilcoxon scores</td>
</tr>
</tbody>
</table>
dbfit

Value

betacov the estimate of var-cov matrix of betas
allbeta the estimates of betas inside of the second bootstrap, not the final estimates of betas. The final estimates of betas are still from boot1.
rhostar the estimates of rho inside of the second bootstrap, not the final estimates of rho. The final estimate(s) of rho are still from boot1.
MSEstar MSE used inside of the second bootstrap.

Note

This function is for internal use. The main function for users is dbfit

---

dbfit

The main function for the double bootstrap method

Description

This function is used to implement the double bootstrap method. It is used to yield estimates of both regression coefficients and autoregressive parameters(rho), and also the inference of them.

Usage

```r
## Default S3 method:
dbfit(x, y, arp, nbs = 500, nbscov = 500,
    conf = 0.95, correction = TRUE, method = "OLS", scores, ...)
```

Arguments

x the design matrix, including intercept, i.e. the first column being ones.
y the response variable.
arp the order of autoregressive errors.
nbs the bootstrap size for the first bootstrap procedure. Default is 500.
nbscov the bootstrap size for the second bootstrap procedure. Default is 500.
conf the confidence level of CI for rho, default is 0.95.
correction logical. Currently, ONLY works for order 1, i.e. for order > 1, this correction will not get involved. If TRUE, uses the correction for cases that the estimate of rho is 0.99. Default is TRUE.
method the method to be used for fitting. If "OLS", uses the ordinary least square lm; If "RANK", uses the rank-based fit rfim.
scores Default is Wilcoxon scores
... additional arguments to be passed to fitting routines
Details

Computes the double bootstrap as discussed in McKnight, McKean, and Huitema (2000). For details, see the references.

Value

coefficients  the estimates of regression coefficients based on the first bootstrap procedure
rho1         the Durbin two-stage estimate of the autoregressive parameter rho
adjar        the estimates of regression coefficients based on the first bootstrap procedure
mse          the mean square error
rho_CI_1     the first type of CI for rho, see the second reference for details.
rho_CI_2     the second type of CI for rho, see the second reference for details.
rho_CI_3     the third type of CI for rho, see the second reference for details.
betacov      the estimate of the variance-covariance matrix of betas
tabbeta      a table of point estimates, SE’s, test statistics and p-values.
flag99       an indicator; if 1, it indicates the original fit yields an estimate of rho to be 0.99. When the correction is requested (default), the correction procedure kicks in, and the final estimates of rho is corrected. Only valid if order 1 is specified.
residuals    the residuals, that is response minus fitted values.
fitted.values the fitted mean values.

Author(s)

Joseph W. McKean and Shaofeng Zhang

References


See Also

dbfit.formula

Examples

# make sure the dependent package Rfit is installed
# To save users time, we set both bootstrap sizes to be 100 in this example.
# Defaults are both 500.

# data(testdata)
# This data is generated by a two-phase design, with autoregressive order being one,
# autoregressive coefficient being 0.6 and all regression coefficients being 0.
# Both the first and second phase have 20 observations.
# y <- testdata[,5]
# x <- testdata[,1:4]
# fit1 <- dbfit(x,y,1, nbs = 100, nbscov = 100) # OLS fit, default
# summary(fit1)
# Note that the CI's of autoregressive coef are not shown in the summary.
# Instead, they are attributes of model fit.
# fit1$rho_CI_1

# fit2 <- dbfit(x,y,1, nbs = 100, nbscov = 100 ,method="RANK") # rank-based fit

# When fitting with autoregressive order 2,
# the estimate of the second order autoregressive coefficient should not be significant,
# since this data is generated with order 1.

# fit3 <- dbfit(x,y,2, nbs = 100, nbscov = 100)
# fit3$rho_CI_1 # The first row is lower bounds, and second row is upper bounds

---

durbin1fit

**Durbin stage 1 fit**

**Description**

Function implements the Durbin stage 1 fit

**Usage**

durbin1fit(y, x, arp, method, scores)

**Arguments**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>the response variable in stage 1, not the original response variable</td>
</tr>
<tr>
<td>x</td>
<td>the model matrix in stage 1, not the original design matrix</td>
</tr>
<tr>
<td>arp</td>
<td>the order of autoregressive errors.</td>
</tr>
<tr>
<td>method</td>
<td>the method to be used for fitting. If &quot;OLS&quot;, uses the ordinary least square; If &quot;RANK&quot;, uses the rank-based fit.</td>
</tr>
<tr>
<td>scores</td>
<td>Default is Wilcoxon scores</td>
</tr>
</tbody>
</table>

**Note**

This function is for internal use. The main function for users is dbfit.

**References**

durbin1xy

Creating New X and Y for Durbin Stage 1

Description

Functions provides the transformed response variable and model matrix for Durbin stage 1 fit. For
details of the transformation, see the reference.

Usage

durbin1xy(y, x, arp)

Arguments

y the orginal response variable
x the orginal design matrix with first column of all one’s (corresponding to the intercept)
arp the order of autoregressive errors.

References

linear models with autoregressive error terms. Psychological methods, 5 (1), 87. Shaofeng Zhang

durbin2fit

Durbin stage 2 fit

Description

Function implements the Durbin stage 1 fit

Usage

durbin2fit(yc, xc, adjphi, method, scores)

Arguments

yc a transformed response variable
xc a transformed design matrix
adjphi the Durbin stage 1 estimate(s) of the autoregressive parameters rho
method the method to be used for fitting. If "OLS", uses the ordinary least square; If "RANK", uses the rank-based fit.
scores Default is Wilcoxon scores
**fullr**

**Value**

- beta: the estimates of regression coefficients
- sigma: the estimate of standard deviation of the white noise

**Note**

This function is for internal use. The main function for users is dbfit.

**References**


---

**fullr**  
*QR decomposition for non-full rank design matrix for Rfit.*

**Description**

With Rfit recent update, it cannot return partial results with singular design matrix (as opposed to lm). This function uses QR decomposition for Rfit to resolve this issue, so that dbfit can run robust version.

**Usage**

`fullr(x, p1)`

**Arguments**

- x: design matrix, including intercept, i.e. the first column being ones.
- p1: number of first few columns of x that are linearly independent.

**Note**

This function is for internal use.
hmdesign2  
the Two-Phase Design Matrix

Description
Returns the design matrix for a two-phase intervention model.

Usage
hmdesign2(n1, n2)

Arguments
n1  number of obs in phase 1
n2  number of obs in phase 2

Details
It returns a matrix of 4 columns. As discussed in Huitema, Mckean, & Mcknight (1999), in two-phase design: beta0 = intercept, beta1 = slope for Phase 1, beta2 = level change from Phase 1 to Phase 2, and beta3 slope change from Phase 1 to Phase 2.

References

Examples
n1 <- 15
n2 <- 15
hmdesign2(n1, n2)

hmmat  
K-Phase Design Matrix

Description
Returns the design matrix for a general k-phase intervention model

Usage
hmmat(vecss, k)
Arguments

vecss  a vector of length k with each element being the number of observations in each phase
k    number of phases

Details

It returns a matrix of 2*k columns. The design can be unbalanced, i.e. each phase has different observations.

References


See Also

hmmdesign2

Examples

# a three-phase design matrix
hmmat(c(10,10,10),3)

hypothesismat  General Linear Tests of the regression coefficients

Description

Performs general linear tests of the regressio coefficients.

Usage

hypothesismat(sfit, mmat, n, p)

Arguments

sfit    the result of a call to dbfit.
mmat   a full row rank q*(p+1) matrix, where q is the row number of the matrix and p is number of independent variables.
n    total number of observations.
p    number of independent variables.

Details

This functions performs the general linear F-test of the form H0: Mb = 0 vs HA: Mb != 0.
**Value**

- `tst` the test statistic
- `pvf` the p-value of the F-test

**References**


**Examples**

```r
# data(testdata)
# y<testdata[,5]
# x<testdata[,1:4]
# fit1<dbfit(x,y,1) # OLS fit, default
# a test that H0: b1 = b3 vs HA: b1 != b3
# mat<matrix(c(1,0,0,-1),nrow=1)
# hypothmat(sfit=fit1,mmat=mat,n=40,p=4)
```

---

**lagx**  

**Lag Functions**

**Description**

For preparing the transformed x and y in the Durbin stage 1 fit

**Usage**

```r
lagx(x, s1, s2)
lagmat(x, p)
```

**Arguments**

- `x` a vector or the design matrix, including intercept, i.e. the first column being ones.
- `s1` starting index of the slice.
- `s2` end index of the slice.
- `p` the order of autoregressive errors.

**Note**

These function are for internal use.
Creating a new response variable for Durbin stage 2

**Description**

It returns a new response variable (vector) for Durbin stage 2.

**Usage**

\[
\text{nurho}(\text{yc, adjphi})
\]

**Arguments**

- \text{yc}: the centered response variable y
- \text{adjphi}: (initial) estimate of rho in Durbin stage 1

**Details**

see reference.

**Note**

This function is for internal use. The main function for users is \text{dbfit}.

**References**


---

**print.dbfit**

**DBfit Internal Print Functions**

**Description**

These functions print the output in a user-friendly manner using the internal R function print.

**Usage**

```r
## S3 method for class 'dbfit'
print(x, ...)
## S3 method for class 'summary.dbfit'
print(x, ...)
```
Arguments

- `x` An object to be printed
- `...` additional arguments to be passed to `print`

See Also

dbfit, summary.dbfit

---

**rhoci2**

*A fisher type CI of the autoregressive parameter rho*

Description

This function returns a Fisher type CI for rho, which is then used to correct the .99 cases.

Usage

```
rhoci2(n, rho, cv)
```

Arguments

- `n` total number of observations
- `rho` final estimate of rho, usually .99.
- `cv` critical value for CI

Details

see reference.

Note

This function is for internal use.

References

**Description**

Generates the simulation data for a two-phase intervention model.

**Usage**

```r
simpgen1hm2(n1, n2, rho, beta = c(0, 0, 0, 0))
```

**Arguments**

- `n1`: number of obs in phase 1
- `n2`: number of obs in phase 2
- `rho`: pre-defined autoregressive parameter(s)
- `beta`: pre-defined regression coefficients

**Details**

This function is used for simulations when developing the package. With pre-defined sample sizes in both phases and parameters, it returns a simulated data.

**Value**

`mat`: a matrix containing the simulation data. The last column is the response variable. All other columns make up the design matrix.

**See Also**

`hmdesign2`

**Examples**

```r
n1 <- 15
n2 <- 15
rho <- 0.6
beta <- c(0,0,0,0)
dat <- simpgen1hm2(n1, n2, rho, beta)
dat
```
simula

**Work Horse Function to implement the Double Bootstrap method**

**Description**

`simula` is the original work horse function to implement the DB method. However, when this function returns an estimate of rho to be .99, another work horse function `simulacorrection` kicks in.

**Usage**

```r
simula(x, y, arp, nbs, nbscov, conf, method, scores)
```

**Arguments**

- `x`: the design matrix, including intercept, i.e. the first column being ones.
- `y`: the response variable.
- `arp`: the order of autoregressive errors.
- `nbs`: the bootstrap size for the first bootstrap procedure. Default is 500.
- `nbscov`: the bootstrap size for the second bootstrap procedure. Default is 500.
- `conf`: the confidence level of CI for rho, default is 0.95.
- `method`: the method to be used for fitting. If "OLS", uses the ordinary least square `lm`; If "RANK", uses the rank-based fit `rfit`.
- `scores`: Default is Wilcoxon scores

**Details**

see `dbfit`.

**Note**

Users should use `dbfit` to perform the analysis.

**References**


**See Also**

`dbfit`. 

**simulacorrection**  
*Work Horse Function to Implement the Double Bootstrap Method For .99 Cases*

**Description**

When function *simula* returns an estimate of rho to be .99, this function kicks in and outputs a corrected estimate of rho. Currently, this only works for order 1, i.e. for order > 1, this correction will not get involved.

**Usage**

```r
simulacorrection(x, y, arp, nbs, nbscov, method, scores)
```

**Arguments**

- `x` the design matrix, including intercept, i.e. the first column being ones.
- `y` the response variable.
- `arp` the order of autoregressive errors.
- `nbs` the bootstrap size for the first bootstrap procedure. Default is 500.
- `nbscov` the bootstrap size for the second bootstrap procedure. Default is 500.
- `method` the method to be used for fitting. If "OLS", uses the ordinary least square lm; If "RANK", uses the rank-based fit rfit.
- `scores` Default is Wilcoxon scores

**Details**

If 0.99 problem is detected, then construct Fisher CI for both initial estimate (in Durbin stage 1) and first bias-corrected estimate (perform only one bootstrap, instead of a loop); if the midpoint of latter is smaller than 0.95, then this midpoint is the final estimate for rho; otherwise the midpoint of the former CI is the final estimate.

By default, when function *simula* returns an estimate of rho to be .99, this function kicks in and outputs a corrected estimate of rho. However, users can turn the auto correction off by setting `correction="FALSE"` in *dbfit*. Users are encouraged to investigate why the stationarity assumption is violated based on their experience of time series analysis and knowledge of the data.

**Note**

Users should use *dbfit* to perform the analysis.

**References**


**See Also**

- *dbfit*
summary.dbfit  
*Summarize the double bootstrap (DB) fit*

**Description**

It summarizes the DB fit in a way that is similar to OLS `lm`.

**Usage**

```r
## S3 method for class 'dbfit'
summary(object, ...)
```

**Arguments**

- `object` a result of the call to `rfit`
- `...` additional arguments to be passed

**Value**

- `call` the call to `rfit`
- `tab` a table of point estimates, standard errors, t-ratios and p-values
- `rho1` the Durbin two-stage estimate of rho
- `adjrho` the DB (final) estimate of rho
- `flag99` an indicator; if 1, it indicates the original fit yields an estimate of rho to be 0.99. Only valid if order 1 is specified.

**Examples**

```r
# data(testdata)
# y<-testdata[,5]
# x<-testdata[,1:4]
# fit1<-dbfit(x,y,1) # OLS fit, default
# summary(fit1)
```

---

**Description**

This data serves as a test data.

**Usage**

```r
data("testdata")
```
**Format**

A data frame with 40 observations. First 4 columns make up the design matrix, while the last column is the response variable. This data is generated by a two-phase design, with autoregressive order being one, autoregressive coefficient being 0.6 and all regression coefficients being 0. Both the first and second phase have 20 observations.

**Examples**

```r
data(testdata)
```

---

**Description**

It returns a new design matrix for Durbin stage 2.

**Usage**

```r
wrho(xc, adjphi)
```

**Arguments**

- `xc`: centered design matrix, no column of ones
- `adjphi`: (initial) estimate of rho in Durbin stage 1

**Details**

see reference.

**Note**

This function is for internal use. The main function for users is `dbfit`.

**References**

Index

* datasets
  testdata, 18

boot1, 3
boot2, 4

DBfit (DBfit-package), 2
dbfit, 5, 14, 16, 17
DBfit-package, 2
durbin1fit, 7
durbin1xy, 8
durbin2fit, 8

fullr, 9

hmdesign2, 10, 11, 15
hmmat, 10
hypothmat, 11

lagmat (lagx), 12
lagx, 12

nurho, 13

print.dbfit, 13
print.summary.dbfit (print.dbfit), 13

rhoci2, 14

simpgen1hm2, 15
simula, 16
simulacorrection, 17
summary.dbfit, 14, 18

testdata, 18

wrho, 19