Package ‘DBfit’

October 25, 2018

Type Package
Title A Double Bootstrap Method for Analyzing Linear Models with Autoregressive Errors
Version 1.0
Date 2018-10-16
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 2018-10-24 22:50:03 UTC

R topics documented:

DBfit-package ......................................................... 2
boot1 ................................................................. 3
boot2 ................................................................. 4
dbfit ............................................................... 5
durbin1fit .......................................................... 7
durbin1xy ........................................................... 8
durbin2fit ........................................................... 8
hmdesign2 ............................................................ 9
hmmat ............................................................... 10
hypothmat .......................................................... 11
lagx ................................................................. 12
nurho ............................................................... 12
print.dbfit .......................................................... 13
**DBfit-package**

**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: 1.0
Date: 2018-10-16
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
- `durbin1xy` Creating New X and Y for Durbin Stage 1
Author(s)
Joseph W. McKean and Shaofeng Zhang
Maintainer: Shaofeng Zhang <shaofeng.zhang@wmich.edu>

References

```
boot1

First Bootstrap Procedure For parameter estimations
```

Description
Function performing the first bootstrap procedure to yield the parameter estimates

Usage

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

Arguments

- `y` the response variable
- `phi1` the Durbin two-stage estimate of the autoregressive parameter rho
- `arp` the order of autoregressive errors
Function performing the second bootstrap procedure to yield the inference of the regression coefficients

Usage

\texttt{boot2(y, xcopy, phi1, beta, nbs, method, scores)}

Arguments

- **y**: the response variable
- **xcopy**: the original design matrix (including intercept), without centering
- **phi1**: the estimate of the autoregressive parameter rho from the first bootstrap procedure
- **beta**: the estimates of the regression coefficients from the first bootstrap procedure
- **nbs**: the bootstrap size
- **method**: If "OLS", uses the ordinary least square; If "RANK", uses rank-based fit
- **scores**: Default is Wilcoxon scores

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 \texttt{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 \texttt{boot1}.
- **MSEstar**: MSE used inside of the second bootstrap.
\textit{dbfit} \hfill 5

\textbf{Note}

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

\begin{center}
\begin{tabular}{ll}
\textbf{dbfit} & \textit{The main function for the double bootstrap method} \\
\end{tabular}
\end{center}

\textbf{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.

\textbf{Usage}

\begin{verbatim}
## Default S3 method:
dbfit(x, y, arp, nbs = 500, nbsov = 500,
conf = 0.95, correction = TRUE, method = "OLS", scores, ...)
\end{verbatim}

\textbf{Arguments}

\begin{itemize}
\item \textbf{x} \hspace{1cm} the design matrix, including intercept, i.e. the first column being ones.
\item \textbf{y} \hspace{1cm} the response variable.
\item \textbf{arp} \hspace{1cm} the order of autoregressive errors.
\item \textbf{nbs} \hspace{1cm} the bootstrap size for the first bootstrap procedure. Default is 500.
\item \textbf{nbsov} \hspace{1cm} the bootstrap size for the second bootstrap procedure. Default is 500.
\item \textbf{conf} \hspace{1cm} the confidence level of CI for \( \rho \), default is 0.95.
\item \textbf{correction} \hspace{1cm} 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.
\item \textbf{method} \hspace{1cm} the method to be used for fitting. If "OLS", uses the ordinary least square \texttt{lm}; If "RANK", uses the rank-based fit \texttt{rfit}.
\item \textbf{scores} \hspace{1cm} Default is Wilcoxon scores
\item \textbf{...} \hspace{1cm} additional arguments to be passed to fitting routines
\end{itemize}

\textbf{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 squared 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, l, 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

y the response variable in stage 1, not the original response variable
x the model matrix in stage 1, not the original design matrix
arp the order of autoregressive errors.
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

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 reponse 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)
ar p the order of autoregressive errors.

References

durbin2fit

Durbin stage 2 fit

Description
Function implements the Durbin stage 1 fit

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

Arguments
yc a transformed reponse 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
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


```
headdesignRHn1L nR
```

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**

```r
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**

`hmdesign2`

**Examples**

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

General Linear Tests of the regression coefficients

Description

Performs general linear tests of the regression coefficients.

Usage

hypothenmat(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

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)
hypothenmat(sfit=fit1,mmat=mat,n=40,p=4)
**lagx**

*Log Functions*

**Description**

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

**Usage**

\[
\text{lagx}(x, s1, s2) \\
\text{lagmat}(x, p)
\]

**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 functions are for internal use.

---

**nurho**

*Creating a new response variable for Durbin stage 2*

**Description**

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

**Usage**

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

**Arguments**

- **yc**: the centered response variable y
- **adjphi**: (initial) estimate of rho in Durbin stage 1

**Details**

see reference.
print.dbfit

Note

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

```r
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

```
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.
**simula**

**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**

`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 simulacorrection 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

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 simulacorrection 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**

dbf.  

---

**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
- `adjar` the DB (final) estimate of rho
- `flagYY` 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

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

data(testdata)

Description

Creating a new design matrix for Durbin stage 2

Usage

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

*Topic **datasets**
  testdata, 18

boot1, 3
boot2, 4

DBfit (DBfit-package), 2
dbfit, 5, 13, 15–17
DBfit-package, 2
durbin1fit, 7
durbin1xy, 8
durbin2fit, 8

hmdes2n2, 9, 10, 15
hmmat, 10
hypothmat, 11

lagmat (lagx), 12
lagx, 12

nurho, 12

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

rhoci2, 13

simpgen1hm2, 14
simula, 15
simulacorrection, 16
summary.dbfit, 13, 17

testdata, 18

wrho, 18