

Package ‘LSEbootLS’

July 1, 2024

Date 2024-06-28

Type Package

Title Bootstrap Methods for Regression Models with Locally Stationary Errors

Version 0.1.0

Maintainer Nicolas Loyola <nloyola2016@udec.cl>

Description Implements bootstrap methods for linear regression models with errors following a time-varying process, focusing on approximating the distribution of the least-squares estimator for regression models with locally stationary errors. It enables the construction of bootstrap and classical confidence intervals for regression coefficients, leveraging intensive simulation studies and real data analysis.

License GPL (>= 3)

Encoding UTF-8

LazyData TRUE

Depends doParallel, R (>= 2.10)

Imports foreach, doRNG, stats, parallel, LSTS, tibble, iterators, rlecuyer

RoxygenNote 7.3.1

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Guillermo Ferreira [aut],
Joel Muñoz [aut],
Nicolas Loyola [aut, cre]

Repository CRAN

Date/Publication 2024-07-01 10:10:03 UTC

Contents

application	2
Coveragelongmemory	4
Coverageshortmemory	6
USinf	9

Index	10
--------------	-----------

application	<i>Calculate the bootstrap LSE for a long memory model</i>
-------------	--

Description

Bootstrap procedure to approximate the sampling distribution of the LSE for time series linear regression with errors following a Locally Stationary process.

Usage

```
application(
  formula,
  data,
  start,
  d.order,
  s.order,
  N,
  S,
  B = 1,
  nr.cores = 1,
  seed = 123
)
```

Arguments

formula	(type: formula) an object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted. The details of model specification are given under 'Details'.
data	(type: data.frame) data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model.
start	(type: numeric) numeric vector, initial values for parameters to run the model.
d.order	(type: numeric) polynomial order, where d is the ARFIMA parameter.
s.order	(type: numeric) polynomial order noise scale factor.
N	(type: numeric) sample size of each block.
S	(type: numeric) shifting places from block to block. Observe that the number of blocks M is determined by the following formula $M = \lfloor \frac{T-N}{S} + 1 \rfloor$, where $\lfloor . \rfloor$ takes a single numeric argument x and returns a numeric vector containing the integers formed by truncating the values in x toward 0.

B	(type: numeric) bootstrap replicates, 1 by default.
nr.cores	(type: numeric) number of CPU cores to be used for parallel processing. 1 by default.
seed	(type: numeric) random number generator seed to generate the bootstrap samples.

Details

This function estimates the parameters in the linear regression model for $t = 1, \dots, T$,

$$Y_{t,T} = X_{t,T}\beta + \epsilon_{t,T},$$

where the error term $\epsilon_{t,T}$ follows a Locally Stationary Autoregressive Fractionally Integrated Moving Average (LS-ARFIMA) structure, given by:

$$\epsilon_{t,T} = (1 - B)^{-d(u)}\sigma(u)\eta_t,$$

where $u=t/T$ in $[0,1]$, $d(u)$ represents the long-memory parameter, $\sigma(u)$ is the noise scale factor, and $\{\eta_t\}$ is a white noise sequence with zero mean and unit variance.

Particularly, we model $d(u)$ and $\sigma(u)$ as polynomials of order $d.order$ and $s.order$ respectively.

$$d(u) = \sum_{i=0}^{d.order} \delta_i u^i,$$

$$\sigma(u) = \sum_{j=0}^{s.order} \alpha_j u^j,$$

For more details, see references.

Value

A list with the following elements:

- **coeff**: A tibble of estimated model coefficients, including intercepts, regression coefficients (β), and coefficients of the δ and α polynomials. Contains columns for coefficient name, estimate, t-value and p-value.
- **estimation**: A matrix of bootstrap replicates for regression coefficients (β).
- **delta**: A matrix of bootstrap replicates for the δ polynomial coefficients.
- **alpha**: A matrix of bootstrap replicates for the α polynomial coefficients.

References

Ferreira G., Mateu J., Vilar J.A., Muñoz J. (2020). Bootstrapping regression models with locally stationary disturbances. TEST, 30, 341-363.

Examples

```

n <- length(USinf)
shift<-201
u1<-c((1:shift)/shift,rep(0, n-shift))
u2<-c(rep(0, shift),(1:(n-shift))/(n-shift))
u<-(1:n)/n
switch <- c(rep(1,shift), rep(0, n-shift))
x1<-switch*u
x2<-(1-switch)*u

test <- data.frame(USinf, x1=x1, x2=x2)

application(formula=USinf~x1+x2,data=test,N=150,S=50,B=10,
start = c(0.16,2.0,-7,8,-3,0.25,-0.25,0.01),
d.order=4,s.order=2,nr.cores=1)

```

Coveragelongmemory *Calculate the coverage of several long-memory models*

Description

Generates coverage metrics for a parameter of interest using a specified long-memory model.

Usage

```

Coveragelongmemory(
  n,
  R,
  N,
  S,
  mu = 0,
  dist,
  method,
  B = NULL,
  nr.cores = 1,
  seed = 123,
  alpha,
  beta,
  start,
  sign = 0.05
)

```

Arguments

n (type: numeric) size of the simulated series.
R (type: numeric) number of realizations of the Monte Carlo experiments.

N	(type: numeric) sample size of each block.
S	(type: numeric) shifting places from block to block. Observe that the number of blocks M is determined by the following formula $M = \lfloor \frac{T-N}{S} + 1 \rfloor$, where $\lfloor \cdot \rfloor$ takes a single numeric argument x and returns a numeric vector containing the integers formed by truncating the values in x toward 0.
mu	(type: numeric) trend coefficient of the regression model.
dist	(type: character) white noise distribution for calculating coverage, it includes the "normal", "exponential" and "uniform" univariate distributions.
method	(type: character) methods are asymptotic ("asym"), bootstrap percentile ("boot") and bootstrap-t ("boot t").
B	(type: numeric) the number of bootstrap replicates, NULL indicates the asymptotic method.
nr.cores	(type: numeric) number of CPU cores to be used for parallel processing. 1 by default.
seed	(type: numeric) random number generator seed to generate the bootstrap samples.
alpha	(type: numeric) numeric vector with values to simulate the time varying autoregressive parameters of model LSAR(1), $\phi(u)$.
beta	(type: numeric) numeric vector with values to simulate the time varying scale factor parameters of model LSAR(1), $\sigma(u)$.
start	(type: numeric) numeric vector, initial values for parameters to run the model.
sign	nominal significance level

Details

This function estimates the parameters in the linear regression model for $t = 1, \dots, T$,

$$Y_{t,T} = X_{t,T}\beta + \epsilon_{t,T},$$

where a locally stationary fractional noise process (LSFN) is described by the equation:

$$\epsilon_{t,T} = \sum_{j=0}^{\infty} \psi_j(u) \eta_{t-j}$$

for $u=t/T$ in $[0,1]$, where $\psi_j(u) = \frac{\Gamma[j+d(u)]}{\Gamma[j+1]\Gamma[d(u)]}$ and $d(u)$ is the smoothly varying long-memory coefficient. This model is referred to as locally stationary fractional noise (LSFN).

In this particular case, $d(u)$ is modeled as a linear polynomial, and $\sigma(u)$ as a quadratic polynomial.

Resampling methods evaluated:

- **asym**: Asymptotic method that uses the asymptotic variance of the estimator, based on the Central Limit Theorem, to construct confidence intervals under the assumption of normality in large samples.
- **boot**: Standard bootstrap that generates replicas of the estimator $\hat{\beta}$ by resampling the adjusted residuals $\hat{\epsilon}_t$. It approximates the distribution of the estimator by the variability observed in the bootstrap replicas of $\hat{\beta}$.

- `boott`: Adjusted bootstrap that scales the bootstrap replicas of the estimator $\hat{\beta}$ by its standard error, aiming to refine the precision of the confidence interval and adjust for the variability in the parameter estimation.

For more details, see references.

Value

A data frame containing the following columns:

- `n`: Size of each simulated series.
- `method`: Statistical method used for simulation.
- `coverage`: Proportion of true parameter values within the intervals.
- `avg_width`: Average width of the intervals.
- `sd_width`: Standard deviation of the interval widths.

References

Ferreira G., Mateu J., Vilar J.A., Muñoz J. (2020). Bootstrapping regression models with locally stationary disturbances. *TEST*, 30, 341-363.

Examples

```
Coveragelongmemory(n=500,R=5,N=60,S=40,mu=0.5,dist="normal",method="asym",
beta=c(0.1,-2),alpha=c(0.15,0.25, 0.1),start = c(0.1,-2,0.15,0.2, 0.1))
```

Coverageshortmemory *Calculate the coverage for several short-memory models*

Description

Generates coverage metrics for a parameter of interest using a specified short-memory model.

Usage

```
Coverageshortmemory(
  n,
  R,
  N,
  S,
  mu,
  dist,
  method,
  alpha,
  beta,
  start,
```

```

Subdivisions = 100,
m = 500,
NN = 100,
B,
case,
sign = 0.05
)

```

Arguments

n	(type: numeric) size of the simulated series.
R	(type: numeric) number of realizations of the Monte Carlo experiments.
N	(type: numeric) sample size of each block.
S	(type: numeric) shifting places from block to block. Observe that the number of blocks M is determined by the following formula $M = \lfloor \frac{T-N}{S} + 1 \rfloor$, where $\lfloor \cdot \rfloor$ takes a single numeric argument x and returns a numeric vector containing the integers formed by truncating the values in x toward \emptyset .
mu	(type: numeric) trend coefficient of the regression model.
dist	(type: character) white noise distribution for calculating coverage, it includes the "normal", "exponential" and "uniform" univariate distributions.
method	(type: character) methods are asymptotic ("asym"), bootstrap percentile ("boot"), bootstrap-t ("boott") and bootstrap-SP ("bootSP").
alpha	(type: numeric) numeric vector with values to simulate the time varying autoregressive parameters of model LSAR(1), $\phi(u)$.
beta	(type: numeric) numeric vector with values to simulate the time varying scale factor parameters of model LSAR(1), $\sigma(u)$.
start	(type: numeric) numeric vector, initial values for parameters to run the model.
Subdivisions	(type: numeric) the number of subintervals produced in the subdivision (integration) process; only required in the asymptotic method.
m	(type: numeric) parameter that allows to remove the first m observations when simulating the LSAR process.
NN	(type: numeric) parameter that allows to remove the first NN observations of noise from the LSAR model.
B	(type: numeric) the number of bootstrap replicates, NULL indicates the asymptotic method.
case	(type: character) nonlinear ("non-linear") and linear cases ("linear").
sign	nominal significance level

Details

This function estimates the parameters in the linear regression model for $t = 1, \dots, T$,

$$Y_{t,T} = X_{t,T}\beta + \epsilon_{t,T},$$

where a locally stationary autoregressive process of order one (LSAR(1)) is described by the equation:

$$\epsilon_{t,T} = \phi(u)\epsilon_{t-1,T} + \sigma(u)\eta_t$$

where $u=t/T$ in $[0,1]$, with $\phi(u)$ is the autoregressive coefficient which is modeled as a linear polynomial, $\sigma(u)$ is modeled as a quadratic polynomial, and η_t is a white noise sequence with zero mean and unit variance. This setup is referred to as a locally stationary autoregressive model (LSAR(1)).

Resampling methods evaluated:

- **asym**: Asymptotic method that uses the asymptotic variance of the estimator, based on the Central Limit Theorem, to construct confidence intervals under the assumption of normality in large samples.
- **boot**: Standard bootstrap that generates replicas of the estimator $\hat{\beta}$ by resampling the adjusted residuals $\hat{\epsilon}_t$. It approximates the distribution of the estimator by the variability observed in the bootstrap replicas of $\hat{\beta}$.
- **boott**: Adjusted bootstrap that scales the bootstrap replicas of the estimator $\hat{\beta}$ by its standard error, aiming to refine the precision of the confidence interval and adjust for the variability in the parameter estimation.
- **bootSP**: Symmetrized Percentile-t method, a variation of the boot-t that symmetrizes the bootstrap distribution around zero to handle skewed distributions or outliers more effectively. This method enhances the accuracy of confidence intervals by adjusting for asymmetries in the bootstrap replicas.

For more details, see references.

Value

A data frame containing the following columns:

- **n**: Size of each simulated series.
- **method**: Statistical method used for simulation.
- **coverage**: Proportion of true parameter values within the intervals.
- **avg_width**: Average width of the intervals.
- **sd_width**: Standard deviation of the interval widths.

References

Ferreira G., Mateu J., Vilar J.A., Muñoz J. (2020). Bootstrapping regression models with locally stationary disturbances. TEST, 30, 341-363.

Examples

```
Coverageshortmemory(n=100,R=10,N=60,S=40,mu=0.5,dist="normal",method="asym",alpha=c(0.25,0.2),
beta=c(1,1,-0.5),start=c(0.15,0.15,1,1,-0.5),case="no-linear")
```

USinf	<i>US Monthly Inflation Data</i>
-------	----------------------------------

Description

Monthly inflation rates for the United States. The data covers the period from January 1965 to December 2011, totaling 564 observations.

Usage

USinf

Format

A time series object with 564 elements
Monthly inflation rate, expressed as a percentage.

Source

International Financial Statistics (IFS)

Index

* **datasets**

USinf, 9

application, 2

Coveragelongmemory, 4

Coverageshortmemory, 6

USinf, 9