Package ‘ForwardSearch’

February 19, 2015

Type Package
Title Forward Search using asymptotic theory
Version 1.0
Date 2014-09-10
Author Bent Nielsen
Maintainer Bent Nielsen <bent.nielsen@nuffield.ox.ac.uk>

URL http://users.ox.ac.uk/~nuff0078/

Description Forward Search analysis of time series regressions. Implements the asymptotic theory developed in Johansen and Nielsen (2013, 2014).

License GPL-3
Depends robustbase

R topics documented:

ForwardSearch-package ............................................. 2
ForwardSearch.fit ................................................. 4
ForwardSearch.plot ................................................ 6
ForwardSearch.pointwise.asymptotics ............................ 8
ForwardSearch.stopped .......................................... 9
Fulton ................................................................. 11

Index 13
ForwardSearch-package

Description

The Forward Search algorithm is an iterative algorithm for multiple (time series) regression suggested by Hadi and Simonoff (1993) and developed further by Atkinson and Riani (2000). The algorithm starts with a robust estimate of the regression parameters and a sub-sample of size $m_0$ and iterates with a sequence of least squares steps. The asymptotic theory developed by Johansen and Nielsen (2013, 2014) is implemented.

Details

<table>
<thead>
<tr>
<th>Package</th>
<th>ForwardSearch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Package</td>
</tr>
<tr>
<td>Version</td>
<td>1.0</td>
</tr>
<tr>
<td>Date</td>
<td>2014-09-09</td>
</tr>
<tr>
<td>License</td>
<td>GPL-3</td>
</tr>
</tbody>
</table>

The Forward Search algorithm is an iterative algorithm for multiple (time series) regression suggested by Hadi and Simonoff (1993) and developed further by Atkinson and Riani (2000). The algorithm starts with a robust estimate of the regression parameters and a sub-sample of size $m_0$. A common choice for the initial estimator is the Least Trimmed Squares estimator of Rousseeuw (1984).

The algorithm is initiated by computing the absolute residuals for all $n$ observations. The initial sub-sample consists of the observations with the smallest $m_0$ absolute residuals. We then run a regression on those $m_0$ observations and compute absolute residuals of all $n$ observations. The observations with $m_0 + 1$ smallest residuals are then selected. The $m_0 + 1$ smallest residual is the forward residual. A new regression is performed on these $m_0 + 1$ observations. This is then iterated. Eventually the least squares estimator based on all $n$ observations is computed.

The algorithm results in a sequence of forward residuals indexed by the sub-sample size $m$ running from $m_0$ to $n - 1$. The idea is to monitor the plot of these and stop when the forward residuals become "large". Johansen and Nielsen (2013, 2014) have developed, respectively, pointwise and simultaneous confidence bands for estimators and forward residuals. These are implemented in the package.

The ForwardSearch package can be used as follows.

1. Execute the full Forward Search using `ForwardSearch.fit`.
2. Create the forward plot of the forward residuals using `ForwardSearch.plot`. This requires the output from above and a choice of reference distribution. The plot shows the scaled forward residuals from above along with simultaneous confidence bands. The user has to choose a "gauge", which is the expected fraction of falsely detected outliers that are tolerable when in fact there are no outliers. For instance a "gauge" of 0.01 indicates that in a sample of $n=110$
observations 1.1 outlier is found on average when there are none. The simultaneous confidence bands are calibrated so that the Forward Search stop when the fitted values exceed the chosen confidence bands the first time. This is a stopping time. The theory for this is given in Johansen and Nielsen.

3. Get the estimates of the stopped Forward Search using `ForwardSearch.stopped`. The user has to input the estimated stopping time. This also gives the rank of the selected and non-selected observations. These are the "good" and the "bad" observations.

Author(s)

Bent Nielsen <bent.nielsen@nuffield.ox.ac.uk> 9 Sep 2014

References


See Also

Forward Search can alternatively be done by the package `forward`. `forward` version 1.0.3 includes functions for the analysis suggested in e.g. Atkinson and Riani (2000), but does not include the asymptotic theory of Johansen and Nielsen (2013, 2014). Matlab code for Forward Search is also available from [www.riani.it](http://www.riani.it).

Examples

```r
# EXAMPLE 1
# using Fulton Fish data,
# see Johansen and Nielsen (2014).

# Call package
library(ForwardSearch)

# Call data
data(Fulton)
mdata <- as.matrix(Fulton)
n <- nrow(mdata)

# Identify variable to reproduce Johansen and Nielsen (2014)
q <- mdata[2:n,9]
q_1 <- mdata[1:(n-1),9]
```
Execute the Forward Search Algorithm.

Description


Usage

ForwardSearch.fit(x.1, y, psi.0 = 0.5, m.0 = NULL, beta.0 = NULL)

Arguments

- **x.1**
  - Matrix of dimension n x (dim.x -1). Design matrix for regressors apart from constant.

- **y**
  - Vector of dimension n. Dependent variable.

- **psi.0**
  - proportion of observations in initial set of set of selected observations. Default is 0.5. Initial set has round(n*psi.0) observations.

- **m.0**
  - Number of observations in initial set of selected observations. Default is NULL. If value is given this overrides psi.0.

- **beta.0**
  - Vector of dimension dim.x. Initial estimator for regression coefficient. Default is NULL, which results in Least Trimmed Squares estimator through beta.0 <- ltsReg(y~x.1, alpha=psi/2).
Details

Dimensions: \( n \) is the number of observations. \( \text{dim.x} \) is the number of regressors (including intercept).

Default is initial estimator is the Least Trimmed Squares estimator of Rousseeuw (1984) implemented as \texttt{ltsReg} in package \texttt{robustbase}.

The breakdown point of the initial Least Trimmed Squares estimator and the size of the initial subsample are both given by \texttt{psi.0}. Alternatively, a Least Trimmed Squares estimator with a particular breakdown point can be entered through the argument \texttt{beta.0}.

Value

- \texttt{forward.beta} Matrix of dimension \( n \times p \). Forward Search estimates of \( \beta \).
- \texttt{forward.sigma2.biased} Matrix of dimension \( n \times 1 \). Forward Search estimates of \( \sigma \). Values are *not* bias corrected.
- \texttt{forward.residual} Matrix of dimension \( n \times 1 \). Forward Search estimates of forward residuals. Values are *not* bias corrected.
- \texttt{m.0} Number of observations in initial set of selected observations.
- \texttt{y} Vector of dimension \( n \). Dependent variable from argument.
- \texttt{x} Matrix of dimension \( n \times \text{dim.x} \). Design matrix for regressors. Dependent variable from argument augmented with constant. First column is constant.

Author(s)

Bent Nielsen <bent.nielsen@nuffield.ox.ac.uk> 9 Sep 2014

References


Examples

```
# EXAMPLE 1
# using Fulton Fish data,
# see Johansen and Nielsen (2014).

# Call package
library(ForwardSearch)

# Call data
```
data(Fulton)
mdata <- as.matrix(Fulton)
n <- nrow(mdata)

# Identify variable to reproduce Johansen and Nielsen (2014)
q <- mdata[2:n,9]
q_1 <- mdata[1:(n-1),9]
s <- mdata[2:n,6]
x.q.s <- cbind(q_1,s)
colnames(x.q.s) <- c("q_1","stormy")

# Fit Forward Search
FS95 <- ForwardSearch.fit(x.q.s,q,psi=0.95)

---

**ForwardSearch.plot**  
*Plots forward residuals with simultaneous confidence bands*

**Description**

Plots forward residuals with simultaneous confidence bands based on Johansen and Nielsen (2013, 2014).

**Usage**

```r
ForwardSearch.plot(FS, ref.dist = "normal", bias.correct = FALSE, return = FALSE, plot.legend = TRUE,
col = NULL, legend = NULL, lty = NULL, lwd = NULL,
main = NULL, type = NULL, xlab = NULL, ylab = NULL)
```

**Arguments**

- **FS**  
  List. Value of the function `ForwardSearch.fit`.
- **ref.dist**  
  Character. Reference distribution.  
  "normal" standard normal distribution.
- **bias.correct**  
  Logical. If FALSE do not bias correct variance, so plots have appearance similar to Atkinson and Riani (2000). If TRUE do bias correct variance, so plots start at origin. Default is FALSE.
- **return**  
  Logical. Default is FALSE: do not return values.
- **plot.legend**  
  Logical. Default is TRUE: include legend in plot.
- **col**  
  plot parameter. Vector of 6 colours.
- **legend**  
  plot parameter. Vector of 6 characters.
- **lty**  
  plot parameter. Vector of 6 line types.
- **lwd**  
  plot parameter. Vector of 6 line widths.
- **main**  
  plot parameter. Character.
type plot parameter. Character for plot type.
xlab plot parameter. Character for x label.
ylab plot parameter. Character for y label.

Value
ref.dist Character. From argument.
bias.correct Logical. From argument.
forward.residual.scaled Vector. Forward residuals scaled by estimated variance. The estimated variance is or is not bias corrected depending on the choice of bias.correct.
forward.asymp.median Vector. Asymptotic median.
cut.off Matrix. Cut-offs taken from Table 3 of Johansen and Nielsen (2014).

Author(s)
Bent Nielsen <bent.nielsen@nuffield.ox.ac.uk> 9 Sep 2014

References

Examples

###

# EXAMPLE 1
# using Fulton Fish data,
# see Johansen and Nielsen (2014).

# Call package
library(ForwardSearch)

# Call data
data(Fulton)
mdata <- as.matrix(Fulton)
n <- nrow(mdata)

# Identify variable to reproduce Johansen and Nielsen (2014)
q <- mdata[2:n,9]
q.1 <- mdata[1:(n-1),9]
s <- mdata[2:n,6]
x.q.s <- cbind(q, q.1, s)
Functions for asymptotic theory of Forward Search

Description
Computes functions appearing in asymptotic theory of Forward Search based on Johansen and Nielsen (2013).

Usage
ForwardSearch.pointwise.asymptotics(psi, ref.dist = "normal")

Arguments
psi Number or vector. Takes value(s) in interval 0,1.
ref.dist Character. Reference distribution
"normal" Standard normal distribution

Details
The asymptotic theory is developed in Johansen and Nielsen (2013), see Section 2.2.

\( c \) and \( \psi \) are linked through \( P(|\epsilon| < c) = \psi \), where \( \epsilon \) is a random variable with the chosen reference distribution.
\( \zeta \) is a consistency factor. Its square is defined as the truncated second moment \( \tau = \int_{-c}^{c} x^2 f(x)dx \) divided by \( \psi \).
\( \varpi \) is the asymptotic standard deviation resulting from Theorem 3.3.

Value
varpi Number or vector. sdv for forward residuals normalized by variance estimator and multiplied by twice the reference density.
zeta Number or vector. Consistency correction factor.
sdv.unbiased Number or vector. varpi/2/f.
sdv.biased Number or vector. varpi/2/f/\psi.
c Number or vector. c (median in unbiased case).
median.biased Number or vector. median (in biased case).
**ForwardSearch.stopped**  

**Description**  

A Forward Search gives a sequence of regression estimators. This function gives the regression estimators when stopped at m.

**Usage**  

ForwardSearch.stopped(FS, m)
Arguments

FS List. Value of the function `ForwardSearch.fit`.
m Integer. Stopping time.

Value

ranks.selected Vector. Ranks of m observations in the selected set.
ranks.outliers Vector. Ranks of n-m observations that are not selected. These are the "outliers". It is the complement to ranks.selected.

beta.m Vector. Least squares estimator based on ranks.selected.
sigma2.biased Scalar. Least squares residual variance based on ranks.selected. Value is *not* bias corrected.

Author(s)

Bent Nielsen <bent.nielsen@nuffield.ox.ac.uk> 9 Sep 2014

References


Examples

```r
#################################
# EXAMPLE 1
# using Fulton Fish data,
# see Johansen and Nielsen (2014).

# Call package
library(ForwardSearch)

# Call data
data(Fulton)
mdata <- as.matrix(Fulton)
n <- nrow(mdata)

# Identify variable to reproduce Johansen and Nielsen (2014)
q <- mdata[2:n,9]
q_1 <- mdata[1:(n-1),9]
s <- mdata[2:n,6]
x.q.s <- cbind(q_1,s)
colnames(x.q.s) <- c("q_1","stormy")

# Fit Forward Search
FS95 <- ForwardSearch.fit(x.q.s,q,psi=0.95)
```
**Fulton fish data**

**Description**


**Usage**

data(Fulton)

**Format**

Matrix with 111 rows of daily data and 13 variables.

**Details**

Documentation on the Fulton Fish market and original data can be found in Graddy (1995, 2006). Documentation for aggregated data used here can be found in Angrist, Graddy and Imbens (2000). Data used as example in Hendry and Nielsen (2007). Downloaded from Econometric Modeling.

The data set comprises aggregated daily prices and quantities of whiting sold in the period 2 December 1991 to 8 May 1992. In particular it has the variables

- **Monday** 1 if Monday, 0 otherwise.
- **Tuesday** 1 if Tuesday, 0 otherwise.
- **Wednesday** 1 if Wednesday, 0 otherwise.
- **Thursday** 1 if Thursday, 0 otherwise.
- **Date**
  - **Stormy** 1 if wave height greater than 4.5 feet wind speed greater than 18 knots. Based on moving averages of the last three days’ wind speed and wave height before the trading day, as measured off the coast of Long Island and reported in the New York Times boating forecast.
  - **Mixed** 1 if wave height greater than 3.8 feet wind speed greater than 13 knots excluding stormy days. Based on moving averages of the last three days’ wind speed and wave height before the trading day, as measured off the coast of Long Island and reported in the New York Times boating forecast.
- **LogPrice** Prices are average prices in US dollars per pound.
- **LogQuantity** Quantities are pounds of whiting per day.
- **Rainy** 1 if rainy weather on shore.
- **Cold** 1 if cold weather on shore.
- **Windspeed**
- **Windspeed**

Square of windspeed.
Source


Examples

data(Fulton)
Index

*Topic datasets
  Fulton, 11
*Topic package
  ForwardSearch-package, 2

ForwardSearch (ForwardSearch-package), 2
ForwardSearch-package, 2
ForwardSearch.fit, 2, 4, 6, 10
ForwardSearch.plot, 2, 6
ForwardSearch.pointwise.asymptotics, 8
ForwardSearch.stopped, 3, 9
Fulton, 11