Package ‘lspartition’

October 13, 2022

Type Package

Title Nonparametric Estimation and Inference Procedures using Partitioning-Based Least Squares Regression

Version 0.4

Date 2019-08-08

Author Matias D. Cattaneo, Max H. Farrell, Yingjie Feng

Maintainer Yingjie Feng <yingjief@princeton.edu>


Depends R (>= 3.1)

License GPL-2

Encoding UTF-8

LazyData true

Imports ggplot2, pracma, mgcv, combinat, matrixStats, MASS, dplyr

RoxygenNote 6.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2019-08-08 22:40:06 UTC

R topics documented:

lspartition-package .................................................. 2
lspkselect .............................................................. 2
lsplincom ............................................................. 5
lsprobust .............................................................. 7
lsprobust.plot ........................................................ 11
Description

This package provides tools for statistical analysis using B-splines, wavelets, and piecewise polynomials as described in Cattaneo, Farrell and Feng (2019a): \texttt{lsprobust} for least squares point estimation with robust bias-corrected pointwise and uniform inference procedures; \texttt{lspkselect} for data-driven procedures for selecting the IMSE-optimal number of partitioning knots; \texttt{lsprobust.plot} for regression plots with robust confidence intervals and confidence bands; \texttt{lsplincom} for estimation and inference for linear combination of regression functions of different groups.

The companion software article, Cattaneo, Farrell and Feng (2019b), provides further implementation details and empirical illustrations.

Author(s)

Matias D. Cattaneo, Princeton University, Princeton, NJ. \texttt{<cattaneo@princeton.edu>}

Max H. Farrell, University of Chicago, Chicago, IL. \texttt{<max.farrell@chicagobooth.edu>}

Yingjie Feng (maintainer), Princeton University, Princeton, NJ. \texttt{<yingjief@princeton.edu>}

References


Description

\texttt{lspkselect} implements data-driven procedures to select the Integrated Mean Squared Error (IMSE) optimal number of partitioning knots for partitioning-based least squares regression estimators. Three series methods are supported: B-splines, compactly supported wavelets, and piecewise polynomials. See Cattaneo and Farrell (2013) and Cattaneo, Farrell and Feng (2019a) for complete details.

Companion commands: \texttt{lsprobust} for partitioning-based least squares regression estimation and inference; \texttt{lsprobust.plot} for plotting results; \texttt{lsplincom} for multiple sample estimation and inference.

A detailed introduction to this command is given in Cattaneo, Farrell and Feng (2019b).

For more details, and related Stata and R packages useful for empirical analysis, visit \texttt{https://sites.google.com/site/nppackages/}.
Usage

```r
lspkselect(y, x, m = NULL, m.bc = NULL, smooth = NULL, bsmooth = NULL, deriv = NULL, method = "bs", ktype = "uni", kselect = "imse-dpi", proj = TRUE, bc = "bc3", vce = "hc2", subset = NULL, rotnorm = TRUE)
```

## S3 method for class 'lspkselect'

print(x, ...)

## S3 method for class 'lspkselect'

summary(object, ...)

Arguments

- `y` Outcome variable.
- `x` Independent variable. A matrix or data frame.
- `m` Order of basis used in the main regression. Default is `m=2`.
- `m.bc` Order of basis used to estimate leading bias. Default is `m.bc=m+1`.
- `smooth` Smoothness of B-splines for point estimation. When `smooth=s`, B-splines have `s`-order continuous derivatives. Default is `smooth=m-2`.
- `bsmooth` Smoothness of B-splines for bias correction. Default is `bsmooth=m.bc-2`.
- `deriv` Derivative order of the regression function to be estimated. A vector object of the same length as `ncol(x)`. Default is `deriv=c(0,...,0)`.
- `method` Type of basis used for expansion. Options are "bs" for B-splines, "wav" for compactly supported wavelets (Cohen, Daubechies and Vial, 1993), and "pp" for piecewise polynomials. Default is `method="bs"`.
- `ktype` Knot placement. Options are "uni" for evenly spaced knots over the support of `x` and "qua" for quantile-spaced knots. Default is `ktype="uni"`.
- `kselect` Method for selecting the number of inner knots used by `lspkselect`. Options are "imse-rot" for a rule-of-thumb (ROT) implementation of IMSE-optimal number of knots, "imse-dpi" for second generation direct plug-in (DPI) implementation of IMSE-optimal number of knots, and "all" for both. Default is `kselect="imse-dpi"`.
- `proj` If TRUE, projection of leading approximation error onto the lower-order approximating space is included for bias correction (splines and piecewise polynomial only). Default is `proj=TRUE`.
- `bc` Bias correction method. Options are "bc1" for higher-order-basis bias correction, "bc2" for least squares bias correction, and "bc3" for plug-in bias correction. Defaults are "bc3" for splines and piecewise polynomials and "bc2" for wavelets.
- `vce` Procedure to compute the heteroskedasticity-consistent (HCK) variance-covariance matrix estimator with plug-in residuals. Options are
  - "hc0" for unweighted residuals (HC0).
  - "hc1" for HC1 weights.
• "hc2" for HC2 weights. Default.
  • "hc3" for HC3 weights.

subset  Optional rule specifying a subset of observations to be used.

rotnorm If TRUE, ROT selection is adjusted using normal densities.

... further arguments

object  class lspkselect objects.

Value

ks  A matrix may contain k.rot (IMSE-optimal number of knots for the main regression through ROT implementation), k.bias.rot (IMSE-optimal number of knots for bias correction through ROT implementation), k.dpi (IMSE-optimal number of knots for the main regression through DPI implementation), k.bias.dpi (IMSE-optimal number of knots for bias correction through DPI implementation)

opt  A list containing options passed to the function.

Methods (by generic)

• print: print method for class "lspkselect".
• summary: summary method for class "lspkselect".

Author(s)

Matias D. Cattaneo, Princeton University, Princeton, NJ. <cattaneo@princeton.edu>.

Max H. Farrell, University of Chicago, Chicago, IL. <max.farrell@chicagobooth.edu>.

Yingjie Feng (maintainer), Princeton University, Princeton, NJ. <yingjief@princeton.edu>.

References


See Also

lsprobust, lsrobust.plot, lsplincom
**Examples**

```r
x <- data.frame(runif(500), runif(500))
y <- sin(4*x[,1])+cos(x[,2])+rnorm(500)
est <- lsplkselect(y, x)
summary(est)
```

---

**Description**

lsplincom implements user-specified linear combinations across different data sub-groups for regression functions estimation, and computes corresponding (pointwise and uniform) robust bias-corrected inference measures. Estimation and inference is implemented using the *lspartition* package. See Cattaneo and Farrell (2013) and Cattaneo, Farrell and Feng (2019a) for complete details. A detailed introduction to this command is given in Cattaneo, Farrell and Feng (2019b).

For more details, and related Stata and R packages useful for empirical analysis, visit [https://sites.google.com/site/nppackages/](https://sites.google.com/site/nppackages/).

**Usage**

```r
lsplincom(y, x, G, R, eval = NULL, neval = NULL, level = 95, 
band = FALSE, cb.method = NULL, cb.grid = NULL, cb.ngrid = 50, 
B = 1000, subset = NULL, knot = NULL, ...)
```

```r
## S3 method for class 'lsplincom'
print(x, ...)
```

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

**Arguments**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>Outcome variable.</td>
</tr>
<tr>
<td>x</td>
<td>Independent variable. A matrix or data frame.</td>
</tr>
<tr>
<td>G</td>
<td>Group indicator. It may take on multiple discrete values.</td>
</tr>
<tr>
<td>R</td>
<td>A numeric vector giving the linear combination of interest. Each element is the coefficient of the conditional mean estimator of one group, and they are ordered ascendingly along the value of G.</td>
</tr>
<tr>
<td>eval</td>
<td>Evaluation points. A matrix or data frame.</td>
</tr>
<tr>
<td>neval</td>
<td>Number of quantile-spaced evaluating points.</td>
</tr>
<tr>
<td>level</td>
<td>Confidence level used for confidence intervals; default is level=95.</td>
</tr>
<tr>
<td>band</td>
<td>If TRUE, the critical value for constructing confidence band is calculated. Default is band=FALSE.</td>
</tr>
</tbody>
</table>
cb.method  Method used to calculate the critical value for confidence bands. Options are "pl" for a simulation-based plug-in procedure, and "wb" for a wild bootstrap procedure. If band=TRUE with cb.method unspecified, default is cb.method="pl".

cb.grid  A matrix containing all grid points used to construct confidence bands. Each row corresponds to the coordinates of one grid point.

cb.ngrid  A numeric vector of the same length as ncol(x). Each element corresponds to the number of grid points for each dimension used to implement uniform inference. Default is uni.ngrid=50.

B  Number of simulated samples used to obtain the critical value for confidence bands. Default is B=1000.

subset  Optional rule specifying a subset of observations to be used.

knot  A list of numeric vectors giving the knot positions (including boundary knots) for each dimension which are used in the main regression. The length of the list is equal to ncol(x). If not specified, it uses the number of knots either specified by users or computed by the companion command lspkselect to generate the corresponding knots according to the rule specified by ktype. See help for lsprobust.

...  Arguments to be passed to the function. See lsprobust.

object  class lsplincom objects.

Value

Estimate  A matrix containing eval (grid points), N (effective sample sizes), tau.cl (point estimates with a basis of order m), tau.bc (bias corrected point estimates with a basis of order m.bc), se.cl (standard error corresponding to tau.cl), and se.rb (robust standard error).

sup.cval  Critical value for constructing confidence bands.

opt  A list containing options passed to the function.

Methods (by generic)

- print: print method for class "lsplincom".
- summary: summary method for class "lsplincom"

Author(s)

Matias D. Cattaneo, Princeton University, Princeton, NJ. <cattaneo@princeton.edu>.
Max H. Farrell, University of Chicago, Chicago, IL. <max.farrell@chicagobooth.edu>.
Yingjie Feng (maintainer), Princeton University, Princeton, NJ. <yingjief@princeton.edu>.

References


See Also

`lsprobust`, `lspkselect`, `lsprobust.plot`.

Examples

```r
x <- runif(500)
y <- sin(4*x)+rnorm(500)
z <- c(rep(0, 250), rep(1, 250))
est <- lsplincom(y, x, z, c(-1, 1))
summary(est)
```

Description

`lsprobust` implements partitioning-based least squares point estimators for the regression function and its derivatives. It also provides robust bias-corrected (pointwise and uniform) inference, including simulation-based confidence bands. Three series methods are supported: B-splines, compact supported wavelets, and piecewise polynomials. See Cattaneo and Farrell (2013) and Cattaneo, Farrell and Feng (2019a) for complete details.

Companion commands: `lspkselect` for data-driven IMSE-optimal selection of the number of knots on rectangular partitions; `lsprobust.plot` for plotting results; `lsplincom` for multiple sample estimation and inference.

A detailed introduction to this command is given in Cattaneo, Farrell and Feng (2019b).

For more details, and related Stata and R packages useful for empirical analysis, visit https://sites.google.com/site/nppackages/.

Usage

```r
lsprobust(y, x, eval = NULL, neval = NULL, method = "bs", m = NULL,
m.bc = NULL, deriv = NULL, smooth = NULL, bsmooth = NULL,
kttype = "uni", knot = NULL, nknot = NULL, same = TRUE,
bknot = NULL, bnknot = NULL, J = NULL, bc = "bc3", proj = TRUE,
kselect = "imse-dpi", vce = "hc2", level = 95, uni.method = NULL,
uni.grid = NULL, uni.ngrid = 50, uni.out = FALSE, band = FALSE,
B = 1000, subset = NULL, rotnorm = TRUE)
```

```r
## S3 method for class 'lsprobust'
print(x, ...)
```

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

\begin{itemize}
\item \texttt{y} \quad Outcome variable.
\item \texttt{x} \quad Independent variable. A matrix or data frame.
\item \texttt{eval} \quad Evaluation points. A matrix or data frame.
\item \texttt{neval} \quad Number of quantile-spaced evaluating points.
\item \texttt{method} \quad Type of basis used for expansion. Options are "bs" for B-splines, "wav" for compactly supported wavelets (Cohen, Daubechies and Vial, 1993), and "pp" for piecewise polynomials. Default is \texttt{method="bs"}.
\item \texttt{m} \quad Order of basis used in the main regression. Default is \texttt{m=2}. For B-splines, if \texttt{smooth} is specified but \texttt{m} is unspecified, default is \texttt{m=smooth+2}.
\item \texttt{m.bc} \quad Order of basis used to estimate leading bias. Default is \texttt{m.bc=m+1}. For B-splines, if \texttt{bssmooth} is specified but \texttt{m.bc} is unspecified, default is \texttt{m.bc=bssmooth+2}.
\item \texttt{deriv} \quad Derivative order of the regression function to be estimated. A vector object of the same length as \texttt{ncol(x)}. Default is \texttt{deriv=c(0,...,0)}.
\item \texttt{smooth} \quad Smoothness of B-splines for point estimation. When \texttt{smooth=s}, B-splines have \texttt{s}-order continuous derivatives. Default is \texttt{smooth=m-2}.
\item \texttt{bssmooth} \quad Smoothness of B-splines for bias correction. Default is \texttt{bssmooth=m.bc-2}.
\item \texttt{ktype} \quad Knot placement. Options are "uni" for evenly-spaced knots over the support of \texttt{x} and "qua" for quantile-spaced knots. Default is \texttt{ktype="uni"}.
\item \texttt{knot} \quad A list of numeric vectors giving the knot positions (including boundary knots) for each dimension which are used in the main regression. The length of the list is equal to \texttt{ncol(x)}. If not specified, it uses the number of knots either specified by users or computed by the companion command \texttt{lspkselect} to generate the corresponding knots according to the rule specified by \texttt{ktype}.
\item \texttt{nknot} \quad A numeric vector of the same length as \texttt{ncol(x)}. Each element corresponds to the number of \texttt{inner} partitioning knots for each dimension used in the main regression. If not specified, \texttt{nknot} is computed by the companion command \texttt{lspkselect}.
\item \texttt{same} \quad If \texttt{TRUE}, the same knots are used for bias correction as that for the main regression. Default is \texttt{same=TRUE}.
\item \texttt{bknot} \quad A list of numeric vectors giving knot positions used for bias correction. If not specified and \texttt{same=FALSE}, it uses the number of knots either specified by users or computed by the companion command \texttt{lspkselect} to generate knots according to the rule specified by \texttt{ktype}.
\item \texttt{bnknot} \quad A numeric vector of the same length as \texttt{ncol(x)}. Each element corresponds to the number of \texttt{inner} partitioning knots for each dimension used for bias correction. If not specified, \texttt{bnknot} is computed by the companion command \texttt{lspkselect}.
\item \texttt{J} \quad A numeric vector containing resolution levels of father wavelets for each dimension.
\item \texttt{bc} \quad Bias correction method. Options are "bc1" for higher-order-basis bias correction, "bc2" for least squares bias correction, and "bc3" for plug-in bias correction. Default are "bc3" for splines and piecewise polynomials and "bc2" for wavelets.
\end{itemize}
proj

If TRUE, projection of leading approximation error onto the lower-order approximation space is included for bias correction (splines and piecewise polynomials only). Default is proj=TRUE.

kselect

Method for selecting the number of inner knots used by lspkselect. Options are "imse-rot" for ROT implementation of IMSE-optimal number of knots and "imse-dpi" for second generation of DPI implementation of IMSE-optimal number of knots. Default is kselect="imse-dpi".

vce

Procedure to compute the heteroskedasticity-consistent (HCk) variance-covariance matrix estimator with plug-in residuals. Options are

- "hc0" for unweighted residuals (HC0).
- "hc1" for HC1 weights.
- "hc2" for HC2 weights. Default.
- "hc3" for HC3 weights.

level

Confidence level used for confidence intervals; default is level=95.

uni.method

Method used to implement uniform inference. Options are "pl" for a simulation-based plug-in procedure, "wb" for a wild bootstrap procedure. If unspecified, neither procedure is implemented. Default is uni.method=NULL.

uni.grid

A matrix containing all grid points used to implement uniform inference. Each row corresponds to the coordinates of one grid point.

uni.ngrid

A numeric vector of the same length as ncol(x). Each element corresponds to the number of grid points for each dimension used to implement uniform inference. Default is uni.ngrid=50.

uni.out

If TRUE, the quantities used to implement uniform inference is outputted. Default is uni.out=FALSE.

band

If TRUE, the critical value for constructing confidence band is calculated. Default is band=FALSE. If band=TRUE with uni.method unspecified, default is uni.method="pl".

B

Number of simulated samples used to obtain the critical value for confidence bands. Default is B=1000.

subset

Optional rule specifying a subset of observations to be used.

rotnorm

If TRUE, ROT selection is adjusted using normal densities.

... further arguments

object
class lsprobust objects.

Value

Estimate

A matrix containing eval (grid points), N (effective sample sizes), tau.cl (point estimates with a basis of order m), tau.bc (bias corrected point estimates with a basis of order m.bc), se.cl (standard error corresponding to tau.cl), and se.rb (robust standard error).

k.num

A matrix containing the number of inner partitioning knots used in the main regression and bias correction for each covariate.

knot

A list of knots for point estimation.
bknob: A list of knots for bias correction.
sup.cval: Critical value for constructing confidence band.
uni.output: A list containing quantities used to implement uniform inference.
opt: A list containing options passed to the function.

Methods (by generic)

- print: print method for class "lsprobust"
- summary: summary method for class "lsprobust"

Author(s)

Matias D. Cattaneo, Princeton University, Princeton, NJ. <cattaneo@princeton.edu>.
Max H. Farrell, University of Chicago, Chicago, IL. <max.farrell@chicagobooth.edu>.
Yingjie Feng (maintainer), Princeton University, Princeton, NJ. <yingjief@princeton.edu>.

References


See Also

lspkselect, lsprobust.plot, lsplincom

Examples

```r
data.frame(runif(500), runif(500))
y <- sin(4*x[,1])+cos(x[,2])+rnorm(500)
est <- lsprobust(y, x)
summary(est)
```
lsprobust.plot plots estimated regression functions and confidence regions using the \texttt{lspartition} package. See Cattaneo and Farrell (2013) and Cattaneo, Farrell and Feng (2019a) for complete details.

Companion command: \texttt{lsprobust} for partitioning-based least squares regression estimation and inference; \texttt{lsprobust.plot} for plotting results; \texttt{lsplincom} for multiple sample estimation and inference.

A detailed introduction to this command is given in Cattaneo, Farrell and Feng (2019b).

For more details, and related Stata and R packages useful for empirical analysis, visit \url{https://sites.google.com/site/nppackages/}.

**Usage**

\begin{verbatim}
lsprobust.plot(..., alpha = NULL, type = NULL, CS = "ci", CStype = NULL, title = ", xlabel = "", ylabel = "", lty = NULL, lwd = NULL, lcol = "", pty = NULL, pwd = NULL, pcol = NULL, CSshade = NULL, CScolor = NULL, legendTitle = NULL, legendGroups = NULL)
\end{verbatim}

**Arguments**

\begin{itemize}
  \item \texttt{...} Objects returned by \texttt{lsprobust}.
  \item \texttt{alpha} Numeric scalar between 0 and 1, the significance level for plotting confidence regions. If more than one is provided, they will be applied to data series accordingly.
  \item \texttt{type} String, one of "line" (default), "points", "binscatter", "none" or "both", how the point estimates are plotted. If more than one is provided, they will be applied to data series accordingly.
  \item \texttt{CS} String, type of confidence sets. Options are "ci" for pointwise confidence intervals, "cb" for uniform confidence bands, and "all" for both.
  \item \texttt{CStype} String, one of "region" (shaded region, default), "line" (dashed lines), "ebar" (error bars), "all" (all of the previous) or "none" (no confidence region), how the confidence region should be plotted. If more than one is provided, they will be applied to data series accordingly. If CS = "all", pointwise confidence intervals are forced to be represented by error bars, and uniform bands are represented by both lines and regions.
  \item \texttt{title} String, title of the plot.
  \item \texttt{xlabel} Strings, labels for x-axis.
  \item \texttt{ylabel} Strings, labels for y-axis.
\end{itemize}
Line type for point estimates, only effective if type is "line" or "both". 1 for solid line, 2 for dashed line, 3 for dotted line. For other options, see the instructions for ggplot2 or par. If more than one is provided, they will be applied to data series accordingly.

Line width for point estimates, only effective if type is "line" or "both". Should be strictly positive. For other options, see the instructions for ggplot2 or par. If more than one is provided, they will be applied to data series accordingly.

Line color for point estimates, only effective if type is "line" or "both". 1 for black, 2 for red, 3 for green, 4 for blue. For other options, see the instructions for ggplot2 or par. If more than one is provided, they will be applied to data series accordingly.

Scatter plot type for point estimates, only effective if type is "points" or "both". For options, see the instructions for ggplot2 or par. If more than one is provided, they will be applied to data series accordingly.

Scatter plot size for point estimates, only effective if type is "points" or "both". Should be strictly positive. If more than one is provided, they will be applied to data series accordingly.

Scatter plot color for point estimates, only effective if type is "points" or "both". 1 for black, 2 for red, 3 for green, 4 for blue. For other options, see the instructions for ggplot2 or par. If more than one is provided, they will be applied to data series accordingly.

Numeric, opaqueness of the confidence region, should be between 0 (transparent) and 1. Default is 0.2. If more than one is provided, they will be applied to data series accordingly.

Color for confidence region. 1 for black, 2 for red, 3 for green, 4 for blue. For other options, see the instructions for ggplot2 or par. If more than one is provided, they will be applied to data series accordingly.

String, title of legend.

String vector, group names used in legend.

Companion command: lsprobust for partition-based least-squares regression estimation.

A standard ggplot2 object is returned, hence can be used for further customization.

Matias D. Cattaneo, Princeton University, Princeton, NJ. <cattaneo@princeton.edu>.
Max H. Farrell, University of Chicago, Chicago, IL. <max.farrell@chicagobooth.edu>.
Yingjie Feng (maintainer), Princeton University, Princeton, NJ. <yingjief@princeton.edu>.
References

See Also
lsprobust, lspkselect, lsplincom, ggplot2.

Examples
x <- runif(500)
y <- sin(4*x)+rnorm(500)
est <- lsprobust(y, x)
lsprobust.plot(est)
Index

ggplot2, 12, 13
lspartition-package, 2
lspkselect, 2, 2, 7, 10, 13
lsplincom, 2, 4, 5, 7, 10, 11, 13
lsprobust, 2, 4, 6, 7, 7, 11–13
lsprobust.plot, 2, 4, 7, 10, 11, 11

par, 12
print.lspkselect (lspkselect), 2
print.lsplincom (lsplincom), 5
print.lsprobust (lsprobust), 7

summary.lspkselect (lspkselect), 2
summary.lsplincom (lsplincom), 5
summary.lsprobust (lsprobust), 7