Package ‘nprobust’

September 15, 2017

Type Package
Title Nonparametric Robust Estimation and Inference Methods using Local Polynomial Regression and Kernel Density Estimation
Version 0.1.1
Date 2017-09-14
Author Sebastian Calonico <scalonico@bus.miami.edu>, Matias D. Cattaneo <cattaneo@umich.edu>, Max H. Farrell <max.farrell@chicagobooth.edu>
Maintainer Sebastian Calonico <scalonico@bus.miami.edu>
Description Tools for data-driven statistical analysis using local polynomial regression and kernel density estimation methods as described in Calonico, Cattaneo and Farrell (2017a): lprobust() for local polynomial point estimation and robust bias-corrected inference and kdrobust() for kernel density point estimation and robust bias-corrected inference. Several optimal bandwidth selection procedures are computed by lpbwselect() and kdbwselect() for local polynomial and kernel density estimation, respectively. Finally, nprobust.plot() for density and regression plots with robust confidence interval.
License GPL-2
Imports Rcpp, ggplot2
LinkingTo Rcpp, RcppArmadillo
NeedsCompilation yes
Repository CRAN
Date/Publication 2017-09-14 22:26:18 UTC

R topics documented:
nprobust-package .................................................. 2
kdbwselect ......................................................... 2
kdrobust ........................................................... 4
lpbwselect .......................................................... 6
lprobust ............................................................ 9
nprobust.plot ....................................................... 11

Index 14
**nprobust-package**  
Nonparametric Robust Estimation and Inference Methods using Local Polynomial Regression and Kernel Density Estimation

**Description**

This package provides tools for data-driven statistical analysis using local polynomial regression (LPR) and kernel density estimation (KDE) methods:  
- `lprobust` for local polynomial point estimation and robust bias-corrected inference,  
- `lpbwselect` for local polynomial bandwidth selection,  
- `kdrobust` for kernel density point estimation and robust bias-corrected inference,  
- `kdbwselect` for kernel density bandwidth selection, and  
- `nprobust.plot` for plotting results.

**Details**

Package: nprobust  
Type: Package  
Version: 0.1.1  
Date: 2017-09-14  
License: GPL-2

Function for LPR estimation and inference: `lprobust`  
Function for LPR bandwidth selection: `lpbwselect`  
Function for KDE estimation and inference: `kdrobust`  
Function for KDE bandwidth selection: `kdbwselect`  
Function for graphical analysis: `nprobust.plot`

**Author(s)**

Sebastian Calonico, University of Miami, Coral Gables, FL. <scalonico@bus.miami.edu>.  
Matias D. Cattaneo, University of Michigan, Ann Arbor, MI. <cattaneo@umich.edu>.  
Max H. Farrell, University of Chicago, Chicago, IL. <max.farrell@chicagobooth.edu>.

**kdbwselect**  
Bandwidth Selection Procedures for Kernel Density Estimation and Inference

**Description**

`kdbwselect` implements bandwidth selectors for kernel density point estimators and inference procedures developed in Calonico, Cattaneo and Farrell (2017a). It also implements other bandwidth selectors available in the literature. See Wand and Jones (1995) for background references.
Companion commands are: `kdrobust` for kernel density point estimation and inference procedures. A detailed introduction to this command is given in Calonico, Cattaneo and Farrell (2017b). For more details, and related Stata and R packages useful for empirical analysis, visit https://sites.google.com/site/nppackages/.

**Usage**

```plaintext
dbwselect(x, eval = NULL, neval = NULL, rho = NULL, kernel = "epa",
bwselect = "mse-dpi", bwcheck= NULL, subset = NULL)
```

**Arguments**

- `x`: independent variable.
- `eval`: vector of evaluation point(s). By default it uses 30 equally spaced points over to support of `x`.
- `neval`: number of quantile-spaced evaluation points on support of `x`. Default is `neval=30`.
- `rho`: `rho=h/b`. Default is `rho = 1` if `h` is specified but `b` is not.
- `kernel`: kernel function used to construct local polynomial estimators. Options are `epa` for the epanechnikov kernel, `tri` for the triangular kernel and `uni` for the uniform kernel. Default is `kernel = epa`.
- `bwselect`: bandwidth selection procedure to be used via `dbwselect`. By default it computes both `h` and `b`, unless `rho` is specified, in which case it only computes `h` and sets `b=h/rho`. Options are:
  - `mse-dpi` second-generation DPI implementation of MSE-optimal bandwidth. Default option.
  - `imse-dpi` second-generation DPI implementation of IMSE-optimal bandwidth (computed using grid of evaluation points selected).
  - `imse-rot` ROT implementation of IMSE-optimal bandwidth (computed using grid of evaluation points selected).
  - `ce-dpi` second generation DPI implementation of CE-optimal bandwidth.
  - `ce-rot` ROT implementation of CE-optimal bandwidth.
  - `all`: reports all available bandwidth selection procedures.
- `bwcheck`: if a positive integer is provided, then the selected bandwidth is enlarged so that at least `bwcheck` effective observations are available at each evaluation point. Default is `bwcheck = NULL`.
- `subset`: optional rule specifying a subset of observations to be used.

**Value**

- **Estimate**: A matrix containing `eval` (grid points), `h` and `b` (bandwidths).
- **opt**: A list containing options passed to the function.
Author(s)
Sebastian Calonico, University of Miami, Coral Gables, FL. <scalonico@bus.miami.edu>.
Matias D. Cattaneo, University of Michigan, Ann Arbor, MI. <cattaneo@umich.edu>.
Max H. Farrell, University of Chicago, Chicago, IL. <max.farrell@chicagobooth.edu>.

References

See Also
kdrobust

Examples
```r
x <- rnorm(500)
est <- kdbwselect(x)
summary(est)
```

Description
kdrobust implements kernel density point estimators, with robust bias-corrected confidence intervals and inference procedures developed in Calonico, Cattaneo and Farrell (2017a). It also implements other estimation and inference procedures available in the literature. See Wand and Jones (1995) for background references.

Companion commands: kdbwselect for kernel density data-driven bandwidth selection, and nprobust.plot for plotting results.

A detailed introduction to this command is given in Calonico, Cattaneo and Farrell (2017b). For more details, and related Stata and R packages useful for empirical analysis, visit https://sites.google.com/site/nppackages/.

Usage
```r
kdrobust(x, eval = NULL, neval = NULL, h = NULL, b = NULL, rho = NULL,
kernel = "epa", bwselect = "mse-dpi", bwcheck = NULL, level = 95, subset = NULL)
```
Arguments

- **x**: independent variable.
- **eval**: vector of evaluation point(s). By default it uses 30 equally spaced points over to support of x.
- **neval**: number of quantile-spaced evaluation points on support of x. Default is neval=30.
- **h**: main bandwidth used to construct the kernel density point estimator. Can be either scalar (same bandwidth for all evaluation points), or vector of same dimension as eval. If not specified, bandwidth h is computed by the companion command `kdbwselect`.
- **b**: bias bandwidth used to construct the bias-correction estimator. Can be either scalar (same bandwidth for all evaluation points), or vector of same dimension as eval. If not specified, b is computed by the companion command `kdbwselect`.
- **rho**: \( \rho = \frac{h}{b} \). Default is \( \rho = 1 \) if h is specified but b is not.
- **kernel**: kernel function used to construct local polynomial estimators. Options are `epa` for the epanechnikov kernel, `tri` for the triangular kernel and `uni` for the uniform kernel. Default is kernel = `epa`.
- **bwselect**: bandwidth selection procedure to be used via `kdbwselect`. By default it computes both h and b, unless \( \rho \) is specified, in which case it only computes h and sets \( b = \frac{h}{\rho} \). Options are:
  - `mse-dpi`: second-generation DPI implementation of MSE-optimal bandwidth. Default option.
  - `imse-dpi`: second-generation DPI implementation of IMSE-optimal bandwidth (computed using grid of evaluation points selected).
  - `imse-rot`: ROT implementation of IMSE-optimal bandwidth (computed using grid of evaluation points selected).
  - `ce-dpi`: second generation DPI implementation of CE-optimal bandwidth.
  - `ce-rot`: ROT implementation of CE-optimal bandwidth.
  - `all`: reports all available bandwidth selection procedures.

Note: MSE = Mean Square Error; IMSE = Integrated Mean Squared Error; CE = Coverage Error; DPI = Direct Plug-in; ROT = Rule-of-Thumb. For details on implementation see Calonico, Cattaneo and Farrell (2017b).

- **bwcheck**: if a positive integer is provided, then the selected bandwidth is enlarged so that at least bwcheck effective observations are available at each evaluation point. Default is bwcheck = NULL.
- **level**: confidence level used for confidence intervals; default is level = 95.
- **subset**: optional rule specifying a subset of observations to be used.

Value

- **Estimate**: A matrix containing eval (grid points), h, b (bandwidths), N (effective sample sizes), f.us (point estimates with p-th order kernel function), f.bc (bias corrected point estimates), se.us (standard error corresponding to f.us), and se.rb (robust standard error).
- **opt**: A list containing options passed to the function.
Author(s)

Sebastian Calonico, University of Miami, Coral Gables, FL. <scalonico@bus.miami.edu>.
Matias D. Cattaneo, University of Michigan, Ann Arbor, MI. <cattaneo@umich.edu>.
Max H. Farrell, University of Chicago, Chicago, IL. <max.farrell@chicagobooth.edu>.

References


See Also

kdbwselect

Examples

```r
x <- rnorm(500)
est <- kdrobust(x)
summary(est)
```

lpbwselect

Bandwidth Selection Procedures for Local Polynomial Regression Estimation and Inference

Description

lpbwselect implements bandwidth selectors for local polynomial regression point estimators and inference procedures developed in Calonico, Cattaneo and Farrell (2017a). It also implements other bandwidth selectors available in the literature. See Wand and Jones (1995) and Fan and Gijbels (1996) for background references.

Companion commands: lprobust for local polynomial point estimation and inference procedures. A detailed introduction to this command is given in Calonico, Cattaneo and Farrell (2017b). For more details, and related Stata and R packages useful for empirical analysis, visit https://sites.google.com/site/nppackages/.

Usage

```r
lpbwselect(y, x, eval = NULL, neval = NULL, p = NULL, deriv = NULL, rho = NULL, kernel = "epa", bwselect = "mse-dpi", bwcheck = NULL, bwregul = 1, imsegrid = 30, vce = "nn", nntmatch = 3, interior = FALSE, subset = NULL)
```
**lpbwselect**

**Arguments**

- **y** dependent variable.
- **x** independent variable.
- **eval** vector of evaluation point(s). By default it uses 30 equally spaced points over to support of x.
- **neval** number of quantile-spaced evaluation points on support of x. Default is neval=30.
- **p** polynomial order used to construct point estimator; default is p = 1 (local linear regression).
- **deriv** derivative order of the regression function to be estimated. Default is deriv=0 (regression function).
- **rho** rho = h/b. Default is rho = 1 if h is specified but b is not.
- **kernel** kernel function used to construct local polynomial estimators. Options are epa for the epanechnikov kernel, tri for the triangular kernel and uni for the uniform kernel. Default is kernel = epa.
- **bwselect** bandwidth selection procedure to be used via lpbwselect. By default it computes both h and b, unless rho is specified, in which case it only computes h and sets b=h/rho. Options are:
  - mse-dpi second-generation DPI implementation of MSE-optimal bandwidth. Default option.
  - mse-rot ROT implementation of MSE-optimal bandwidth.
  - imse-dpi second-generation DPI implementation of IMSE-optimal bandwidth (computed using grid of evaluation points selected).
  - imse-rot ROT implementation of IMSE-optimal bandwidth (computed using grid of evaluation points selected).
  - ce-dpi second generation DPI implementation of CE-optimal bandwidth.
  - ce-rot ROT implementation of CE-optimal bandwidth.
  - all reports all available bandwidth selection procedures.
- **bwcheck** if a positive integer is provided, then the selected bandwidth is enlarged so that at least bwcheck effective observations are available at each evaluation point. Default is bwcheck = NULL.
- **bwregul** specifies scaling factor for the regularization term added to the denominator of bandwidth selectors. Setting bwregul = 0 removes the regularization term from the bandwidth selectors. Default is bwregul = 1.
- **imsegrid** number of evaluations points used to compute the IMSE bandwidth selector. Default is imsegrid = 30.
- **vce** procedure used to compute the variance-covariance matrix estimator. Options are:
  - nn heteroskedasticity-robust nearest neighbor variance estimator with nnmatch the (minimum) number of neighbors to be used. Default choice.
  - hc0 heteroskedasticity-robust plug-in residuals variance estimator without weights.

Note: MSE = Mean Square Error; IMSE = Integrated Mean Squared Error; CE = Coverage Error; DPI = Direct Plug-in; ROT = Rule-of-Thumb. For details on implementation see Calonico, Cattaneo and Farrell (2017b).
hc1 heteroskedasticity-robust plug-in residuals variance estimator with hc1 weights.
hc2 heteroskedasticity-robust plug-in residuals variance estimator with hc2 weights.
hc3 heteroskedasticity-robust plug-in residuals variance estimator with hc3 weights.

nnmatch to be combined with for vce=nn for heteroskedasticity-robust nearest neighbor variance estimator with nnmatch indicating the minimum number of neighbors to be used. Default is nnmatch=3.

interior if TRUE, all evaluation points are assumed to be interior points. This option affects only data-driven bandwidth selection via lpbwselect. Default is interior = FALSE.

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

Value

Estimate A matrix containing grid (grid points), h and b (bandwidths), N (sample size)

opt A list containing options passed to the function.

Author(s)

Sebastian Calonico, University of Miami, Coral Gables, FL. <scalonico@bus.miami.edu>.
Matias D. Cattaneo, University of Michigan, Ann Arbor, MI. <cattaneo@umich.edu>.
Max H. Farrell, University of Chicago, Chicago, IL. <max.farrell@chicagobooth.edu>.

References


See Also

lprobust

Examples

```r
x <- runif(500)
y <- sin(4*x) + rnorm(500)
est <- lpbwselect(y,x)
summary(est)
```
**Description**

`lprobust` implements local polynomial regression point estimators, with robust bias-corrected confidence intervals and inference procedures developed in Calonico, Cattaneo and Farrell (2017a). It also implements other estimation and inference procedures available in the literature. See Wand and Jones (1995) and Fan and Gijbels (1996) for background references.

Companion commands: `lpbwselect` for local polynomial data-driven bandwidth selection, and `nprobust.plot` for plotting results.

A detailed introduction to this command is given in Calonico, Cattaneo and Farrell (2017b). 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**

```
lprobust(y, x, eval = NULL, neval = NULL, p = NULL, deriv = NULL, h = NULL, b = NULL, rho = NULL, kernel = "epa", bwselect = "mse-dpi", bwcheck = NULL, bwregul = 1, imseggrid = 30, vce = "nn", nnmatch = 3, level = 95, interior = FALSE, subset = NULL)
```

**Arguments**

- `y` dependent variable.
- `x` independent variable.
- `eval` vector of evaluation point(s). By default it uses 30 equally spaced points over to support of `x`.
- `neval` number of quantile-spaced evaluation points on support of `x`. Default is `neval=30`.
- `p` polynomial order used to construct point estimator; default is `p = 1` (local linear regression).
- `deriv` derivative order of the regression function to be estimated. Default is `deriv=0` (regression function).
- `h` main bandwidth used to construct local polynomial point estimator. Can be either scalar (same bandwidth for all evaluation points), or vector of same dimension as `eval`. If not specified, bandwidth `h` is computed by the companion command `lpbwselect`.
- `b` bias bandwidth used to construct the bias-correction estimator. Can be either scalar (same bandwidth for all evaluation points), or vector of same dimension as `eval`. If not specified, `b` is computed by the companion command `lpbwselect`.
- `rho` `rho=h/b`. Default is `rho = 1` if `h` is specified but `b` is not.
- `kernel` kernel function used to construct local polynomial estimators. Options are `epa` for the epanechnikov kernel, `tri` for the triangular kernel and `uni` for the uniform kernel. Default is `kernel = epa`.
bwselect bandwidth selection procedure to be used via `lpbwselect`. By default it computes both h and b, unless rho is specified, in which case it only computes h and sets b=h/rho. Options are:

- **mse-dpi** second-generation DPI implementation of MSE-optimal bandwidth. Default option.
- **mse-rot** ROT implementation of MSE-optimal bandwidth.
- **imse-dpi** second-generation DPI implementation of IMSE-optimal bandwidth (computed using grid of evaluation points selected).
- **imse-rot** ROT implementation of IMSE-optimal bandwidth (computed using grid of evaluation points selected).
- **ce-dpi** second generation DPI implementation of CE-optimal bandwidth.
- **ce-rot** ROT implementation of CE-optimal bandwidth.
- **all** reports all available bandwidth selection procedures.

Note: MSE = Mean Square Error; IMSE = Integrated Mean Squared Error; CE = Coverage Error; DPI = Direct Plug-in; ROT = Rule-of-Thumb. For details on implementation see Calonico, Cattaneo and Farrell (2017b).

bwcheck if a positive integer is provided, then the selected bandwidth is enlarged so that at least bwcheck effective observations are available at each evaluation point. Default is `bwcheck = NULL`.

bwregul specifies scaling factor for the regularization term added to the denominator of bandwidth selectors. Setting `bwregul = 0` removes the regularization term from the bandwidth selectors. Default is `bwregul = 1`.

imsegrid number of evaluations points used to compute the IMSE bandwidth selector. Default is `imsegrid = S0`.

vce procedure used to compute the variance-covariance matrix estimator. Options are:

- **nn** heteroskedasticity-robust nearest neighbor variance estimator with `nnmatch` the (minimum) number of neighbors to be used. Default choice.
- **hc0** heteroskedasticity-robust plug-in residuals variance estimator without weights.
- **hc1** heteroskedasticity-robust plug-in residuals variance estimator with `hc1` weights.
- **hc2** heteroskedasticity-robust plug-in residuals variance estimator with `hc2` weights.
- **hc3** heteroskedasticity-robust plug-in residuals variance estimator with `hc3` weights.

nnmatch to be combined with for vce=nn for heteroskedasticity-robust nearest neighbor variance estimator with `nnmatch` indicating the minimum number of neighbors to be used. Default is `nnmatch=3`.

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

interior if `TRUE`, all evaluation points are assumed to be interior points. This option affects only data-driven bandwidth selection via `lpbwselect`. Default is `interior = FALSE`.

subset optional rule specifying a subset of observations to be used.
**nprobust.plot**

**Value**

**Estimate**
A matrix containing `eval` (grid points), `h`, `b` (bandwidths), `N` (effective sample sizes), `m.us` (point estimates with `p`-th order local polynomial), `m.bc` (bias corrected point estimates with `(p+1)`-th order local polynomial), `se.us` (standard error corresponding to `m.us`), and `se.rb` (robust standard error).

**opt**
A list containing options passed to the function.

**Author(s)**
Sebastian Calonico, University of Miami, Coral Gables, FL. `<scalonico@bus.miami.edu>`.
Matias D. Cattaneo, University of Michigan, Ann Arbor, MI. `<cattaneo@umich.edu>`.
Max H. Farrell, University of Chicago, Chicago, IL. `<max.farrell@chicagobooth.edu>`.

**References**

**See Also**

*lpbwselect*

**Examples**

```r
x <- runif(500)
y <- sin(4*x) + rnorm(500)
est <- lprobust(y,x)
summary(est)
```

---

**nprobust.plot**

*Graphical Presentation of Results from nprobust Package.*

**Description**

*nprobust.plot* plots estimated density and regression function using the *nprobust* package. A detailed introduction to this command is given in Calonico, Cattaneo and Farrell (2017b).

Companion commands: *lprobust* for local polynomial point estimation and inference procedures, and *kdrobust* for kernel density point estimation and inference procedures.

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/).
nprobust.plot

Usage

nprobust.plot(..., alpha = NULL, type = NULL, CItype = NULL,
   title = "", xlabel = "", ylabel = "", lty = NULL, lwd = NULL,
   lcol = NULL, pty = NULL, pwd = NULL, pcol = NULL, CIscale = NULL,
   Cicol = NULL, legendTitle = NULL, legendGroups = NULL)

Arguments

... Objects returned by kdrobust or lprobust.

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.

type String, one of "line" (default), "points" or "both", how the point estimates are plotted. If more than one is provided, they will be applied to data series accordingly.

CItype 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.

title, xlabel, ylabel Strings, title of the plot and labels for x- and y-axis.

lty 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.

lwd 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.

lcol 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.

pty 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.

pwd 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.

pcol 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.

CIscale 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.
nprobust.plot

CIIcol 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.

legendTitle String, title of legend.
legendGroups String Vector, group names used in legend.

Details

Companion command: lprobust for local polynomial-based regression functions and derivatives estimation.

Value

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

Author(s)

Sebastian Calonico, University of Miami, Coral Gables, FL. <scalonico@bus.miami.edu>.
Matias D. Cattaneo, University of Michigan, Ann Arbor, MI. <cattaneo@umich.edu>.
Max H. Farrell, University of Chicago, Chicago, IL. <max.farrell@chicagobooth.edu>.

References


See Also

lprobust, kdrobust, ggplot2

Examples

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

*Topic **LPR**
  kdrobust, 4
  lprobust, 9

*Topic **Robust Estimation**
  kdrobust, 4
  lprobust, 9

**ggplot2**, 12, 13

**kdbwselect**, 2, 2, 3–6
**kdrobust**, 2–4, 4, 11–13

**lpbwselect**, 2, 6, 6, 7–11
**lprobust**, 2, 6, 8, 9, 9, 11–13

**nprobust (nprobust-package)**, 2
**nprobust-package**, 2
**nprobust.plot**, 2, 4, 9, 11

**par**, 12, 13