Package 'nprobust'

November 7, 2016

Type Package

Title Robust Data-Driven Statistical Inference for Local Polynomial Regression and Kernel Density Estimation

Version 0.0.1

Date 2016-11-06

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 analytical statistical inference for Local Polynomial Regression estimators and Kernel Density Estimation.

License GPL-2

Imports Rcpp

LinkingTo Rcpp, RcppArmadillo

NeedsCompilation yes

Repository CRAN

Date/Publication 2016-11-07 17:22:14

R topics documented:

nprobust-package *Robust Data-Driven Statistical Inference for Local Polynomial Regression and Kernel Density Estimation*

Description

This package provides tools for data-driven analytical statistical inference for Local Polynomial Regression (LPR) estimators and Kernel Density Estimation (KDE): `lprobust` to construct local-polynomial point estimators and robust confidence intervals, `lpbwselect` to perform bandwidth selection for the different procedures implemented.

Details

Package:	nprobust
Type:	Package
Version:	0.0.1
Date:	2016-11-06
License:	GPL-2

Function for LPR statistical inference: `lprobust`
Function for LPR bandwidths selection: `lpbwselect`
Function for KDE statistical inference: `kdrobust`
Function for KDE bandwidths selection: `kdbwselect`

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*

Description

`kdbwselect` implements bandwidth selectors for kernel density estimation and inference procedures developed in Calonico, Cattaneo and Farrell (2016a).

Companion commands are: `kdrobust` for point estimation and inference procedures.

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

Usage

```
kdbwselect(x, c, deriv = 0, p = 2, kernel = "epan",
bwselect = "mse", all = FALSE, subset = NULL)
```

Arguments

x	the data from which the estimate is to be computed.
c	specifies the evaluation point in x.
p	specifies the order of the kernel used to construct the point-estimator; default is p = 2.
deriv	specifies the order of the derivative of the density function to be estimated. Default is deriv=0.
kernel	is the kernel function used to construct the kernel density estimator(s). Options are triangular (default option), epanechnikov and uniform.
bwselect	specifies the bandwidth selection procedure to be used. Options are: mse MSE-optimal bandwidth selector for the kernel density estimator. cer CER-optimal bandwidth selector for the kernel density estimator. Note: MSE = Mean Square Error; CER = Coverage Error Rate. Default is bwselect=mse. For details on implementation see Calonico, Cattaneo and Farrell (2016a), and the companion software articles.
all	if specified, kdbwselect reports all available bandwidth selection procedures.
subset	an optional vector specifying a subset of observations to be used.

Value

bws	matrix containing the estimated bandwidths for each selected procedure.
bwselect	bandwidth selection procedure employed.
kernel	kernel function used to construct the kernel density estimator(s).
p	order of the kernel used to construct the point-estimator.

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

Calonico, S., M. D. Cattaneo, and M. H. Farrell. 2016. On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. Working Paper. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Farrell_2016_JASA.pdf.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. 2016b. rdrobust: Software for Regression Discontinuity Designs. Working Paper. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Farrell-Titiunik_2016_Stata.pdf.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014a. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. Econometrica 82(6): 2295-2326. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Titiunik_2014_ECMA.pdf.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014b. Robust Data-Driven Inference in the Regression-Discontinuity Design. Stata Journal 14(4): 909-946. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Titiunik_2014_Stata.pdf.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2015b. rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs. R Journal 7(1): 38-51. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Titiunik_2015_R.pdf.

See Also

kdrobust

Examples

```
x<-runif(1000,-1,1)
kdbwselect(x, c=0)
```

---

kdrobust *Kernel Density Estimation with Robust Confidence Intervals*

Description

kdrobust implements kernel density estimators with robust bias-corrected confidence intervals and inference procedures developed in Calonico, Cattaneo and Farrell (2016a). It also computes alternative estimation and inference procedures available in the literature.

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

Usage

```
kdrobust(x, c , deriv = 0, p = 2, h = NULL, b = NULL, rho = NULL,
kernel = "epan", bwselect = "mse", level = 95, all = FALSE, subset = NULL)
```

Arguments

x	the data from which the estimate is to be computed.
c	specifies the evalution point in x.
p	specifies the order of the kernel used to construct the point-estimator; default is p = 2.
deriv	specifies the order of the derivative of the density function to be estimated. Default is deriv=0.
h	specifies the main bandwidth used to construct the KD point estimator. If not specified, bandwidth h is computed by the companion command kdbwselect.

b	specifies the bias bandwidth used to construct the bias-correction estimator. If not specified, bandwidth b is computed by the companion command kdbwselect.
rho	specifies the value of rho, so that the bias bandwidth b equals h/rho. Default is rho = 1 if h is specified but b is not.
kernel	is the kernel function used to construct the kernel density estimator(s). Options are triangular (default option), epanechnikov and uniform.
bwselect	specifies the bandwidth selection procedure to be used. 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 MSE-optimal bandwidth selector for the kernel density estimator. cer CER-optimal bandwidth selector for the kernel density estimator. Note: MSE = Mean Square Error; CER = Coverage Error Rate. Default is bwselect=mse. For details on implementation see Calonico, Cattaneo and Titiunik (2014a), Calonico, Cattaneo and Farrell (2016a), and Calonico, Cattaneo, Farrell and Titiunik (2016), and the companion software articles.
level	sets the confidence level for confidence intervals; default is level = 95.
all	if specified, kdrobust reports three different procedures: (i) conventional LPR estimates with conventional standard errors. (ii) bias-corrected estimates with conventional standard errors. (iii) bias-corrected estimates with robust standard errors.
subset	an optional vector specifying a subset of observations to be used.

Value

c	cutoff value.
N	sample size used.
p	order of the kernel used for estimation of the density function.
h	bandwidth used for estimation of the KDE.
b	bandwidth used for estimation of the bias of the KDE.
coef	vector containing conventional and bias-corrected KDE.
se	vector containing conventional and robust standard errors of the KDE.
pv	vector containing the p-values associated with conventional, bias-corrected and robust KDE.
ci	matrix containing the confidence intervals associated with conventional, bias-corrected and robust KDE.

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

Calonico, S., M. D. Cattaneo, and M. H. Farrell. 2016. On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. Working Paper. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Farrell_2016_JASA.pdf.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. 2016b. rdrobust: Software for Regression Discontinuity Designs. Working Paper. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Farrell-Titiunik_2016_Stata.pdf.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014a. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. Econometrica 82(6): 2295-2326. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Titiunik_2014_ECMA.pdf.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2015b. rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs. R Journal 7(1): 38-51. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Titiunik_2015_R.pdf.

See Also

lpbwselect

Examples

```
x<-runif(1000,-1,1)
kdrobust(x, c=0)
```

---

lpbwselect *Bandwidth Selection Procedures for Local Polynomial Regression Estimators*

Description

lpbwselect implements bandwidth selectors for local polynomial regression point estimators and inference procedures developed in Calonico, Cattaneo and Farrell (2016a).

Companion commands are: lprobust for point estimation and inference procedures.

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

Usage

```
lpbwselect(y, x, c, p = 1, q = 2,
deriv = 0, kernel = "epa", bwselect = "mse", scaleregul = 1,
vce = "nn", nnmatch = 3, all = FALSE, subset = NULL)
```

Arguments

y	is the dependent variable.
x	is the independent variable).
c	specifies the evaluation point in x.
p	specifies the order of the local-polynomial used to construct the point-estimator; default is p = 1 (local linear regression).
q	specifies the order of the local-polynomial used to construct the bias-correction; default is q = 2 (local quadratic regression).
deriv	specifies the order of the derivative of the regression functions to be estimated. Default is deriv=0.
kernel	is the kernel function used to construct the local-polynomial estimator(s). Options are triangular (default option), epanechnikov and uniform.
bwselect	specifies the bandwidth selection procedure to be used. Options are: mse MSE-optimal bandwidth selector for the local polynomial regression estimator. cer CER-optimal bandwidth selector for the local polynomial regression estimator. Note: MSE = Mean Square Error; CER = Coverage Error Rate. Default is bwselect=mse. For details on implementation see Calonico, Cattaneo and Farrell (2016a), and the companion software articles.
scaleregul	specifies scaling factor for the regularization term added to the denominator of the bandwidth selectors. Setting scaleregul = 0 removes the regularization term from the bandwidth selectors; default is scaleregul = 1.
vce	specifies the procedure used to compute the variance-covariance matrix estimator. Options are: nn for heteroskedasticity-robust nearest neighbor variance estimator with nnmatch the (minimum) number of neighbors to be used. hc0 for heteroskedasticity-robust plug-in residuals variance estimator without weights. hc1 for heteroskedasticity-robust plug-in residuals variance estimator with hc1 weights. hc2 for heteroskedasticity-robust plug-in residuals variance estimator with hc2 weights. hc3 for heteroskedasticity-robust plug-in residuals variance estimator with hc3 weights. Default is vce=nn.
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
all	if specified, lpbwselect reports all available bandwidth selection procedures.
subset	an optional vector specifying a subset of observations to be used.

Value

bws	matrix containing the estimated bandwidths for each selected procedure.
bwselect	bandwidth selection procedure employed.
kernel	kernel function used to construct the local-polynomial estimator(s).
p	order of the local-polynomial used to construct the point-estimator.
q	order of the local-polynomial used to construct the bias-correction estimator.

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

Calonico, S., M. D. Cattaneo, and M. H. Farrell. 2016. On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. Working Paper. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Farrell_2016_JASA.pdf.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. 2016a. Regression Discontinuity Designs using Covariates. Working Paper. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Farrell-Titiunik_2016_wp.pdf.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. 2016b. rdrobust: Software for Regression Discontinuity Designs. Working Paper. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Farrell-Titiunik_2016_Stata.pdf.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014a. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. Econometrica 82(6): 2295-2326. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Titiunik_2014_ECMA.pdf.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014b. Robust Data-Driven Inference in the Regression-Discontinuity Design. Stata Journal 14(4): 909-946. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Titiunik_2014_Stata.pdf.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2015b. rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs. R Journal 7(1): 38-51. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Titiunik_2015_R.pdf.

See Also

lprobust

Examples

```
x<-runif(1000,-1,1)
y<-5+3*x+rnorm(1000)
lpbwselect(y,x, c=0)
```

lprobust *Local-Polynomial Estimation with Robust Confidence Intervals*

Description

lprobust implements local polynomial point estimators with robust bias-corrected confidence intervals and inference procedures developed in Calonico, Cattaneo and Farrell (2016a). It also computes alternative estimation and inference procedures available in the literature.

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

Usage

```
lprobust(y, x, c, p = 1, q = 2, deriv = 0,
h = NULL, b = NULL, rho = NULL, kernel = "epa", bwselect = "mse", scaleregul = 1,
vce = "nn", nnmatch = 3, level = 95, all = FALSE, subset = NULL)
```

Arguments

y	is the dependent variable.
x	is the independent variable.
c	specifies the evalution point in x.
p	specifies the order of the local-polynomial used to construct the point-estimator; default is p = 1 (local linear regression).
q	specifies the order of the local-polynomial used to construct the bias-correction; default is q = 2 (local quadratic regression).
deriv	specifies the order of the derivative of the regression functions to be estimated. Default is deriv=0
h	specifies the main bandwidth used to construct the LPR point estimator. If not specified, bandwidth h is computed by the companion command lpbwselect.
b	specifies the bias bandwidth used to construct the bias-correction estimator. If not specified, bandwidth b is computed by the companion command lpbwselect.
rho	specifies the value of rho, so that the bias bandwidth b equals h/rho. Default is rho = 1 if h is specified but b is not.
kernel	is the kernel function used to construct the local-polynomial estimator(s). Options are triangular (default option), epanechnikov and uniform.
bwselect	specifies the bandwidth selection procedure to be used. 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 MSE-optimal bandwidth selector for the local polynomial regression estimator. cer CER-optimal bandwidth selector for the local polynomial regression estimator.

Note: MSE = Mean Square Error; CER = Coverage Error Rate. Default is `bwselect=mse`. For details on implementation see Calonico, Cattaneo and Titiunik (2014a), Calonico, Cattaneo and Farrell (2016a), and Calonico, Cattaneo, Farrell and Titiunik (2016), and the companion software articles.

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

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

`nn` for heteroskedasticity-robust nearest neighbor variance estimator with `nnmatch` the (minimum) number of neighbors to be used.

`hc0` for heteroskedasticity-robust plug-in residuals variance estimator without weights.

`hc1` for heteroskedasticity-robust plug-in residuals variance estimator with `hc1` weights.

`hc2` for heteroskedasticity-robust plug-in residuals variance estimator with `hc2` weights.

`hc3` for heteroskedasticity-robust plug-in residuals variance estimator with `hc3` weights.

Default is `vce=nn`.

`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` sets the confidence level for confidence intervals; default is `level = 95`.

`all` if specified, `lprobust` reports three different procedures:

(i) conventional LPR estimates with conventional standard errors.

(ii) bias-corrected estimates with conventional standard errors.

(iii) bias-corrected estimates with robust standard errors.

`subset` an optional vector specifying a subset of observations to be used.

Value

`c` cutoff value.

`N` sample size used.

`N` overall sample size.

`p` order of the polynomial used for estimation of the regression function.

`q` order of the polynomial used for estimation of the bias of the regression function.

`h` bandwidth used for estimation of the regression function.

`b` bandwidth used for estimation of the bias of the regression function estimator.

`coef` vector containing conventional and bias-corrected local-polynomial estimates.

`se` vector containing conventional and robust standard errors of the local-polynomial estimates.

pv	vector containing the p-values associated with conventional, bias-corrected and robust local-polynomial estimates.
ci	matrix containing the confidence intervals associated with conventional, bias-corrected and robust local-polynomial estimates.

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

Calonico, S., M. D. Cattaneo, and M. H. Farrell. 2016. On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. Working Paper. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Farrell_2016_JASA.pdf.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. 2016a. Regression Discontinuity Designs using Covariates. Working Paper. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Farrell-Titiunik_2016_wp.pdf.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. 2016b. rdrobust: Software for Regression Discontinuity Designs. Working Paper. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Farrell-Titiunik_2016_Stata.pdf.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014a. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. Econometrica 82(6): 2295-2326. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Titiunik_2014_ECMA.pdf.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014b. Robust Data-Driven Inference in the Regression-Discontinuity Design. Stata Journal 14(4): 909-946. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Titiunik_2014_Stata.pdf.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2015b. rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs. R Journal 7(1): 38-51. http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Titiunik_2015_R.pdf.

See Also

lpbwselect

Examples

```
x<-runif(1000,-1,1)
y<-5+3*x+rnorm(1000)
lprobust(y,x, c=0)
```

nprobust-internal *Package Internal Functions*

Description

Internal functions which are not part of the package application programming interface.

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>.

Index