Statistics/Data analysis

Title

kdrobust —— Kernel Density Estimation with Robust Bias-Corrected Confidence Intervals and Inference Procedures.

Syntax

kdrobust *varname* [*if*] [*in*] [, **eval(***gridvar***) neval(***#***) h(***#***) b(***#***) rho(***#***) kernel(***kernelfn***) bwselect(***bwmethod***) bwcheck(***#***) imsegrid(***#***) level(***#***) separator(***#***)** *genvars plot* **graph_options(***gphopts***)**]

Description

kdrobust implements kernel density point estimators with robust bias-corrected confidence intervals and inference procedures developed in Calonico, Cattaneo and Farrell (2018). See also Calonico, Cattaneo and Farrell (2020) for related optimality results. 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.

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

Companion command is: kdbwselect for data-driven bandwidth selection.

Related Stata and R packages useful for empirical analysis are described in the following website:

https://nppackages.github.io/

Options

eval(*gridvar*) specifies the grid of evaluation points for *xvar*. By default it uses 30 equally spaced points over to support of *xvar*.

neval(*#*) specifies the number of evaluation points to estimate the regression functions. Default is 30 evaluation points.

h(*hvar*) specifies the main bandwidth (*h*) used to construct the point estimator for each evaluation point. If not specified, it is computed by the companion command kdbwselect.

b(*bvar*) specifies the bias bandwidth (*b*) used to construct the bias-correction estimator for each evaluation point. If not specified, it is computed by the companion command kdbwselect.

rho(*#*) specifies the value of *rho*, so that the bias bandwidth *b* equals *b=h/rho*. Default is **rho(***1***)** if *h* is specified but *b* is not.

kernel(*kernelfn*) specifies the kernel function used to construct the kernel density estimator(s). Options are: **epanechnikov**, and **uniform**. Default is **kernel(***epanechnikov***)**.

bwselect(*bwmethod*) 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-dpi second-generation DPI implementation of MSE-optimal bandwidth. Default choice.
mse-rot ROT implementation of MSE-optimal bandwidth.
imse-dpi second-generation DPI implementation of IMSE-optimal bandwidth.
imse-rot ROT implementation of IMSE-optimal bandwidth.
ce-dpi second generation DPI implementation of CE-optimal bandwidth.
ce-rot ROT implementation of CE-optimal bandwidth.
Note: MSE = Mean Square Error; IMSE = Integrated Mean Squared Error; CE = Coverage Error; DPI = Direct Plug-in; ROT = Rule-of-Thumb.
Default is **bwselect(***mse-dpi***)**. For details on implementation see Calonico, Cattaneo and Farrrell (2019).

bwcheck(#) specifies an optional positive integer so that the selected bandwidth is enlarged to have at least # effective observations available for each evaluation point.

imsegrid(#) number of evaluations points used to compute the IMSE bandwidth selector. Default is 30 points.

level(#) specifies confidence level for confidence intervals. Default is **level(***95***)**.

separator(#) draws separator line after every # variables; default is separator(5).

plot generates the local polynomial regression plot.

genvars generates new variables storing the following results.
kdrobust_eval evaluation points.
kdrobust_h bandwidth h.
kdrobust_b bandwidth b.
kdrobust_nh effective sample size.
kdrobust_gx_us conventional local polynomial estimate.
kdrobust_se_us conventional standard error for the local polynomial estimator.
kdrobust_gx_bc bias-corrected local polynomial regression estimate.
kdrobust_se_rb robust standard error for the local polynomial estimator.
kdrobust_ci_l_rb lower end value of the robust confidence interval.
kdrobust_ci_r_rb upper end value of the robust confidence interval.

graph_options(*gphopts*) specifies graphical options to be passed on to the underlying graph command.

Setup
. sysuse auto

Kernel density estimates for length
. kdrobust length

Kernel density estimates for length
. kdrobust length, plot genvars

Saved results

kdrobust saves the following in **e()**:

Scalars

e(N)	original number of observations

Macros

e(varname)	name of variable
e(bwselect)	bandwidth selection choice
e(kernel)	kernel choice

Matrices

e(Result)	estimation result

References

Calonico, S., M. D. Cattaneo, and M. H. Farrell. 2018. On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. *Journal of the American Statistical Association*, 113(522): 767-779.

Calonico, S., M. D. Cattaneo, and M. H. Farrell. 2019. nprobust: Nonparametric Kernel-Based Estimation and Robust Bias-Corrected Inference. *Journal of Statistical Software*, 91(8): 1-33. doi: 10.18637/jss.v091.i08.

Calonico, S., M. D. Cattaneo, and M. H. Farrell. 2020. Coverage Error Optimal Confidence Intervals for Local Polynomial Regression, working paper.

Fan, J., and Gijbels, I. 1996. Local Polynomial Modelling and Its Applications, London: Chapman and Hall.

Wand, M., and Jones, M. 1995. Kernel Smoothing, Florida: Chapman & Hall/CRC.

Authors

Sebastian Calonico, Columbia University, New York, NY. sebastian.calonico@columbia.edu.

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

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