Statistics/Data Analysis

Title

lprobust —— Local Polynomial Regression Estimation with Robust Bias-Corrected Confidence Intervals and Inference Procedures.

Syntax

lprobust *yvar xvar* [*if*] [*in*] [, **eval(***gridvar***) neval(***#***) deriv(***#***) p(***#***) h(***hvar***) b(***bvar***) rho(***#***) kernel(***kernelfn***) bwselect(***bwmethod***) bwcheck(***#***) imsegrid(***#***) vce(***vcetype [vceopt]***) level(***#***) bwregul(***#***) separator(***#***)** *interior genvars covgrid plot* **graph_options(***gphopts***)**]

Description

lprobust implements local polynomial regression 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: lpbwselect for data-driven bandwidth selection.

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

https://sites.google.com/site/nppackages/

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.

deriv(*#*) specifies the order of the derivative of the regression functions to be estimated. Default is **deriv(***0***)**.

p(*#*) specifies the order of the local polynomial used to construct the point estimator. Default is **p(***1***)** (local linear regression).

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 lpbwselect.

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 lpbwselect.

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 local-polynomial estimator(s). Options are: **tri**angular, **epa**nechnikov, **uni**form and **gau**ssian. 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.

vce(*vcetype [vceopt1]*) specifies the procedure used to compute the variance-covariance matrix estimator. Options are:
vce(*nn [nnmatch]*) for heteroskedasticity-robust nearest neighbor variance estimator with *nnmatch* indicating the minimum number of neighbors to be used.
vce(*hc0***)** for heteroskedasticity-robust plug-in residuals variance estimator without weights.
vce(*hc1***)** for heteroskedasticity-robust plug-in residuals variance estimator with *hc1* weights.
vce(*hc2***)** for heteroskedasticity-robust plug-in residuals variance estimator with *hc2* weights.
vce(*hc3***)** for heteroskedasticity-robust plug-in residuals variance estimator with *hc3* weights.
vce(nncluster *clustervar [nnmatch]***)** for cluster-robust nearest neighbor variance estimation using with *clustervar* indicating the cluster ID variable and *nnmatch* matches indicating the minimum number of neighbors to be used.
vce(cluster *clustervar***)** for cluster-robust plug-in residuals variance estimation with degrees-of-freedom weights and *clustervar* indicating the cluster ID variable.
Default is **vce(***nn 3***)**.

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

bwregul(#) specifies scaling factor for the regularization term added to the denominator of the bandwidth selectors. Setting **bwregul(***0***)** removes the regularization term from the bandwidth selectors. Default is **bwregul(***1***)**.

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

interior option to set all evaluation points to be interior points. This option affects only data-driven bandwith selection via lpbwselect.

covgrid option to compute two covariance matrices (cov_us and cov_rb) for classical and robust covariances across point estimators over the grid of evaluation points.}

plot generates the local polynomial regression plot.

genvars generates new variables storing the following results.
lprobust_eval evaluation points.
lprobust_h bandwidth h.
lprobust_b bandwidth b.
lprobust_nh effective sample size.
lprobust_gx_us conventional local polynomial estimate.
lprobust_se_us conventional standard error for the local polynomial estimator.
lprobust_gx_bc bias-corrected local polynomial regression estimate.
lprobust_se_rb robust standard error for the local polynomial estimator.
lprobust_ci_l_rb lower end value of the robust confidence interval.
lprobust_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
. webuse motorcycle

Local linear regression with second-generation DPI implementation of MSE-optimal bandwidth
. lprobust accel time

Same as above, but generating a plot and the corresponding output variables
. lprobust accel time, plot genvars

Saved results

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

Scalars

e(N)	original number of observations
e(p)	order of the polynomial used for estimation of the regression function

Macros

e(varname)	name of variable
e(clustvar)	name of cluster variable
e(bwselect)	bandwidth selection choice
e(kernel)	kernel choice
e(vce)	vce 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.