Title

lpbwselect —— Bandwidth Selection Procedures for Local Polynomial Regression Estimation and Inference.

Syntax

lpbwselect *yvar xvar* [*if*] [*in*] [, **eval(***gridvar***) neval(***#***) deriv(***#***) p(***#***) rho(***#***) kernel(***kernelfn***) bwselect(***bwmethod***) bwcheck(***#***) imsegrid(***#***) vce(***vcetype [vceopt]***) bwregul(***#***) separator(***#***)** *interior*]

Description

lpbwselect implements bandwidth selectors for local polynomial regression point estimators 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 bandwidth selectors 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: lprobust for local polynomial point estimation and inference procedures.

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.

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).

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: **triangular**, **epanechnikov**, **uniform** and **gaussian**. Default is **kernel(***epanechnikov***)**.

bwselect(*bwmethod*) bandwidth selection procedure to be used. 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***)**.

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).

Example:

Setup
. webuse motorcycle

Second-generation DPI implementation of MSE-optimal bandwidth
. lpbwselect accel time

Saved results

lpbwselect 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(bws)	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.