



Title

lpbwdensity — Bandwidth Selection for Local Polynomial Density Estimation and Inference.

Syntax

```
lpbwdensity var [if] [in] [, grid(var) p(#) v(#) bwselect(BwMethod)
kernel(KernelFn) cweights(var) pweights(var) genvars(VarName) separator(#)
regularize ]
```

Description

lpdensity implements the bandwidth selector for local polynomial based density (and derivatives) estimation, proposed in [Cattaneo, Jansson and Ma \(2017a\)](#). See [Cattaneo, Jansson and Ma \(2017b\)](#) for more implementation details and illustrations.

Companion R functions are also available [here](#).

Related Stata and R packages useful for nonparametric estimation and inference are described in the following website:

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

Options

grid(var) specifies the grid on which density is estimated. When set to default, grid points will be chosen as 0.05–0.95 percentiles of the data, with 0.05 step size.

p(#) specifies the the order of the local-polynomial used to construct point estimates. Default is **p(2)** (local quadratic regression).

v(#) specifies the derivative of distribution function to be estimated. **v(0)** for the distribution function, **v(1)** (default) for the density function, etc.

bwselect(*BwMethod*) specifies method for data-driven bandwidth selection. This option will be ignored if **bw(var)** is provided. Options are:
mse-dpi for mean squared error-optimal bandwidth selected for each grid point. This is the default option.
imse-dpi for integrated MSE-optimal bandwidth, common for all grid points.
mse-rot for rule-of-thumb bandwidth with Gaussian reference model.
imse-rot for integrated rule-of-thumb bandwidth with Gaussian reference model.

kernel(*KernelFn*) specifies the kernel function used to construct the local-polynomial estimator(s). Options are: **triangular**, **epanechnikov**, and **uniform**. Default is **triangular**.

cweights(var) specifies weights used for counterfactual distribution construction.

pweights(var) specifies weights used in sampling. Should be nonnegative.

genvars(*VarName*) specified if new variables should be generated to store estimation results. If *VarName* is provided, the following new variables will be generated:
VarName_grid (grid points),
VarName_bw (bandwidth),
VarName_nh (effective sample size).

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

noregularize suppresses the feature that bandwidth is chosen such that at least 20 + **p(#)** + 1 observations will be included.

Examples

```
Generate artifitial data:
. set obs 1000
. set seed 42
. gen lpd_data = rnormal()

MSE-optimal bandwidths for empirical quantiles:
. lpbwdensity lpd_data

Save estimation results to variables:
. capture drop temp_*
. lpbwdensity lpd_data, genvars(temp)
```

Saved results

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

Scalars	
e(N)	sample size
e(p)	option p (#)
e(v)	option v (#)
Macros	
e(bwselect)	option bwselect (<i>BwMethod</i>)
e(kernel)	option kernel (<i>KernelFn</i>)
Matrices	
e(result)	estimation result

References

Cattaneo, M. D., Michael Jansson, and Xinwei Ma. 2017a. Simple Local Polynomial Density Estimators. Working paper, University of Michigan.

Cattaneo, M. D., Michael Jansson, and Xinwei Ma. 2017b. lpdensity: Local Polynomial Density Estimation and Inference. Working paper, University of Michigan.

Authors

Matias D. Cattaneo, University of Michigan, Ann Arbor, MI. cattaneo@umich.edu.

Michael Jansson, University of California at Berkeley, Berkeley, CA. mjansson@econ.berkeley.edu.

Xinwei Ma, University of Michigan, Ann Arbor, MI. xinweima@umich.edu.