



## Title

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

## Syntax

```
lpbwdensity Var [if] [in] [,
    grid(Var) p(#) v(#) kernel(KernelFn) bwselect(BwMethod) nomasspoints
    nostdvar
    nlocalmin(#) nuniquemin(#) noregularize
    cweights(Var) pweights(Var)
    genvars(NewVarName)
    rgrid(Var) rindex(Var) separator(#)
    ]
```

## Description

**lpbwdensity** implements the bandwidth selection methods for local polynomial based density (and derivatives) estimation proposed and studied in [Cattaneo, Jansson and Ma \(2020\)](#) and [Cattaneo, Jansson and Ma \(2023\)](#). See [Cattaneo, Jansson and Ma \(2022\)](#) for more implementation details and illustrations.

Companion command: [lpdensity](#) for estimation and robust bias-corrected inference.

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

Related Stata and R packages are available in the following website:

<https://nppackages.github.io/>

## Options

### Bandwidth Selection

**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 funtion, etc.

**kernel**(*KernelFn*) specifies the kernel function used to construct the local-polynomial estimator(s).

**triangular**  $K(u) = (1 - |u|) * (|u| \leq 1)$ . This is the default option.

**epanechnikov**  $K(u) = 0.75 * (1 - u^2) * (|u| \leq 1)$ .

**uniform**  $K(u) = 0.5 * (|u| \leq 1)$ .

**bwselect**(*BwMethod*) specifies method for data-driven bandwidth selection. This option will be ignored if **bw**(*Var*) is provided.

**mse-dpi** mean squared error optimal bandwidth for each grid point. This is the default option.

**imse-dpi** integrated mean squared error optimal bandwidth which is common for all grid points.

**mse-rot** rule-of-thumb bandwidth based on a Gaussian reference model.

**imse-rot** integrated rule-of-thumb bandwidth based on a Gaussian reference model which is common for all grid points.

**nomasspoints** will not adjust for mass points in the data.

**nostdvar** will not standardize the data for bandwidth selection. Note that this may lead to unstable performance of the numerical optimization procedure.

---

### Local Sample Size Checking

---

**nlocalmin**(#) specifies the minimum number of observations in each local neighborhood. This option will be ignored if set to 0, or if **noregularize** is used. The default value is **20+p(#)+1**.

**nuniquemin**(#) specifies the minimum number of unique observations in each local neighborhood. This option will be ignored if set to 0, or if **noregularize** is used. The default value is **20+p(#)+1**.

**noregularize** suppresses local sample size checking.

---

### Weights

---

**cweights**(Var) specifies weights used for counterfactual distribution construction.

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

---

### Storing and displaying results

---

**genvars**(NewVarName) specifies if new variables should be generated to store estimation results. If *newVarName* is provided, the following new variables will be generated:

```
NewVarName_grid grid points,
NewVarName_bw   bandwidths,
NewVarName_nh   local/effective sample sizes.
```

**rgrid**(var) specifies a set of grid points to display the results. When omitted, this will be the same as **grid**(Var).

**rindex**(var) specifies a set of indices to display the results. This option will be ignored if **rgrid**(Var) is provided.

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

### 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 (BwMethod)
e(kernel)     option kernel (KernelFn)
```

Matrices

```
e(result)     estimation result
```

### References

- Cattaneo, M. D., Michael Jansson, and Xinwei Ma. 2020. Simple Local Polynomial Density Estimators.  
*Journal of the American Statistical Association* 115(531): 1449-1455.
- Cattaneo, M. D., Michael Jansson, and Xinwei Ma. 2022. lpdensity: Local Polynomial Density Estimation and Inference.  
*Journal of Statistical Software* 101(2): 1-25.
- Cattaneo, M. D., Michael Jansson, and Xinwei Ma. 2023. Local Regression Distribution Estimators.  
*Journal of Econometrics*, forthcoming.

### **Authors**

- Matias D. Cattaneo, Princeton University, Princeton, NJ. [cattaneo@princeton.edu](mailto:cattaneo@princeton.edu).
- Michael Jansson, University of California Berkeley, Berkeley, CA.  
[mjansson@econ.berkeley.edu](mailto:mjansson@econ.berkeley.edu).
- Xinwei Ma, University of California San Diego, La Jolla, CA. [xlma@ucsd.edu](mailto:xlma@ucsd.edu).