

Title

lwcate - Estimation of Conditional Average Treatment Effects using locally-weighted regressions.

Syntax

`lwcate` *command* *depvar* *treatvar* [*indepvars*] [*if*] , *xtilde* [*options*]

Inputs	Description
<i>command</i>	Equation command (i.e. <i>reg</i> , <i>probit</i> , <i>logit</i> , <i>tobit</i> , <i>nbreg</i> , <i>oaxaca</i>)
<i>depvar</i>	Dependent Variable, y
<i>treatvar</i>	Treatment Dummy, x_τ (0 = Control, 1 = Treated)
<i>indepvars</i>	Independent Variables, \mathbf{x} ; Excluding x_τ and \tilde{x}
<i>xtilde</i> (varname)	Conditional Variable of Interest, \tilde{x}

Description

`lwcate` uses locally-weighted regressions to estimate Conditional Average Treatment Effects (CATEs), predict individual-level potential outcomes and conveniently plot useful outputs. It is applicable to experimental, quasi-experimental and natural experimental designs; where researchers want to estimate treatment effects conditional on a variable of interest \tilde{x} . The command is flexible, estimating the following regression:

$$\mathbf{y} = \beta_0(\tilde{\mathbf{x}}) + \beta_\tau(\tilde{\mathbf{x}})\mathbf{D} + \boldsymbol{\beta}(\tilde{\mathbf{x}})\mathbf{X} + \mathbf{u} \quad (1)$$

Where \mathbf{y} is the outcome variable, \mathbf{D} the treatment dummy, \mathbf{X} any additional explanatory variables and \mathbf{u} the error term. The coefficients are functions of the observable characteristic(s) of interest $\tilde{\mathbf{x}}$: the constant, $\beta_0(\tilde{\mathbf{x}})$, CATE, $\beta_\tau(\tilde{\mathbf{x}})$, and additional coefficients, $\boldsymbol{\beta}(\tilde{\mathbf{x}})$.

It does so by running a set of M locally-weighted regressions, conditional on each $j \in M$ quantile of the conditional variable of interest, $q(\tilde{x})$. Each j local regression provides coefficient estimates $\boldsymbol{\beta}(\tilde{x}_i)$ for each individual $i \in N$ for all independent variables \mathbf{x}_i . Counterfactual predictions of \hat{y}_i are further estimated (the potential outcomes of belonging to control and treatment groups) for each individual i . The command is designed primarily for $\tilde{\mathbf{x}}$ values which are continuous, but is flexible in allowing for bunching and has options if \tilde{x} is discrete.

A variety of kernel weighting functions are available, to estimate the weights for each local regression. The Beta distribution, is proposed and used as the default option; but the standard weighting functions (Uniform, Triangular, Normal and Epanechnikov) are included as options. A “discrete” option is also available to explicitly run a stratified analysis. Estimations can be ran using kernel, local-linear or local-polynomial methods.

Additionally, by using the *oaxaca* command, `lwcate` allows for the decomposition of CATEs. Estimating the separate components which constitute the total CATE, across $\tilde{\mathbf{x}}$.

To use `lwcate` download “lwcate.ado” and save into the directory “c:\ado\personal\”. Alternatively, type “sysdir” into the command bar to find the PERSONAL ado location and save it there. Then run the command.

<i>options</i>	Description
<u>xstrat</u>(string)	Additional discrete \tilde{x} variable on which to use for stratified analysis.
<u>quantiles</u>(#)	Number of Quantiles, M , (i.e. 100 = Percentiles, 10 = Deciles) *Default = 100
<u>minx</u>(#)	Minimum quantile of \tilde{x} at which to begin estimation (between 0 and 1) *Default = 0
<u>maxx</u>(#)	Maximum quantile of \tilde{x} at which to stop estimation (between 0 and 1) *Default = 1
<u>exoptions</u>(string)	Any extra options for equation
<u>vce</u>(vcetype)	Standard vce options, or “bootstrap” for bootstrapped SEs (computationally heavy).
<u>bsiter</u>(#)	How many bootstrap iterations are wanted for bootstrapped vce’s
Non-Parametric	
<u>kweight</u>(string)	The type of Kernel Weights used *Default = “beta”, Options = “uniform”, “triangular”, “normal” and “epanechnikov”. “discrete” is also available for discrete \tilde{x} variables.
<u>bvar</u>(string)	Allows an alternative variance assumption for the Beta distribution to be used. *Default = “”, Options = “constant”.
<u>estimation</u>(string)	Local estimator: *Default = “kernel”, Option = “linear”, “linearint”, “poly” or “polyint”.
<u>polydegree</u>(#)	Degree of polynomial, if local-polynomial is chosen. *Default = 2.
<u>s</u>(#)	Precision Parameter when using Beta Weights; must be ≥ 3 . *Default = 10
<u>kbandwidth</u>(#)	The relative bandwidth used for optional kernel weights. *Default = 0.1.
Regression Discontinuity	
<u>bandwidth</u>(#)	Bandwidth, around which observations are excluded
<u>running</u>(string)	Forcing/Running Variable (which equals zero at the discontinuity)
Oaxaca-Blinder Decomposition	
<u>oblinder</u>(string)	Include string of all options to be used with oaxaca command; within this all usual oaxaca options should be included i.e. ob(by(treatvar) vce(robust) swap detail weight(0))
<u>obexclusion</u>(string)	Variables to be excluded when calculating total and component effects.
<u>obcutoff</u>(#)	Excludes individuals components from the decomposition graph for which the mean P-Value is greater than the stated value.
Graph	
<u>prediction</u>(string)	Predicted graphs; *Default = “scatter” (scatter of counterfactual predictions), Options: “line” (single line with CIs, of counterfactuals), “actual” (actual predictions, rather than counterfactual) and “both” (an expected line plus scatter)
<u>gtitle</u>(string)	Graph Title
<u>gsave</u>(string)	Name to Save Graph to
<u>gxaxis</u>(string)	What is Plotted on the X Axis. *Default quantiles of \tilde{x} ; alternative is “actual” which plots the expected value of \tilde{x} within the quantile of interest
<u>gminx</u>(#)	Minimum X which the Graph is plotted from
<u>gmaxx</u>(#)	Maximum X which the Graph is plotted from
<u>nodraw</u>(#)	Do not draw any graphs.

Quick Examples

The following examples are shown to give a quick intuition to the command. Here the following base case variables will be used, with stated tweaks for the particular model run. The dependent variable, $y = \text{health}$, the treatment dummy, $x_\tau = \text{treatment}$ and conditional variable, $\tilde{x} = \text{income}$.

Alternative Econometric Specifications

A. *Experimental Design. OLS regression.*

```
lwcate reg health treatment , x(income)
```

B. *Experimental Design. OLS regression, with added controls (age, sex and ethnicity).*

```
lwcate reg health treatment age sex ethnicity , x(income)
```

C. *Experimental Design. OLS regression, with additional stratified analysis on sex.*

```
lwcate reg health treatment , x(income) xstrat(sex)
```

D. *Experimental Design. Probit regression, with $y = \text{healthdummy}$; a binary variable.*

```
lwcate probit healthdummy treatment , x(income)
```

E. *Regression Discontinuity Design. OLS Regression.*

```
lwcate reg health treatment runvar i.treatment#runvar , x(income) r(runvar) b(50)
```

With the running/forcing variable **runvar** and a bandwidth of 50. Additional x variables **runvar** and **i.treatment#runvar** need to be included within the RDD econometric specification.

F. *Difference-in-Differences Design. OLS Regression.*

```
lwcate reg health groupXtime i.group i.time , x(income)
```

With the DID estimator variable **groupXtime** (the treatment effect of interest), alongside **group** and **time**, as dummy variables for belonging to the treated group and being observed after the intervention, respectively.

G. *Decomposing CATEs using oaxaca. Experimental Design.*

```
lwcate oaxaca health age sex ethnicity , x(income) ob(by(treatment) swap detail weight(0))
```

To decompose CATEs use **oaxaca** as the command variable. Options for **oaxaca** need to be included within **ob(string)**. Note the **treatment** dummy is only included within the **ob(string)**, not immediately following the dependent variable.

Alternative Options

I. *Different Quantiles; estimate results for 50 quantiles, for only $0.2 \leq q \leq 0.8$.*

```
lwcate reg health treatment , x(income) q(50) minx(0.2) maxx(0.8)
```

II. *Local-Linear Estimation and Lower Precision for Beta Weights.*

```
lwcate reg health treatment , x(income) est(linear) s(5)
```

III. *Normal Distribution for Kernel Weights, with a relative bandwidth of 0.25.*

```
lwcate reg health treatment , x(income) kw(normal) kb(0.25)
```

IV. *Discrete \tilde{x} variable; stratified econometric specification.*

```
lwcate reg health treatment , x(discretincome) kw(discrete)
```

V. Only estimate for Females.

```
lwcate reg health treatment if sex == 1 , x(income)
```

*VI. Add Option to **regress**; noconstant.*

```
lwcate reg health treatment , x(income) exop(noconstant)
```

VII. VCE Robust

```
lwcate reg health treatment , x(income) vce(robust)
```

VIII. Bootstrapped VCEs; with 200 bootstrap iterations.

```
lwcate reg health treatment , x(income) vce(bootstrap) bsi(200)
```

IX. Specific Graph Options.

```
lwcate reg health treatment , x(income) gt("CATEs") gs("GR_1") gminx(0.1) gmaxx(0.9) pr(line)
```