Honest Inference in Sharp Regression Discontinuity in STATA

Kwok Hao Lee

May 27, 2019

Contents

The STATA package `rdhonest` implements confidence intervals for the regression discontinuity (RD) parameter considered in Armstrong and Kolesar (2017), Armstrong and Kolesar (2016), and Kolesar and Rothe (2017). In this vignette, we demonstrate the implementation of these confidence intervals using datasets from Lee (2008) and Oreopoulos (2006).

Much of this help file follows that of Michal Kolesar's `rdhonest` package implemented in R. You can view that help file at `https://github.com/kolesarm/RDHonest/blob/master/inst/doc/RDhonest.pdf`.

1 Sharp RD model

In the sharp RD model, we observe units $i = 1, \dots, n$. The outcome y_i for the i-th unit is given by

$$y_i = f(x_i) + u_i,$$

where $f(x_i)$ is the expectation of y_i conditional on the running variable x_i and u_i is the regression error. A unit is treated if and only if the running variable x_i lies above a known cutoff c_0. The parameter of interest is given by the jump of f at the cutoff,

$$\beta = \lim_{x \downarrow c_0} f(x) - \lim_{x \uparrow c_0} f(x).$$

Let $\sigma^2(x_i)$ denote the conditional variance of u_i.

In the Lee dataset, the running variable corresponds to the margin of victory of a Democratic candidate in a US House election, and the treatment corresponds to winning the election. Therefore, the cutoff is zero. The outcome of interest is the Democratic vote share in the following election.

2 Plots

See R package help file for plots. Plotting functions are implemented in other packages; see, for example, `RDRobust`: `https://sites.google.com/site/rdpackages/rdrobust`.

3 Inference based on local polynomial estimates

The function `rdhonest` constructs one- and two-sided confidence intervals (CIs) around local linear and local quadratic estimators using either a user-supplied bandwidth (which is allowed to differ on either side of the cutoff), or bandwidth that is optimized for a given performance criterion. The sense of honesty is that, if the regression errors are normally distributed with known variance, the CIs are guaranteed to achieve correct coverage *in finite samples*, and achieve correct coverage asymptotically uniformly over the parameter space otherwise. Furthermore, because the CIs explicitly take into account the possible bias of the estimators, the asymptotic approximation doesn't rely on the bandwidth to shrink to zero at a particular rate.

To characterize the structure of the CIs, let $\hat{\beta}_{h_+,h_-}$ denote a local polynomial estimator with bandwidth equal to h_+ above the cutoff and h_- below the cutoff. Let $\beta_{h_+,h_-}(f)$ denote the expectation of this estimator, conditional on the covariates when the regression function equals f. Then the estimator has a bias given by $\beta_{h_+,h_-}(f) - \beta$.

We are interested in the worst-case bias over the parameter space $\mathcal{F}$. Let

$$B(\hat{\beta}_{h_+,h_-}) = \sup_{f \in \mathcal{F}} \beta_{h_+,h_-}(f) - \beta$$

denote this worst-case bias. Then the lower limit of a one-sided CI is given by

$$\hat{\beta}_{h_+,h_-} - B\left(\hat{\beta}_{h_+,h_-}\right) - z_{1-\alpha}\widehat{se}\left(\hat{\beta}_{h_+,h_-}\right),$$

where $z_{1-\alpha}$ is the $1-\alpha$ quantile of a standard normal distribution, and $\widehat{se}(\hat{\beta}_{h_+,h_-})$ is the standard error (an estimate of the standard deviation of the estimator). Contrast this with the lower limit of the usual CI: instead of only subtracting the usual critical value times the standard error, we also subtract the worst-case bias. This ensures that our one-sided CI has correct coverage at all points in the parameter space.

A two-sided CI is given by

$$\hat{\beta}_{h_+,h_-} \pm cv_{1-\alpha}\left(B\left(\hat{\beta}_{h_+,h_-}\right)/\widehat{se}\left(\hat{\beta}_{h_+,h_-}\right)\right) \times \widehat{se}\left(\hat{\beta}_{h_+,h_-}\right),$$

where the critical value function $cv_{1-\alpha}(b)$ corresponds to the $1-\alpha$ quantile of the $|N(b,1)|$ distribution. To see why using this critical value ensures honesty, decompose the t-statistic as

$$\frac{\hat{\beta}_{h_+,h_-} - \beta}{\widehat{se}\left(\hat{\beta}_{h_+,h_-}\right)} = \frac{\hat{\beta}_{h_+,h_-} - \beta_{h_+,h_-}(f)}{\widehat{se}\left(\hat{\beta}_{h_+,h_-}\right)} + \frac{\beta_{h_+,h_-}(f) - \beta}{\widehat{se}\left(\hat{\beta}_{h_+,h_-}\right)}.$$

By a central limit theorem, the first term on the right-hand side will be distributed standard normal, no matter the bias. The second term is bounded in absolute value by $B\left(\hat{\beta}_{h_+,h_-}\right)/\widehat{se}\left(\hat{\beta}_{h_+,h_-}\right)$; hence in large samples, the $1-\alpha$ quantile of the absolute value of the t-statistic will be bounded by $cv_{1-\alpha}(B(\hat{\beta}_{h_+,h_-})/\widehat{se}(\hat{\beta}_{h_+,h_-}))$. This approach gives tighter CIs than simply adding and subtracting $B(\hat{\beta}_{h_+,h_-})$ from the point estimate, in addition to adding and subtracting $z_{1-\alpha}\widehat{se}(\hat{\beta}_{h_+,h_-})$.

We exhibit a table of these critical values below. For R code that generates these critical values, see the help file for the `RDHonest` R package.

Table 1: Critical values

bias	alpha	cv
0.25	0.01	2.65224
0.25	0.05	2.01971
0.25	0.10	1.69558

3.1 Parameter space

To implement the honest CIs, one needs to specify the parameter space $\mathcal{F}$. The function `rdhonest` computes honest CIs when the parameter space $\mathcal{F}$ corresponds ot a second-order Taylor or second-order Holder smoothness class, each of which captures a different type of smoothness restriction.The second-order Taylor class assumes that f lies in the class of functions

$$\mathcal{F}_{\text{Taylor}}(M) = \{f_+ - f_- : f_+ \in \mathcal{F}_T(M;[c_0,\infty)), f_- \in \mathcal{F}_T(M;(-\infty,c_0))\},$$

where $\mathcal{F}_T(M;\mathcal{X})$ consists of functions f such that the approximation error from the second-order Taylor expansion of $f(x)$ about c_0 is bounded by $M|x|^2/2$, and this bound is uniform over $\mathcal{X}$:

$$\mathcal{F}_T(M;\mathcal{X}) = \left\{f : \left|f(x) - f(c_0) - f'(c_0)\,x\right| \leq M|x|^2/2 \text{ all } x \in \mathcal{X}\right\}.$$

The class $\mathcal{F}_T(M;\mathcal{X})$ formalizes the idea that the second derivative of f at zero should be bounded by M. See Section 2 in Armstrong and Kolesar (2017) (note the constant C in that paper equals $C = M/2$ here). This class does not impose smoothness away from the boundary, which may be undesirable in some empirical applications. The Holder class addresses this problem by bounding the second derivative globally. In particular, it assumes that f lies in the class of functions

$$\mathcal{F}_{\text{Holder}}(M) = \{f_+ - f_- : f_+ \in \mathcal{F}_H(M;[c_0,\infty)) \quad \text{and} \quad f_- \in \mathcal{F}_H(M;(-\infty,c_0))\},$$

where

$$\mathcal{F}_H(M;\mathcal{X}) = \{f : |f'(x) - f'(y)| \leq M|x-y|, \quad \forall\, x,y \in \mathcal{X}\}.$$

This smoothness class is specified using the option `sclass`.

We now provide two code examples for CIs around a local linear estimator. Here we assume the bandwidth is equal to 10 on either side of the cutoff. In each example, the parameter space is given by a Holder and Taylor smoothness class, respectively, with $M = 0.1$:

```
CODE:
rdhonest voteshare margin, m(0.1) hp(10) ///
```

```
kernel("uni") se_method("nn") sclass("H") j(3)

OUTPUT:
Call: voteshare is dependent variable; margin is running variable.
Inference by se_method: nn
Estimate: 6.0567735
Maximum Bias: 1.7237683
Std. Error: 1.190527

Confidence intervals:
(2.3747303, 9.7388168), (2.3747627, Inf), (-Inf, 9.7387844)

Bandwidth below cutoff: 10
Bandwidth above cutoff: 10
Number of effective observations: 292.32459


CODE:
rdhonest voteshare margin, m(0.1) hp(10) ///
kernel("uni") se_method("nn") sclass("T") j(3)

OUTPUT:
Call: voteshare is dependent variable; margin is running variable.
Inference by se_method: nn
Estimate: 6.0567735
Maximum Bias: 3.7822376
Std. Error: 1.190527

Confidence intervals:
(.31629332, 11.797254), (.31629332, Inf), (-Inf, 11.797254)

Bandwidth below cutoff: 10
Bandwidth above cutoff: 10
Number of effective observations: 292.32459
```

The confidence intervals use the nearest-neighbor method to estimate the standard error by default (this can be changed using the option `se.method`. This includes "EHW" for Eicker-Huber-White standard errors, "demeaned" for demeaned standard errors, "supplied_var" if a vector of variances is supplied, and "nn" if standard errors are estimated by nearest neighbors.) The package reports two-sided as well as one-sided CIs (with lower and upper limits) by default.

Instead of specifing a bandwidth, one can just specify the smoothness class and smoothness constant M, and the bandwidth will be chosen optimally for a given optimality criterion:

```
CODE:
rdhonest voteshare margin, m(0.1) kernel("tri") ///
opt_criterion("MSE") sclass("H")
```

```
OUTPUT:
Call: voteshare is dependent variable; margin is running variable.
Inference by se_method: nn
Estimate: 5.9366484
Maximum Bias: .83225827
Std. Error: 1.2944207

Confidence intervals:
(2.9548288, 8.918468), (2.9752576, Inf), (-Inf, 8.8980393)

Bandwidth below cutoff: 8.8485087
Bandwidth above cutoff: 8.8485087
Number of effective observations: 213.46289
```

We can also allow for different bandwidths on either side of the cutoff.

```
CODE:
rdhonest voteshare margin, m(0.1) kernel("tri") ///
opt_criterion("FLCI") sclass("H") bw_equal(0)

OUTPUT:
Call: voteshare is dependent variable; margin is running variable.
Inference by se_method: nn
Estimate: 5.9601242
Maximum Bias: .88123277
Std. Error: 1.2764276

Confidence intervals:
(2.9644416, 8.9558068), (2.9793549, Inf), (-Inf, 8.9408935)

Bandwidth below cutoff: 8.8079065
Bandwidth above cutoff: 9.3795503
Number of effective observations: 220.36986
```

Sometimes, when the "opt_criterion("FLCI")" option is used, STATA's optimizer may terminate prematurely, so it is highly recommended that you run the command several times to ensure that the optimal bandwidths have converged to the right values.

3.2 Inference when the running variable is discrete

The confidence intervals described above can also be used when the running variable is discrete, with G support points: when constructing them, we do not assume the running variable is continuous (see Section 5.1 in Kolesar and Rothe (2018) for a more detailed discussion).

As an example, we will use the Oreopoulos (2006) data. In this data set, the running variable is age in years:

```
CODE:
rdhonest logEarn yearat14, m(0.04) c(1947) ///
kernel("uni") opt_criterion("FLCI") sclass("H")

OUTPUT:
Call: logEarn is dependent variable; yearat14 is running variable.
Inference by se_method: nn
Estimate: .07909462
Maximum Bias: .04736585
Std. Error: .06784089

Confidence intervals:
(-.08061322, .23880246), (-.07985958, Inf), (-Inf, .23804881)

Bandwidth below cutoff: 2
Bandwidth above cutoff: 2
Number of effective observations: 2017.0753

CODE:
rdhonest logEarn yearat14, m(0.04) c(1947) ///
kernel("tri") opt_criterion("FLCI") sclass("H")

OUTPUT:
Call: logEarn is dependent variable; yearat14 is running variable.
Inference by se_method: nn
Estimate: .07327069
Maximum Bias: .05947672
Std. Error: .05638951

Confidence intervals:
(-.0790059, .22554728), (-.07895851, Inf), (-Inf, .2254999)

Bandwidth below cutoff: 3.2020732
Bandwidth above cutoff: 3.2020732
Number of effective observations: 2265.8264
```

4 References

Armstrong, Timothy B., and Michal Kolesar. 2018. “Simple and Honest Confidence Intervals in Nonparametric Regression.”

———. 2018. “Optimal Inference in a Class of Regression Models.” *Econometrica* 86: 655-83.

Kolesar, Michal, and Christoph Rothe. 2018. “Inference in Regression Discontinuity Designs with a Discrete Running Variable.” *American Economic Review* 108: 2277-2304.

Lee, David S. 2008. “Randomized Experiments from Non-Random Selection in U.S. House Elections.” *Journal of Econometrics* 142 (2): 675–97.

Oreopoulos, Philip. 2006. "Estimating Average and Local Average Treatment Effects When Compulsory Education Schooling Laws Really Matter." *American Economic Review* 96 (1): 152–75.