Honest inference in Sharp Regression Discontinuity

Michal Kolesár

2017-04-23

Contents

The package `RDHonest` implements confidence intervals for the regression discontinuity parameter considered in Armstrong and Kolesár (2016a), Armstrong and Kolesár (2016b), and Kolesár and Rothe (2016). In this vignette, we demonstrate the implementation of these confidence intervals using datasets from Lee (2008) and Oreopoulos (2006), which are included in the package as a data frame `lee08` and `cghs`.

Sharp RD model

In the sharp regression discontinuity model, we observe units $i = 1, \ldots, n$, with the outcome y_i for the ith unit 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.

The Oreopoulos dataset consists of a subsample of British workers, and it exploits a change in minimum school leaving age in the UK from 14 to 15, which occurred in 1947. The running variable is the year in which the individual turned 14, with the cutoff equal to 1947 so that the "treatment" is being subject to a higher minimum school-leaving age. The outcome is log earnings in 1998.

Plots

The package provides a function `plot_RDscatter` to plot the raw data. To remove some noise, the function plots averages over `avg` number of observations.

```r
library("RDHonest")
## transform data to an RDdata object
dt <- RDData(lee08, cutoff = 0)
## plot 25-bin averages in a window equal to 50 around
## the cutoff, see Figure 1
plot_RDscatter(dt, avg = 25, window = 50, xlab = "Margin of victory",
    ylab = "Vote share in next election")
```

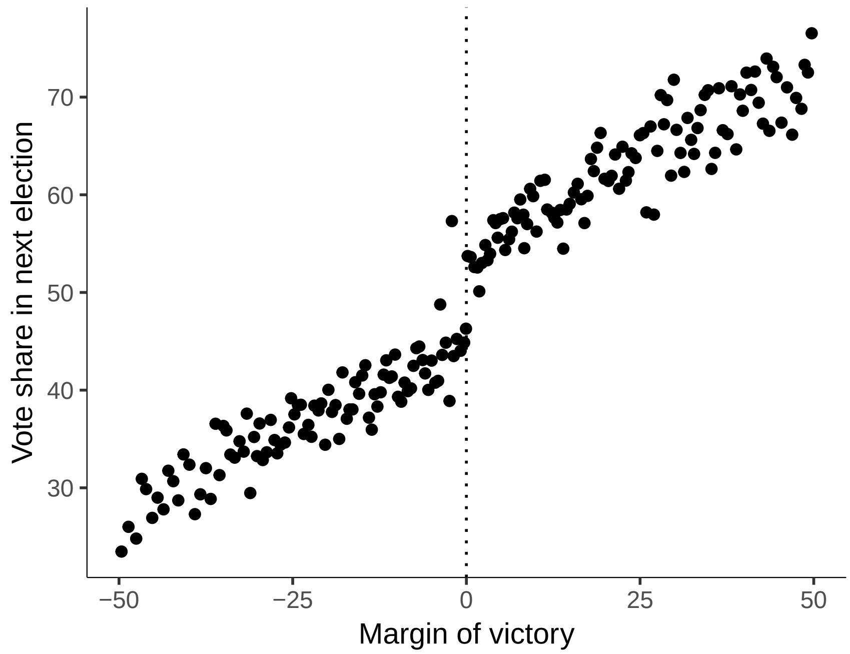

Figure 1: Lee (2008) data

The running variable in the Oreopoulos dataset is discrete. It is therefore natural to plot the average outcome by each value of the running variable, which is achieved using by setting `avg=Inf`. The option `dotsize="count"` makes the size of the points proportional to the number of observations that the point averages over.

```r
## transform data to an RDdata object, and transform
## earnings to log-earnings
do <- RDData(data.frame(logearn = log(cghs$earnings), year14 = cghs$yearat14),
    cutoff = 1947)
## see Figure 2
plot_RDscatter(do, avg = Inf, xlab = "Year aged 14", ylab = "Log earnings",
    propdotsize = TRUE)
```

Inference in Local polynomial regression

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

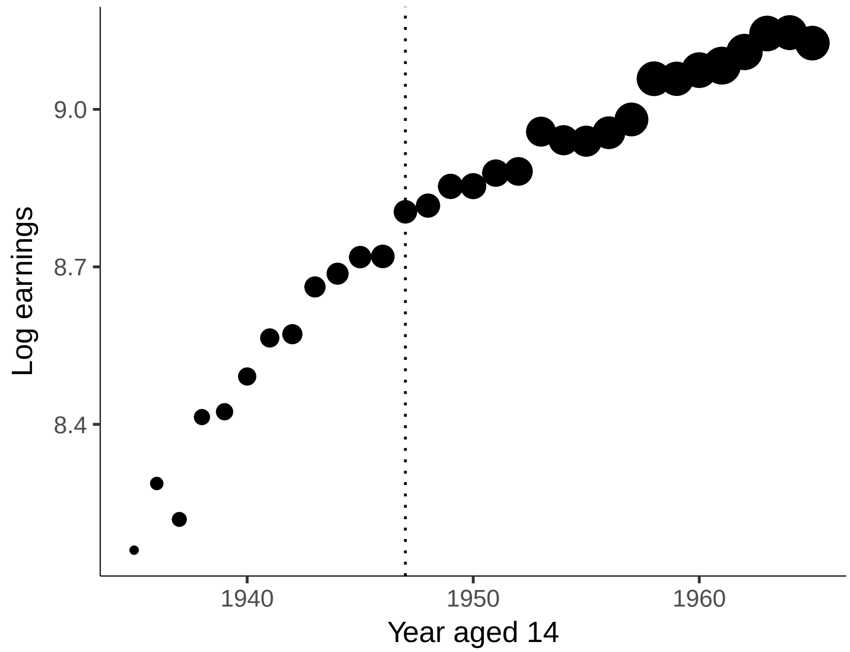

Figure 2: Oreopoulos (2006) data

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

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

denote the worst-case bias over the parameter space $\mathcal{F}$. Then the lower limit of a one-sided CI is given by

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

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). Subtracting the worst-case bias in addition to the usual critical value times standard error ensures correct coverage at all points in the parameter space.

A two-sided CI is given by

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

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

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

By a central limit theorem, the first term on the right-hand side will by distributed standard normal, irrespective of the bias. The second term is bounded in absolute value by $B(\hat{\beta}_{h_+,h_-})/\widehat{se}(\hat{\beta}_{h_+,h_-})$, so that, in large samples, the $1-\alpha$ quantile of the absolute value of the t-statistic will be bounded by $cv_{1-\alpha}(B(\hat{\beta}_{h_+,h_-})/\widehat{se}(\hat{\beta}_{h_+,h_-}))$. The function `CVb` gives these critical values:

```
## Usual critical value
CVb(0, alpha = 0.05)
#>   bias alpha      cv TeXDescription
#> 1    0  0.05 1.95996 $\\alpha=0.05$
## Tabulate critical values for different significance
## levels when bias-sd ratio equals 1/4
knitr::kable(CVb(1/4, alpha = c(0.01, 0.05, 0.1)), caption = "Critical values")
```

Table 1: Critical values

bias	alpha	cv	TeXDescription
0.25	0.01	2.65224	$\alpha = 0.01$
0.25	0.05	2.01971	$\alpha = 0.05$
0.25	0.10	1.69558	$\alpha = 0.1$

The field `TeXDescription` is useful for plotting, or for exporting to LaTeX, as in the table above.

Parameter space

The function `RDHonest` computes honest CIs when the parameter space $\mathcal{F}$ corresponds to a second-order Taylor or second-order Hölder smoothness class, which capture two different types of smoothness restrictions. The second-order Taylor class assumes that f lies in the the class of functions

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

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

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

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 Kolesár (2016a). A disadvantage of this class is that it doesn't impose smoothness away from boundary, which may be undesirable in many empirical applications. The Hölder class addresses this problem directly by bounding the second derivative globally. In particular, it assumes that f lies in the class of functions

$$\mathcal{F}_{\text{Hölder}}(M) = \{f_+ - f_- \colon f_+ \in \mathcal{F}_H(M; [c_0, \infty)),\ f_- \in \mathcal{F}_H(M; (-\infty, c_0))\},$$

where

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

Discrete running variable TODO

The confidence intervals described can also be used when the running variable is discrete, with G support points (see Section 5.1 in Kolesár and Rothe (2016)).

In addition, the package provides function `bmeCI` that calculates honest confidence intervals under the assumption that the specification bias at zero is no worse at the cutoff than away from the cutoff as in Section 5.2 in Kolesár and Rothe (2016). To describe the implementation of these confidence intervals, suppose the point estimate is given by the first element of the regression of the outcome y_i on $m(x_i)$. For instance, local linear regression with uniform kernel and bandwidth h corresponds to $m(x) = I(|x| \leq h) \cdot (I(x > c_0), 1, x, x \cdot I(x > c_0))'$. Let $\theta = Q^{-1}E[m(x_i)y_i]$, where $Q = E[m(x_i)m(x_i)']$, denote the estimand for this regression (treating the bandwidth as fixed), and let $\delta(x) = f(x) - m(x)'\theta$ denote the specification error at x. The RD estimate is given by first element of the least squares estimator $\hat{\theta} = \hat{Q}^{-1}\sum_i m(x_i)y_i$, where $\hat{Q} = \sum_i m(x_i)m(x_i)'$.

Let $w(x_i)$ denote a vector of indicator (dummy) variables for all support points of x_i within distance h of the cutoff, so that $\mu(x_g)$, where x_g is the gth support point of x_i, is given by the gth element of the regression estimand $S^{-1}E[w(x_i)y_i]$, where $S = E[w(x_i)w(x_i)']$. Let $\hat{\mu} = \hat{S}^{-1}\sum_i w(x_i)y_i$, where $\hat{S} = \sum_i w(x_i)w(x_i)'$ denote the least squares estimator. Then an estimate of $(\delta(x_1), \dots, \delta(x_G))'$ is given by $\hat{\delta}$, the vector with elements $\hat{\mu}_g - x_g\hat{\theta}$.

By standard regression results, the asymptotic distribution of $\hat{\theta}$ and $\hat{\mu}$ is given by

$$\sqrt{n}\begin{pmatrix}\hat{\theta} - \theta \\ \hat{\mu} - \mu\end{pmatrix} \xrightarrow{d} N(0, V),$$

where

$$V = \begin{pmatrix} Q^{-1}E[(\epsilon_i^2 + \delta(x_i)^2)m(x_i)m(x_i)']Q^{-1} & Q^{-1}E[\epsilon_i^2 m(x_i)w(x_i)']S^{-1} \\ S^{-1}E[\epsilon_i^2 w(x_i)m(x_i)']Q^{-1} & S^{-1}E[\epsilon_i^2 w(x_i)w(x_i)']S^{-1} \end{pmatrix}.$$

Let $\hat{u}_i$ denote the regression residual from the regression of y_i on $m(x_i)$, and let $\hat{\epsilon}_i$ denote the regression residuals from the regression of y_i on $w(x_i)$. Then a consistent estimator of the asymptotic variance V is given by

$$\hat{V} = n\sum_i V_i V_i', \qquad V_i' = \begin{pmatrix}\hat{u}_i m(x_i)'\hat{Q}^{-1} & \hat{\epsilon}_i w(x_i)'\hat{S}^{-1}\end{pmatrix}.$$

Note that the upper left block and lower right block correspond simply to the Eicker-Huber-White estimators of the asymptotic variance of $\hat{\theta}$ and $\hat{\mu}$. By the delta method, a consistent estimator of the asymptotic variance of $(\hat{\delta}, \hat{\theta}_1)$ is given by

$$\hat{\Omega} = \begin{pmatrix} -X & I \\ e_1' & 0 \end{pmatrix} \hat{V} \begin{pmatrix} -X & I \\ e_1' & 0 \end{pmatrix}',$$

where X is a matrix with gth row equal to x_g', and e_1 is the first unit vector.

In the notation of Kolesár and Rothe (2016), let $\tau_L(\mathbf{g}, \mathbf{s}) = \theta_1 - s^+\delta(x_{g^+}) - s^-\delta(x_{g^-})$, where g^+ and g^- are such that $x_{g^-} \leq c_0 \leq x_{g^+}$, and $s^+, s^- \in \{-1, 1\}$. A left-sided CI for $\tau_L(\mathbf{g}, \mathbf{s})$ is then given by

$$CI_L(\mathbf{g}, \mathbf{s}) = \hat{\theta}_1 - s^+\hat{\delta}(x_{g^+}) - s^-\hat{\delta}(x_{g^-}) - z_{1-\alpha}$$

Optimal inference

For the second-order Taylor smoothness class, the function `RDHonest`, with `kernel="optimal"`, computes finite-sample optimal estimators and confidence intervals, as described in Section 2.2 in Armstrong and Kolesár (2016a).

Empirical examples

Lee data

CIs around a local linear estimator with bandwidth that equals to 10 on either side of the cutoff when the parameter space is given by a Taylor and Hölder smoothness class, respectively, with $M = 0.1$:

```
RDHonest(voteshare ~ margin, data = lee08, kern = "uniform",
    M = 0.1, hp = 10, sclass = "T")
#> Call:
#> RDHonest(formula = voteshare ~ margin, data = lee08, M = 0.1,
#>     kern = "uniform", hp = 10, sclass = "T")
#>
#>
#> Inference by se.method:
#>    Estimate Maximum Bias Std. Error
#> nn  6.05677      3.78224    1.19053
#>
#> Confidence intervals:
#> nn    (0.316293, 11.7973), (0.316293, Inf), (-Inf, 11.7973)
#>
#> Bandwidth below cutoff: 10
#> Bandwidth above cutoff: 10 (Bandwidths are the same)
#> Number of effective observations: 292.325
RDHonest(voteshare ~ margin, data = lee08, kern = "uniform",
    M = 0.1, hp = 10, sclass = "H")
#> Call:
#> RDHonest(formula = voteshare ~ margin, data = lee08, M = 0.1,
#>     kern = "uniform", hp = 10, sclass = "H")
#>
#>
#> Inference by se.method:
#>    Estimate Maximum Bias Std. Error
#> nn  6.05677      1.72377    1.19053
#>
#> Confidence intervals:
#> nn    (2.37473, 9.73882), (2.37476, Inf), (-Inf, 9.73878)
#>
#> Bandwidth below cutoff: 10
#> Bandwidth above cutoff: 10 (Bandwidths are the same)
#> Number of effective observations: 292.325
```

The confidence intervals use the nearest-neighbor method to estimate the standard error by default. The package reports two-sided as well one-sided CIs (with lower as well as upper limit) by default.

CIs around MSE-optimal bandwidth:

```
RDHonest(voteshare ~ margin, data = lee08, kern = "uniform",
    M = 0.1, opt.criterion = "MSE", sclass = "T")
#> Call:
#> RDHonest(formula = voteshare ~ margin, data = lee08, M = 0.1,
#>     kern = "uniform", opt.criterion = "MSE", sclass = "T")
#>
#>
#> Inference by se.method:
```

```
#>    Estimate Maximum Bias Std. Error
#> nn  4.93326     0.995479    1.53039
#>
#> Confidence intervals:
#> nn    (1.39748, 8.46904), (1.42052, Inf), (-Inf,   8.446)
#>
#> Bandwidth below cutoff: 5.05908
#> Bandwidth above cutoff: 5.05908 (Bandwidths are the same)
#> Number of effective observations: 140.767
RDHonest(voteshare ~ margin, data = lee08, kern = "uniform",
    M = 0.1, opt.criterion = "MSE", sclass = "H")
#> Call:
#> RDHonest(formula = voteshare ~ margin, data = lee08, M = 0.1,
#>     kern = "uniform", opt.criterion = "MSE", sclass = "H")
#>
#>
#> Inference by se.method:
#>    Estimate Maximum Bias Std. Error
#> nn  5.88207     0.840511    1.43903
#>
#> Confidence intervals:
#> nn    (2.64206, 9.12208), (2.67456, Inf), (-Inf, 9.08957)
#>
#> Bandwidth below cutoff: 6.89685
#> Bandwidth above cutoff: 6.89685 (Bandwidths are the same)
#> Number of effective observations: 198.086
```

It is possible to compute the MSE-optimal bandwidth directly using the function `RDOptBW`

```
RDOptBW(voteshare ~ margin, data = lee08, kern = "uniform",
    M = 0.1, opt.criterion = "MSE", sclass = "T")
#> Call:
#> RDOptBW(formula = voteshare ~ margin, data = lee08, M = 0.1,
#>     kern = "uniform", opt.criterion = "MSE", sclass = "T")
#>
#>
#> Bandwidth below cutoff:  5.05908
#> Bandwidth above cutoff:  5.05908 (Bandwidths are the same)
RDOptBW(voteshare ~ margin, data = lee08, kern = "uniform",
    M = 0.1, opt.criterion = "MSE", sclass = "H")
#> Call:
#> RDOptBW(formula = voteshare ~ margin, data = lee08, M = 0.1,
#>     kern = "uniform", opt.criterion = "MSE", sclass = "H")
#>
#>
#> Bandwidth below cutoff:  6.89685
#> Bandwidth above cutoff:  6.89685 (Bandwidths are the same)
```

For Taylor smoothness class, instead of using local polynomial estimators, one can use estimators with optimally chosen weights. This typically yields tighter CIs. Comparing the lenghts of two-sided CIs with optimally chosen bandwidths:

```
r <- RDHonest(voteshare ~ margin, data = lee08, kern = "optimal",
    M = 0.1, opt.criterion = "FLCI", se.initial = "Silverman",
    se.method = "nn")
```

```r
2 * r$hl
#> [1] 6.29408
2 * RDHonest(voteshare ~ margin, data = lee08, kern = "triangular",
    M = 0.1, opt.criterion = "FLCI", se.initial = "Silverman",
    se.method = "nn", sclass = "T")$hl
#>      nn
#> 6.64827
```

References

Armstrong, Timothy B., and Michal Kolesár. 2016a. "Optimal Inference in a Class of Regression Models."

———. 2016b. "Simple and Honest Confidence Intervals in Nonparametric Regression."

Kolesár, Michal, and Christoph Rothe. 2016. "Inference in Regression Discontinuity Designs with a Discrete Running Variable."

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.