USC, Fall 2016, Economics 513

Lecture 13: Nonparametric regression and regression discontinuity II

Instruments and forcing variables

We are interested in the effect of an intervention on an outcome variable Y. The outcomes are

$$\begin{aligned} Y_{0i} &= \text{ if } i \text{ is treated} \\ Y_{1i} &= \text{ if } i \text{ is not treated} \end{aligned}$$

The treatment indicator is D_i.

We have seen that if the treatment is assigned randomly then

$$\mathrm{E}(Y|D=1) - \mathrm{E}(Y|D=0) = \mathrm{E}(Y_1|D=1) - \mathrm{E}(Y_0|D=0) = \mathrm{E}(Y_1 - Y_0) = \mathrm{ATE}$$

and if the dif-in-diff assumption

$$\mathrm{E}(Y_{01} - Y_{00}|D=1) = \mathrm{E}(Y_{01} - Y_{00}|D=0)$$

holds, then

$$\mathrm{E}(Y_1|D=1) - \mathrm{E}(Y_1|D=0) - (\mathrm{E}(Y_0|D=1) - \mathrm{E}(Y_0|D=0)) =$$

$$\mathrm{E}(Y_{11} - Y_{01}|D=1) = \mathrm{ATET}$$

With random assignment the ATE can be estimated using the regression

$$Y_i = \alpha + \beta D_i + \varepsilon_i$$

The OLS estimator for β estimates the ATE.

If we have have an instrumental variable Z then we can estimate β by IV. Such a variable Z is correlated with D but has no direct effect on Y. It should be noted that if the treatment effect is not the same for all i then the IV estimator does not estimate the ATE. It estimates the Local Average Treatment Effect (LATE), i.e. the treatment effect for the subpopulation that changes treatment status by the change in Z. For instance, if D is the indicator of a college education and Z is the distance to the nearest four year college, then a decrease in Z makes it more likely that i completes college education. The LATE is the treatment effect for those i who would attend college if they would live closer to college. The individuals for who distance is a deciding factor is probably small, so that it may be a special subpopulation.

Now assume that Z does affect whether a unit is treated, i.e. Z affects D, but Z also has a direct effect on the outcome. Such a Z is not a valid instrument.

The effect of Z on D may have a special form. Consider the examples

1. Healthcare is provided to children below a threshold age under SCHIP (in many states)and to older individuals above a threshold age under MEDICARE.

2. Certain rules or laws apply only to firms that have a number of employees that exceeds a threshold.

3. In a two party election the party that has the candidate who gets more than 50% of the vote is elected.

In all these examples there is a variable Z, age in 1., number of employees in 2., and vote share in 3. that has a discontinuous effect on D, the provision of health care by the government in 1., being subject to the rule or law in 2. , and being elected in 3. In 2. and 3. the treatment dummy D satisfies

$$D = I(Z \geq c)$$

In all examples Z is likely to have a direct effect on the outcome variables, e.g. some measure of health in example 1.

The idea of the Regression Discontinuity (RD) estimator is that units just below and above the threshold are comparable, i.e. have similar characteristics, except for their treatment status. A comparison of the outcome just below and above c would show the effect of the treatment. In particular, if the intervention has an effect, we expect to find a discontinuity in the relation between the outcome and Z.

Example

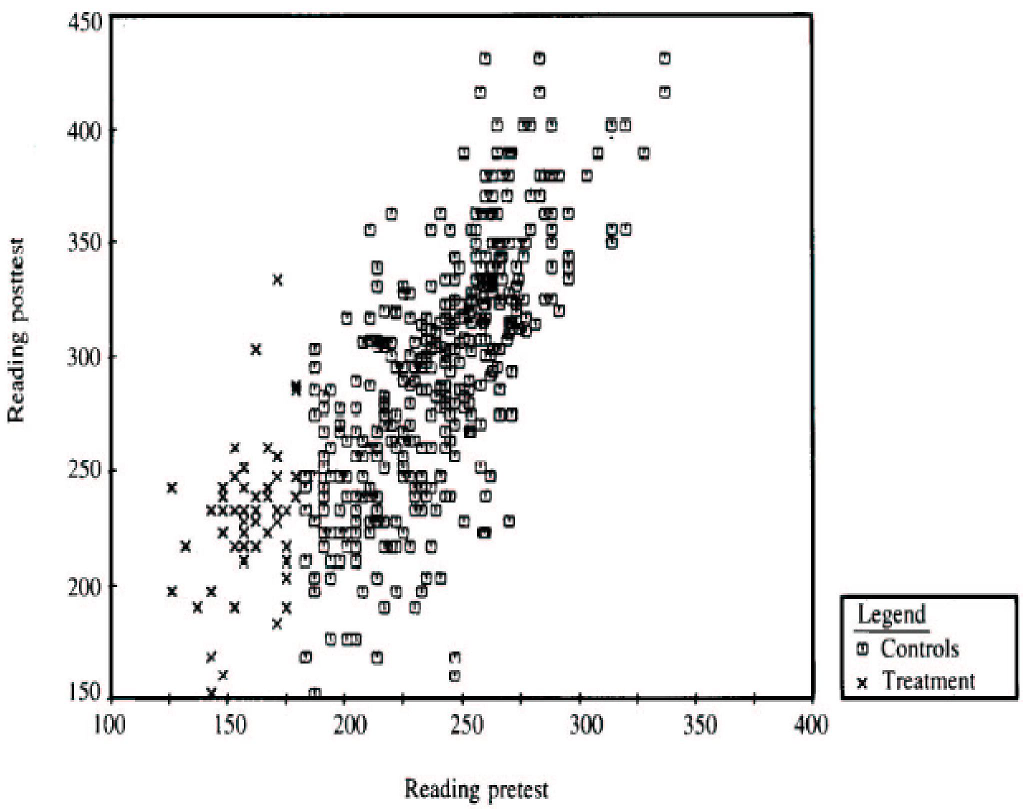
Compensatory education program in Providence on basis of test outcome: See figure.

- Compensatory program if pre-program score is below 179.
- Outcome variable is test score after program
- Figure 1 has the pre- and post-program test scores.
- Notation:
  - $X - 179$ is pre-program score minus cut-off.
  - $Y - 179$ is post-program score minus cut-off.
- Model is

$$Y - 179 = \gamma_0 + \gamma_1(X - 179) + \beta I(X < 179) + u$$

- OLS estimate of β is 44.61 (s.e. 7.50).

Figure 9. Example analysis: Bivariate distribution - Providence second grade reading



- $(X - 179)I(X < 179)$ not significant.

The Regression Discontinuity estimator

We assume that

$$D = I(Z \geq c)$$

If units just below c are comparable, i.e. have the same characteristics except for the intervention, to units just above c, then we have for small h

$$\mathrm{E}(Y_0|c - h \leq Z < c) \approx \mathrm{E}(Y_0|c \leq Z \leq c + h)$$

and

$$\mathrm{E}(Y_1|c - h \leq Z < c) \approx \mathrm{E}(Y_1|c \leq Z \leq c + h)$$

What we have in mind is illustrated in the figure.

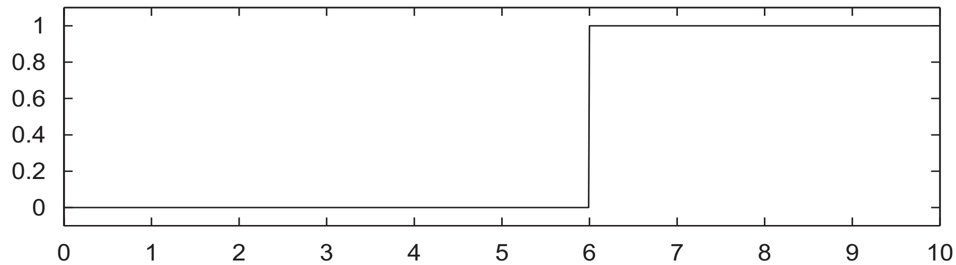


Fig. 1. Assignment probabilities (SRD).

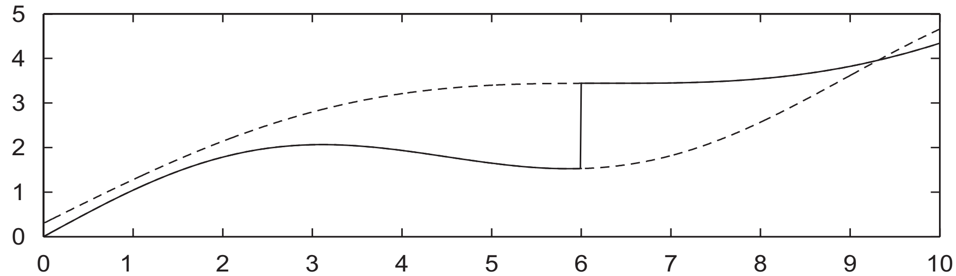


Fig. 2. Potential and observed outcome regression functions.

Formally we assume

$$\mathrm{E}(Y_0|Z=z)$$

and

$$\mathrm{E}(Y_1|Z=z)$$

are continuous in z. We only need these functions to be continuous in c, but discontinuities at other points would be suspicious.

With this assumption

$$\mathrm{E}(Y_0|Z=c) = \lim_{z\uparrow c} \mathrm{E}(Y_0|Z=z) = \lim_{z\uparrow c} \mathrm{E}(Y_0|Z=z, D=0) = \lim_{z\uparrow c} \mathrm{E}(Y|Z=z)$$

and

$$\mathrm{E}(Y_1|Z=c) = \lim_{z\downarrow c} \mathrm{E}(Y_1|Z=z) = \lim_{z\downarrow c} \mathrm{E}(Y_1|Z=z, D=1) = \lim_{z\downarrow c} \mathrm{E}(Y|Z=z)$$

The RD effect of the intervention is

$$\tau = \mathrm{E}(Y_1|Z=c) - \mathrm{E}(Y_0|Z=c)$$

i.e. the ATE at $Z = c$, and in the population

$$\tau = \lim_{z \downarrow c} \mathrm{E}(Y|Z = z) - \lim_{z \uparrow c} \mathrm{E}(Y|Z = z)$$

This is an average treatment effect but as with the case that Z is a valid instrument it is an effect for a subpopulation: those units who change treatment status by a change in Z, i.e. the units with $Z = c$. This is a small subpopulation that hopefully is representative of the whole population. Note that in moving from randomized assignment to RD via IV, we obtain treatment effects for smaller subpopulations.

Estimating the RD treatment effect

The first step is to graphical: if there is a treatment effect at c the average outcome should be discontinuous at c.

The figure is from David Lee, "The Electoral Advantage of Incumbency and Voters' Valuation of Politicians' Experience: A Regression Discontinuity Analysis of Elections to the U.S. House", Journal of Econometrics. It plots the probability of a candidate winning the election for a seat in the US Congress in year $t + 1$ as a function of the fraction of votes in year t, Z. The intervention is being elected and in a two-party election

$$D = I(Z > .5)$$

The ATE at .5 is interpreted as the incumbency effect. The idea is that districts and candidates who have a fraction of the vote just above/below .5 are comparable.

As a check the bottom panel plots the number of elections won by the same candidate in earlier years and there is no apparent discontinuity here as one would expect if there is an incumbency effect.

A kernel regression estimator of the ATE at $Z = c$ is

$$\hat{\tau} = \hat{\mu}_n, r(c) - \hat{\mu}_n, l(c)$$

with

$$\hat{\mu}_n, r(c) = \frac{\sum_{i=1}^n I(Z_i \geq c) Y_i K\left(\frac{Z_i - c}{h}\right)}{\sum_{i=1}^n I(Z_i \geq c) K\left(\frac{Z_i - c}{h}\right)}$$

and

$$\hat{\mu}_n, l(c) = \frac{\sum_{i=1}^n I(Z_i \leq c) Y_i K\left(\frac{Z_i - c}{h}\right)}{\sum_{i=1}^n I(Z_i \leq c) K\left(\frac{Z_i - c}{h}\right)}$$

With a uniform kernel on $[-1, 1]$ we have

$$\hat{\mu}_n, r(c) = \frac{\sum_{i=1}^n I(c \leq Z_i \leq c + h) Y_i}{\sum_{i=1}^n I(c \leq Z_i \leq c + h)}$$

Figure IIa: Candidate's Probability of Winning Election t+1, by Margin of Victory in Election t: local averages and parametric fit

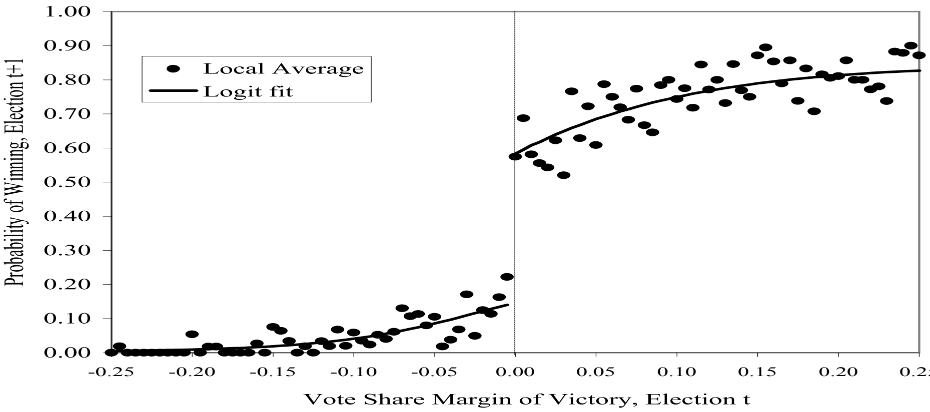
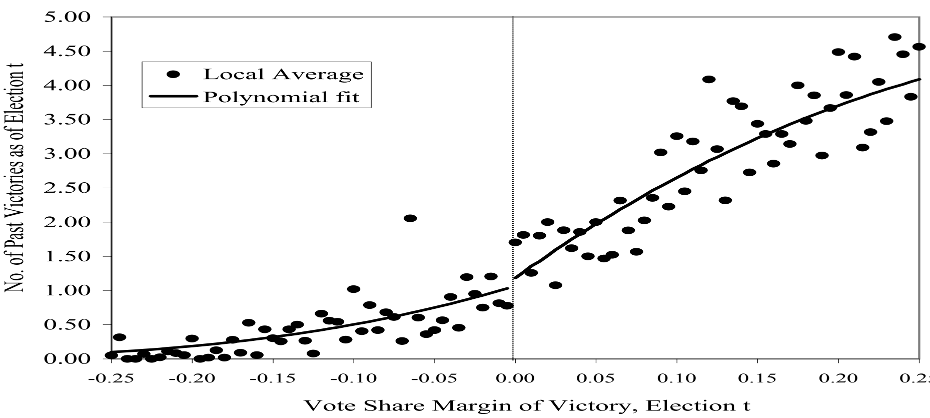


Figure IIb: Candidate's Accumulated Number of Past Election Victories, by Margin of Victory in Election t: local averages and parametric fit



and

$$\hat{\mu}_n, l(c) = \frac{\sum_{i=1}^n I(c-h \le Z_i \le c) Y_i}{\sum_{i=1}^n I(c-h \le Z_i \le c)}$$

For fixed h we have

$$\hat{\mu}_n, r(c) \xrightarrow{p} \frac{\int_c^{c+h} \mu(z) f(z) \mathrm{d}z}{\int_c^{c+h} f(z) \mathrm{d}z}$$

We have

$$\mu(z) \approx \mu(c) + \mu'(c)h + O(h^2)$$

so that

$$\frac{\int_c^{c+h} \mu(z) f(z) \mathrm{d}z}{\int_c^{c+h} f(z) \mathrm{d}z} - \mu(c+h) = \frac{\int_c^{c+h} \left(\mu(z) - \mu(c+h)\right) f(z) \mathrm{d}z}{\int_c^{c+h} f(z) \mathrm{d}z} \approx$$

$$\frac{\int_c^{c+h} \left(\mu(c) + \mu'(c)h - \mu(c+h)\right) f(z)\mathrm{d}z}{\int_c^{c+h} f(z)\mathrm{d}z} = O(h^2)$$

We conclude

$$\frac{\int_c^{c+h} \mu(z)f(z)\mathrm{d}z}{\int_c^{c+h} f(z)\mathrm{d}z} = \mu(c+h) + O(h^2) = \mu(c) + \mu'(c)h + O(h^2)$$

The bias at c goes to 0 at a slower rate than at points $z > c$. The reason is that we average one sided, i.e. for $\mu_r(c)$ observations with $Z_i < c$ are not relevant.

For that reason we use a local linear regression estimator that does not have this problem. We solve (we take the same uniform kernel)

$$\min_{\alpha_l, \beta_l} \sum_{i=1}^{n} I(c - h \leq X_i < c)(Y_i - \alpha_l - \beta_l(X_i - c))^2$$

and

$$\min_{\alpha_r, \beta_r} \sum_{i=1}^{n} I(c - h \leq X_i < c)(Y_i - \alpha_r - \beta_r(X_i - c))^2$$

The value of the regression functions at c is estimated as

$$\hat{\mu}_l(c) = \hat{\alpha}_l + \hat{\beta}_l(c - c) = \hat{\alpha}_l$$

and

$$\hat{\mu}_r(c) = \hat{\alpha}_r + \hat{\beta}_r(c - c) = \hat{\alpha}_r$$

The local linear RD estimator is

$$\hat{\tau} = \hat{\alpha}_r - \hat{\alpha}_l$$

The choice of h can be made by cross-validation (although that is tricky here).

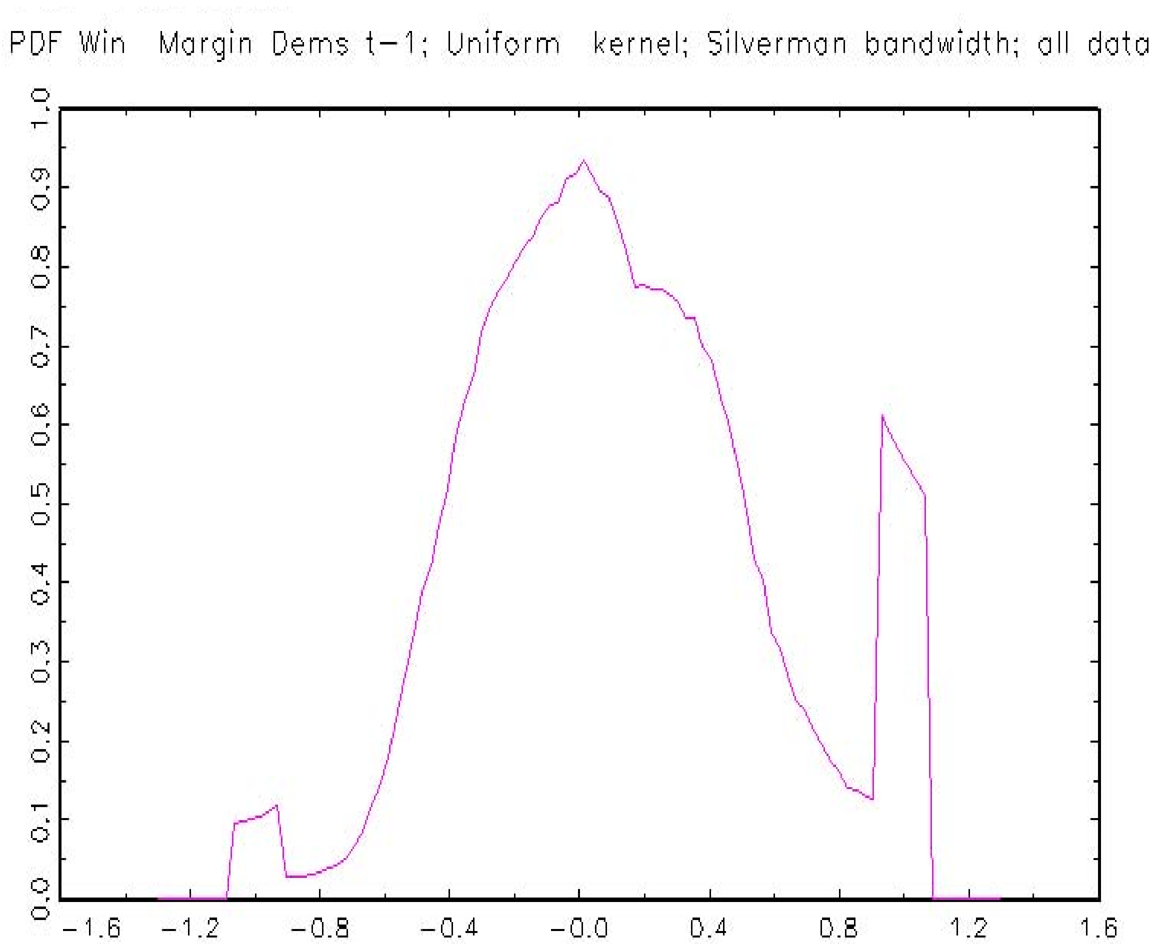
Application

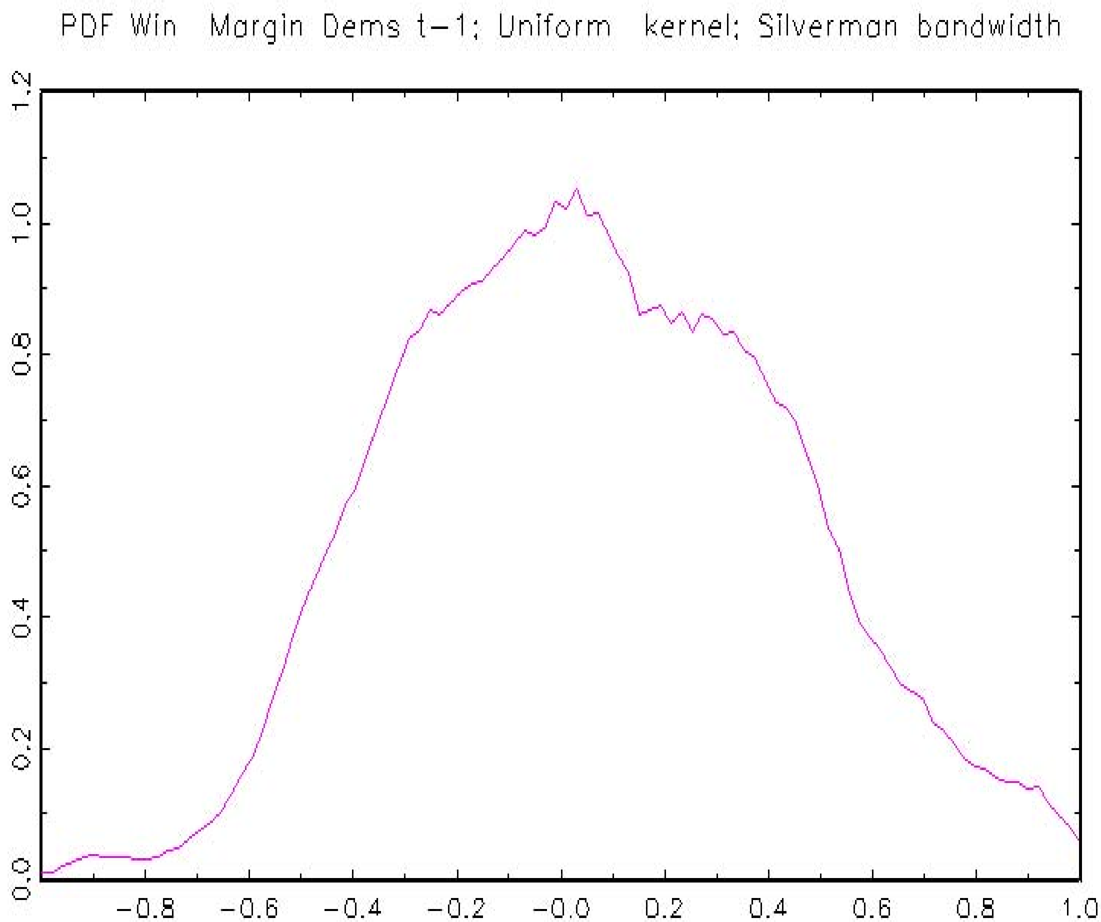
Lee uses 6558 observation on congressional elections to estimate the effect of incumbency on the probability of winning as the discontinuity of the regression of the margin of victory of the candidate of the Democrats in election $t-1$ on the candidate of that party winning in year t.

The estimates are for kernel regression with a uniform kernel and the bandwidth $h = 1.06\hat{\sigma}_z n^{-1/5}$ that is optimal if we estimate a normal density for z. I also consider bandwidths half and two times as large. Using the same bandwidths I also report local linear estimates (with uniform kernel). Standard errors are in parentheses.

	Uniform kernel	Local linear
$h_r = .0221, h_l = .0160$	.474 (.061)	.604 (.121)
$h_r = .0443, h_l = .0320$	.427 (.043)	.525 (.089)
$h_r = .0887, h_l = .0640$	.529 (.028)	.388 (.057)

As a check I show kernel density estimates of the density of z. There is no indication of an irregularity at $z = 0$, e.g. due to change in voter behavior.




PDF Win Margin Dems t-1; Uniform kernel; Silverman bandwidth
0.0
0.2
0.4
0.6
0.8
1.0
1.2
-0.8
-0.6
-0.4
-0.2
-0.0
0.2
0.4
0.6
0.8
1.0