Problem Set #8 Sarah Bass

18.2.a This is a regression of Y on D with state fixed effects. So

$$\hat{\theta} = \frac{\sum_{t=0}^{T} \sum_{i=0}^{N} (D_{it} - \bar{D}_i)(Y_{it} - \bar{Y})}{\sum_{t=0}^{T} \sum_{i=0}^{N} (D_{it} - \bar{D}_i)^2}$$

* where $\bar{D}_i$ and $\bar{Y}_i$ are sample averages over time of D_{it} and Y_{it}, respectively.

18.2.b For the untreated sub-sample, $D_{0t} = 0$, so $\bar{D}_0 = 0$ and for the treated sub-sample $D_{10} = 0$, $D_{11} = 1$, $\bar{D}_1 = 0.5$

$$\hat{\theta} = \frac{(D_{00} - \bar{D}_0)(Y_{00} - \bar{Y}_0) + (D_{01} - \bar{D}_0)(Y_{01} - \bar{Y}_0) + (D_{10} - \bar{D}_1)(Y_{10} - \bar{Y}_1) + (D_{11} - \bar{D}_1)(Y_{11} - \bar{Y}_1)}{(D_{00} - \bar{D}_0)^2 + (D_{01} - \bar{D}_0)^2 + (D_{10} - \bar{D}_1)^2 + (D_{11} - \bar{D}_1)^2}$$

$$= \frac{(D_{10} - \bar{D}_1)(Y_{10} - \bar{Y}_1) + (D_{11} - \bar{D}_1)(Y_{11} - \bar{Y}_1)}{(D_{10} - \bar{D}_1)^2 + (D_{11} - \bar{D}_1)^2}$$

$$= \frac{(-0.5)(Y_{10} - \bar{Y}_1) + (0.5)(Y_{11} - \bar{Y}_1)}{(0.25) + (0.25)}$$

$$= Y_{11} - Y_{10}$$

18.2.c. No, since $\hat{\theta}$ doesn't depend on the untreated sub-sample, this is only a difference estimator.

18.2.d. If there is no time trend, e.g. $E[Y_{00}] = E[Y_{01}]$.

18.4. If all dummy variables were included, there would be perfect multicollinearity. One needs to be ommitted as a reference group.

18.5.a.

	WI	MN	
Pre	15.23	16.42	1.19
Post	16.72	18.10	1.38
Diff	+1.49 -	+1.68	= -0.19

18.5.b. -0.19

18.5.c. - 1.19

17.1.a. $E[X^*] = \int x \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) dx$

$= \frac{1}{nh} \int x \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) dx$

Let $y = \left(\frac{x_i - x}{h}\right)$, $h\,dy = dx$, $x = x_i - hy$

$= \frac{1}{n} \sum_{i=1}^{n} \int (x_i - yh) \, K(y) \, dy$

$= \frac{-h}{n} \sum_{i=1}^{n} \int y K(y) dy + \frac{1}{n} \sum_{i=1}^{n} \int x_i K(y) \, dy$

$= \frac{1}{n} \sum_{i=1}^{n} x_i \int K(y) \, dy$

$= \frac{1}{n} \sum_{i=1}^{n} x_i$

$= \bar{X}_n$

17.1.6. $\mathrm{Var}(X^*) = E[X^{*2}] - E[X^*]^2$

$$= \int x^2 \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) dx - \bar{X}_n^2$$

$$= \frac{1}{nh} \int x^2 \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) dx - \bar{X}_n^2$$

Let $y = \left(\frac{X_i - x}{h}\right)$, $h\,dy = dx$, $x = X_i - yh$

$$= \frac{1}{n} \sum_{i=1}^{n} \int (X_i - yh)^2 K(y)\, dy - \bar{X}_n^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \int (X_i^2 - 2X_i yh + y^2h^2) K(y)\, dy - \bar{X}_n^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \int X_i^2 K(y) dy - \frac{2h}{n} \sum_{i=1}^{n} \int X_i y K(y) dy + \frac{h^2}{n} \sum_{i=1}^{n} \int y^2 K(y) dy - \bar{X}_n^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} X_i^2 + h^2 - \bar{X}_n^2$$

$$= \hat{\sigma}^2 + h^2$$

17.3 The optimal bandwidth is $h_0 = \left(\frac{R_K}{R(f'')}\right)^{1/5} n^{-1/5}$.

Since the distribution is uniform, $f''=0$, so the optimal bandwidth is whatever is the largest feasible.

17.4 Using the same bandwidth will result in a much wider, flatter plot than is appropriate since the scaling changed. The bandwidth should be divided by 1,000,000 to get the same shape density plot.

19.3. When $m(x)$ is increasing and concave → neg. bias.
increasing and convex → pos. bias.
decreasing and concave → neg. bias
decreasing and convex → pos. bias

This is because you're averaging around the point of interest.

19.4. Bias is:

$$B(x) = \frac{1}{2} m''(x) + m'(x) \frac{f'(x)}{f(x)}$$

$$= \frac{\beta f'(x)}{f(x)}$$

If $\beta > 0$, then $B(x) > 0$ if $f'(x) > 0$ and $B(x) < 0$ if $f'(x) < 0$.

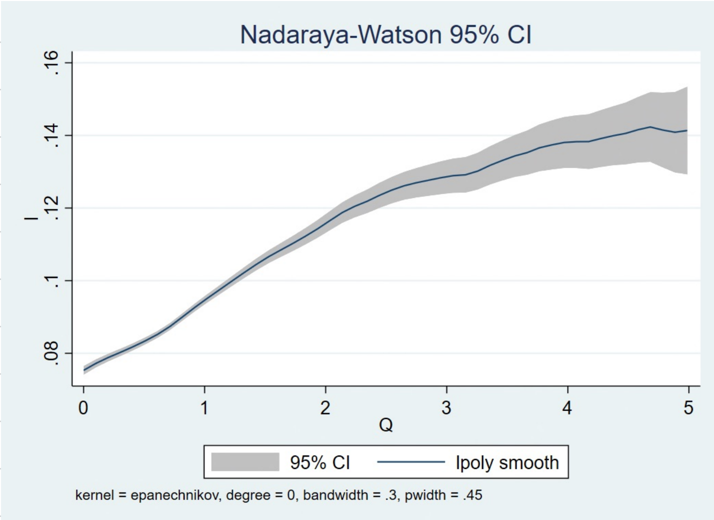
If $\beta < 0$, then $B(x) > 0$ if $f'(x) < 0$ and $B(x) < 0$ if $f'(x) > 0$.

$f(x)$ is the marginal density of X. For $\beta > 0$, if $f'(x) > 0$, then the mass of x is concentrated to the right of x, so the bias is positive. If $f'(x) < 0$, then the mass of x is concentrated to the left of x, so the bias is negative.

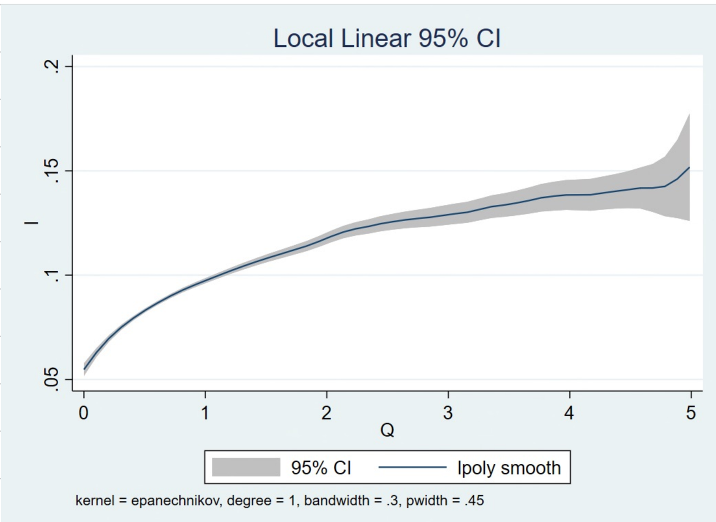
For $\beta < 0$, if $f'(x) > 0$, then the mass of x is concentrated to the left of x → negative bias. If $f'(x) < 0$, the mass is concentrated to the right → positive bias.

In short, if the mass is to the left of x, there's negative bias. If the mass is concentrated to the right, there is positive bias.

19.9.a



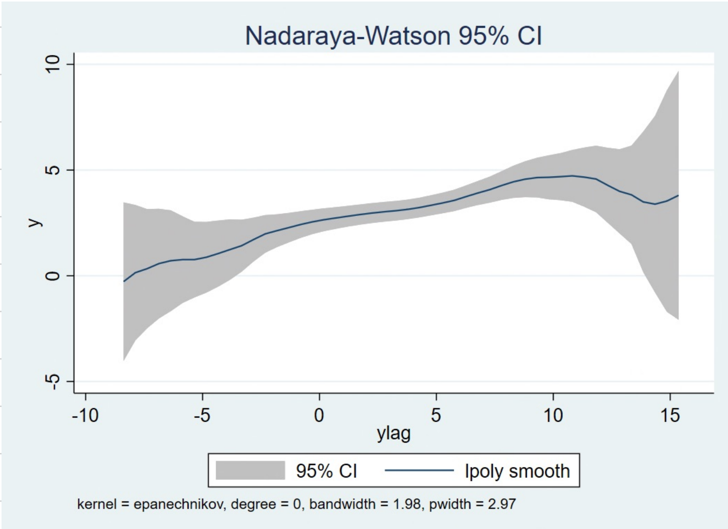
19.9.b.



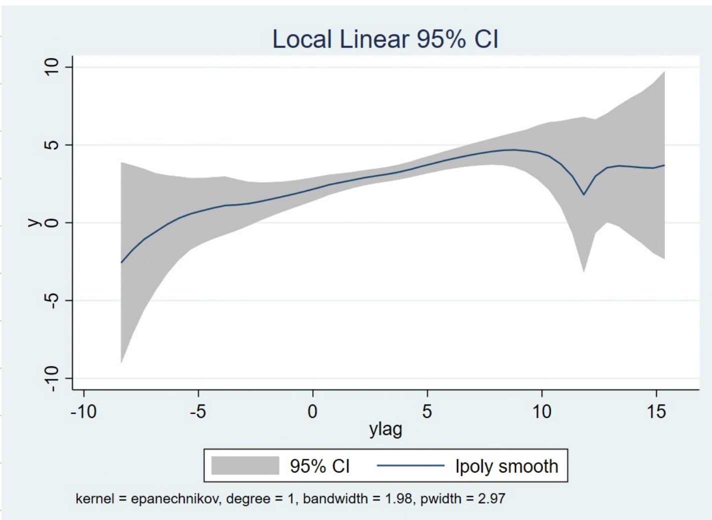
19.9.c. It appears that there is non-linearity in the relationship.

19.11.a. Done in stata.

19.11.b.



19.11.c.



19.11.d It appears that there is non-linearity in the relationship.