# Problem Set #8

Danny Edgel
Econ 710: Economic Statistics and Econometrics II
Spring 2021

March 25, 2021

*Discussed and/or compared answers with Sarah Bass, Emily Case, Katherine Kwok, Michael Nattinger, and Alex Von Hafften*

## Exercise 18.2

(a) Since there are four observations with four possible outcomes, each outcome is determined in this model. Let $i$ and $t$ each take on values of 0 and 1, with $State_1 = 1$ and $Time_1 = 1$. Then we can remove time and state fixed effects from the model by taking averages across the state variable:

$$Y_{it} - \overline{Y}_i = \theta(D_{it} - \overline{D}_i) + \varepsilon_{it} - \overline{\varepsilon}_i$$

Then this model can be estimated with OLS, where

$$\hat{\theta} = \frac{\sum_{i=0}^{1} \sum_{t=0}^{1} (Y_{it} - \overline{Y}_i)(D_{it} - \overline{D}_i)}{\sum_{i=0}^{1} \sum_{t=0}^{1} (D_{it} - \overline{D}_i)^2}$$

(b) Recall how our $Time$ and $State$ are defined. The sample has only four observations, with $\{it\} \in \{\{00\}, \{10\}, \{01\}, \{11\}\}$. Then,

$$D_{0t} = State_0 Time_t = 0 Time_t = 0 \; \forall t$$

Furthermore,

$$\sum_{t=0}^{1}(D_{1t} - \overline{D}_i)^2 = \sum_{t=0}^{1}\left(State_1 Time_t - \frac{1}{2}\sum_{t=0}^{1} State_1 Time_t\right)^2$$

$$= \left(1*0 - \frac{1}{2}\right)^2 + \left(1*1 - \frac{1}{2}\right)^2 = \frac{1}{2}$$

Thus, our estimate for $\hat{\theta}$ simplifies to:

$$\hat{\theta} = \frac{(Y_{10} - \overline{Y}_1)\left(-\frac{1}{2}\right) + (Y_{11} - \overline{Y}_1)\left(\frac{1}{2}\right)}{\frac{1}{2}} = Y_{11} - Y_{10}$$

(c) No. This is a single difference estimator.

(d) $\hat{\theta}$ would be an appropriate estimator of the treatment effect if there is no omitted variable, such as a time trend, causing a change in $Y_{1t}$ from $t = 0$ to $t = 1$.

## Exercise 18.4

If $N_2$ interaction dummies are included, then no observation will be used as the omitted group in the fixed effects estimation, leading to multicollinearity. The regression will fail because the independent variable matrix will not be invertible. This is also the reason that only $N_1 - 1$ interaction dummy variables are included in the regression test for equal control effects.

## Exercise 18.5

(a) The table below displays how the economist would calculate a difference in difference estimate for this data, with her point estimate in bold.

|  | Wisconsin | Minnesota | Difference |
| --- | --- | --- | --- |
| Before | 15.23 | 16.42 | -1.19 |
| After | 16.72 | 18.10 | -1.38 |
| Difference | 1.49 | 1.68 | **-0.19** |

Thus, her point estimate is -0.19.

(b) Since the economist does not add any fixed effects, her point estimate of the difference-in-difference is the same, and $\beta$ is the value for Minnesota in the "After" period. Thus, $\hat{\beta} = 18.10$.

(c) $\gamma$ represents the value for Wisconsin in the "Before" period. Thus, $\hat{\gamma} = 15.23$.

# Exercise 17.1

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

(a)

$$\mathbb{E}\left[X^*\right] = \int x \hat{f}(x) dx = \int \frac{x}{nh} \sum_{i}^{n} K\left(\frac{x - X_i}{h}\right) dx$$

$$= \int \left(\frac{X_i + uh}{n}\right) \sum_{i}^{n} K(u) du$$

$$= \frac{1}{n} \sum_{i}^{n} X_i \int K(u) du + \int \left(\frac{1}{n}\right) \sum_{i}^{n} huK(u) du$$

$$= \overline{X}_n$$

(b)

$$Var(X^*) = \mathbb{E}\left[(X^*)^2\right] - \mathbb{E}\left[X^*\right]^2 = \mathbb{E}\left[(X^*)^2\right] - \overline{X}_n^2$$

$$\mathbb{E}\left[(X^*)^2\right] = \int \frac{x^2}{nh} \sum_{i}^{n} K\left(\frac{x - X_i}{h}\right) dx$$

$$= \frac{1}{n} \sum_{i}^{n} \int (X_i + uh)^2 K(u) du$$

$$= \frac{1}{n} \sum_{i}^{n} X_i^2 + 2\frac{1}{n} \sum_{i}^{n} X_i h \int uK(u) du + h^2 \int u^2 K(u) du$$

$$= \frac{1}{n} \sum_{i}^{n} X_i^2 + h^2$$

$$\Rightarrow Var(X^*) = \frac{1}{n} \sum_{i}^{n} X_i^2 - \overline{X}_n^2 + h^2 = \hat{\sigma}^2 + h^2$$

# Exercise 17.3

(17.11) provides the following optimal bandwidth equation (and definitions for secondary values):

$$h_0 = \left(\frac{R_K}{R(f'')}\right)^{1/5} n^{-1/5}$$

$$R_K = \int K(u)^2 du$$

$$R(f'') = \int \left(f''(x)\right)^2 dx$$

Uniform density on $[0, 1]$ has the pdf $f(x) = 1$, which results in $R(f'') = 0$ and an optimal bandwidth approaching positive infinity. $h_0$ is derived as the bandwidth that optimally trades off the variance gains of having a higher bandwidth against the increased bias of doing so. On a uniform density, there is no bias from increasing bandwidth, so this formula recommends minimizing variance by using the entire range of $X$ as a single bin.

## Exercise 17.4

The bandwidth itself will not change because it is given as a percentile value of the range of $X$. However, the variance of the distribution in the second estimation is a substantially smaller number, so a lower bandwidth will be necessary to obtain an optimal trade-off between variance and bias. Using the same bandwidth for each estimation will lead to a large bias in the estimation that uses millions of dollars.

## Exercise 19.3

## Exercise 19.4

The asymptotic bias of Nadaraya-Watson is $h^2 B_{nw}(x)$, where

$$B_{nw}(x) = \frac{1}{2}m''(x) + f(x)^{-1}f'(x)m'(x)$$

Thus, in this case,

$$B_{nw}(x) = \frac{1}{2} + \frac{f'(x)\beta}{\alpha + \beta x}$$

Where $f(x)$ is the marginal density of $x$.

## Exercise 19.9

## Exercise 19.11