# Applied Microeconometrics Problem Set 1

Yixin Sun

October 20, 2020

## Problem 1

**Set Up:**

- $Y(d) - Y(0) = \alpha$
- $Y = DY(d) + (1 - D)Y(0) = \tau D + Y(0)$
- Regression 1: $Y = a_1 + \beta D + \epsilon_1$
- Regression 2: $Y = a_2 + \gamma D + \psi X + \epsilon_2$

### a)

Let $\hat{D} = D - \bar{D}$, and solve

$$\begin{aligned}
\beta &= \frac{cov(\hat{D}), Y)}{var(\hat{D})} \\
&= \frac{cov(\hat{D}, \tau D + Y(0))}{var(\hat{D})} \\
&= \frac{cov(\hat{D}, \tau D)}{var(\hat{D})} + \frac{cov(\hat{D}, Y(0))}{var(\hat{D})} \\
&= \tau + \frac{cov(D, Y(0))}{var(D)}
\end{aligned}$$

Where the last line uses the fact that $var(D - \bar{D}) = var(D)$ and $cov(D - \bar{D}, Y(0)) = cov(D, Y(0))$

Now let's solve for $\gamma$, using $\tilde{D} = D - \pi_0 - \pi_1 X$,

$$\begin{aligned}
\gamma &= \frac{cov(\tilde{D}, Y)}{var(\tilde{D})} \\
&= \frac{cov(\tilde{D}, \tau D + Y(0))}{var(\tilde{D})} \\
&= \frac{cov(\tilde{D}, \tau D) + cov(\tilde{D}, Y(0))}{var(\tilde{D})} \\
&= \tau + \frac{cov(\tilde{D}, Y(0))}{var(\tilde{D})} \\
&= \tau + \frac{cov(D - \pi_0 - \pi_1 X, Y(0)}{var(\tilde{D})} \\
&= \tau + \frac{cov(D - \pi_1 X, Y(0))}{var(\tilde{D})}
\end{aligned}$$

From this, we see that

$$|\alpha - \gamma| = \Big|\frac{cov(D - \pi_1 X, Y(0))}{var(\tilde{D})}\Big| \tag{1}$$

$$|\alpha - \beta| = \Big|\frac{cov(D, Y(0))}{var(D)}\Big| \tag{2}$$

and it is ambiguous which term is bigger, depending on the covariance between $X$ and $Y(0)$.

## b)

If $D$ and $X$ are uncorrelated, then $\pi_1 = 0$ and (1) would simplify to

$$\begin{aligned} |\alpha - \gamma| &= \Big|\frac{cov(D, Y(0))}{var(D)}\Big| \\ &= |\alpha - \beta| \end{aligned}$$

so the two would be equal.

## c)

If $X$ is uncorrelated with $Y(0)$ and $Y(d)$, then we have

$$\alpha - \gamma = \frac{cov(D, Y(0))}{var(\tilde{D})}$$

Expanding out the denominator,

$$\begin{aligned} var(\tilde{D}) &= var(D - \pi_0 - \pi_1 X) \\ &= var(D) + \pi_1 var(X) - 2\pi_1 cov(D, X) \end{aligned}$$

So whether $|\alpha - \gamma|$ is bigger or smaller than $|\alpha - \beta|$ is dependent on whether $\pi_1 var(X) - 2\pi_1 cov(D, X)$ is greater or less than zero.

## d)

If $E[Y(0) \mid D = d, X = x] = E[Y(0) \mid X = x]$, then

$$\begin{aligned} \alpha - \gamma &= \frac{cov(D - \pi_1 X, Y(0))}{var(\tilde{D})} \\ &= \frac{cov(\pi_1 X, Y(0))}{var(\tilde{D})} \end{aligned}$$

which is still ambiguous when compared to $\alpha - \beta = \frac{cov(D, Y(0))}{var(D)}$

### e)

Let $Y(0) = \delta_0 + \delta_1 X$ and plug in,

$$\begin{aligned}
\alpha - \gamma &= \frac{cov(D - \pi_1 X, \delta_0 + \delta_1 X)}{var(\tilde{D})} \\
&= \frac{cov(D, \delta_1 X) - cov(\pi_1 X, \delta_1 X)}{var(\tilde{D})} \\
&= \frac{\delta_1 cov(D, X) - \pi_1 \delta_1 var(X)}{var(\tilde{D})}
\end{aligned}$$

This is again ambiguous when compared to $\alpha - \beta$

## Problem 2

### a)

Note that with $X, Z \perp W$, we can rewrite $G(y)$ as:

$$\begin{aligned}
G(y) &= P(Y \leq a) \\
&= P(Y \leq a|W = 0)P(W = 0) + P(Y \leq a|W = 1)P(W = 1) \\
&= P(Z \leq a)P(W = 0) + P(X \leq a)P(W = 1) \\
&= F(y)(1 - \pi) + P(X \leq a)P(W = 1)
\end{aligned}$$

We know that

$$0 \leq P(X \leq a)P(W = 1) \leq P(W = 1) = \pi$$

So we can get both inequalities that we need:

$$\begin{aligned}
G(y) &\leq F(y)(1 - \pi) + \pi \\
&\Rightarrow F(y) \geq \frac{G(y) - \pi}{1 - \pi} \\
G(y) &\leq F(y)(1 - \pi) + 0 \\
&\Rightarrow F(y) \leq \frac{G(y)}{1 - \pi}
\end{aligned}$$

$F(y)$ must be in the interval [0,1] to be a probability, so we add the min and max operators to give us:

$$\max\left\{\frac{G(y) - \pi}{1 - \pi}, 0\right\} \leq F(y) \leq \min\left\{\frac{G(y)}{1 - \pi}, 1\right\}$$

In order to make lower bound sharp, for any $y$, fix $X$ to be always larger than $y$, which will give us:

$$\begin{aligned}
P(Y \leq y) &= P(X \leq y)\pi + P(Z \leq y)(1 - \pi) \\
&= \pi + P(Z \leq y)(1 - \pi) \\
&\Rightarrow \frac{G(y) - \pi}{1 - \pi} = F(y)
\end{aligned}$$

To make the upper bound sharp, for any $y$, set $X$ to be larger than y, so we have

$$\begin{aligned}
P(Y \leq y) &= P(X \leq y)\pi + P(Z \leq y)(1-\pi) \\
&= P(Z \leq y)(1-\pi) \\
&\Rightarrow \frac{G(y)}{1-\pi} = F(y)
\end{aligned}$$

## b)

Let $a = G^{-1}(1-\pi)$.

$$\begin{aligned}
E[Y \mid Y \leq a] &= E[WX + (1-W)Z \mid Y \leq a] \\
&= P(W=1 \mid Y \leq a)E[X \mid X \leq a] + P(W=0 \mid Y \leq a)E[Z \mid Z \leq a]
\end{aligned}$$

Applying Bayes Rule, we have

$$\begin{aligned}
P(W=1 \mid Y \leq a) &= \frac{P(Y \leq a|W=1)P(W=1)}{P(Y \leq a} \\
&= \frac{P(X \leq a)\pi}{P(Y \leq a)} \\
P(W=0 \mid Y \leq a] &= \frac{P(Y \leq a|W=0)P(W=0)}{P(Y \leq a} \\
&= \frac{P(Z \leq a)(1-\pi)}{P(Y \leq a)}
\end{aligned}$$

Plugging this in, we have

$$E[Y \mid Y \leq a] = \frac{\pi P(X \leq a)E[X \mid X \leq a] + (1-\pi)P(Z \leq a)E[Z \mid Z \leq a]}{P(Y \leq a)}$$

By definition, $1-\pi = G(a) = P(Y \leq a)$ and $P(X \leq a)\pi = P(Y \leq a) - P(Z \leq a)(1-\pi)$, so we can plug this in to get

$$\begin{aligned}
E[Y \mid Y \leq a] &= E[X \mid X \leq a]\frac{1-\pi-(1-\pi)P(Z \leq a)}{1-\pi} + E[Z \mid Z \leq a]\frac{P(Z \leq a)(1-\pi)}{1-\pi} \\
&\leq E[X \mid X \leq a](1-P(Z \leq a)) + E[Z \mid Z \leq a]P(Z \leq a)) \\
&\leq E[Z \mid Z > a]P(Z > a) + E[Z \mid Z \leq a]P(Z \leq a)) \\
&= E[Z]
\end{aligned}$$

The same argument holds for $E[Z] \leq E\left[Y \mid Y \geq G^{-1}(\pi)\right]$.

As we increase $\pi$, we are upweighting $X$, and the opposite happens as $\pi$ approaches 0. When $\pi = 0$, the expectations hold with equality, making these bounds sharp.

## Problem 3

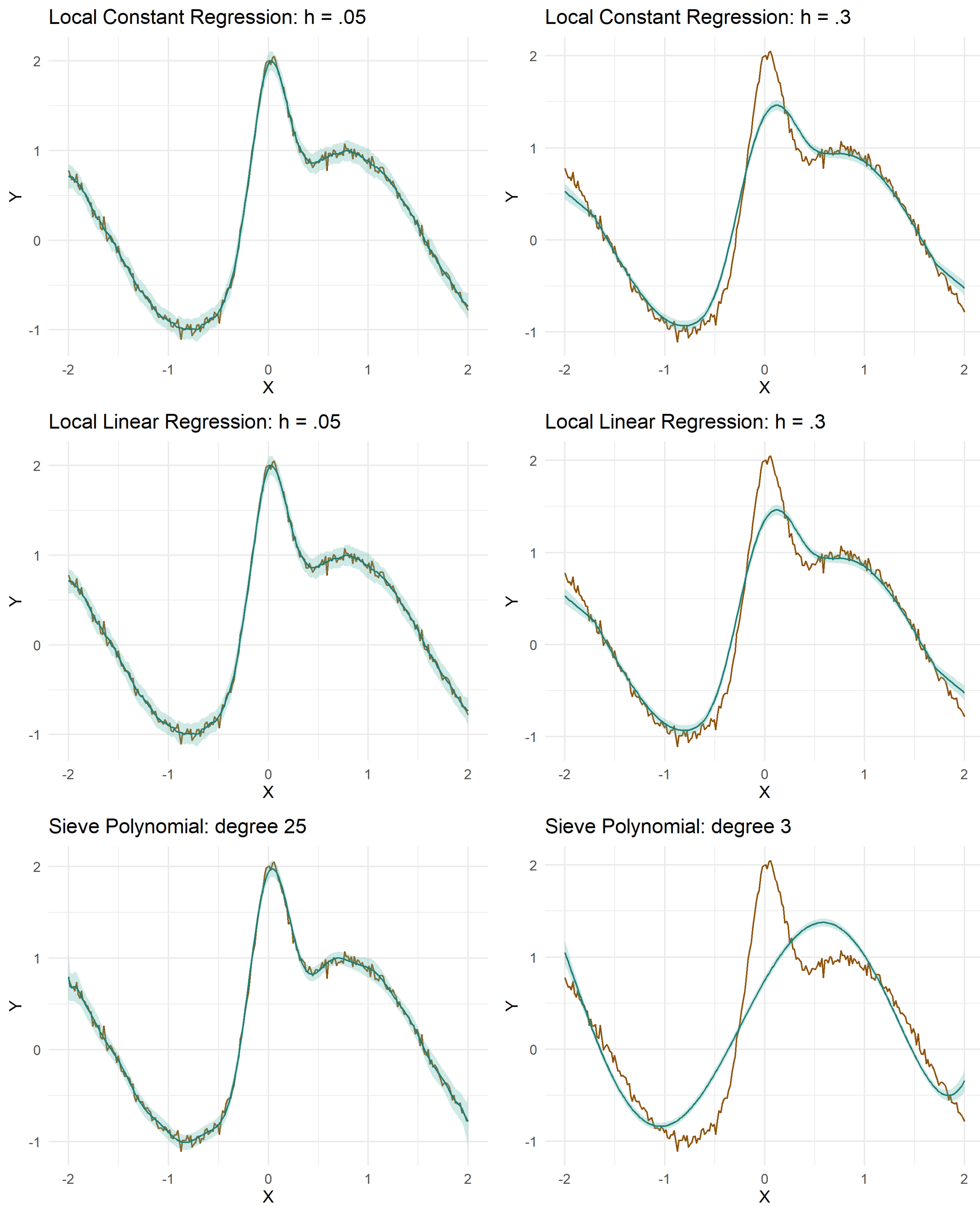
$$
\begin{aligned}
E\left[Y \mid D=d, P_d=p\right] &= E\left[Y(d) \mid D=d, P_d=p\right] \\
&= E\left[E\left[Y(d) \mid D=d, P_d=p, X=x\right] \mid D=d, P_d=p\right] \\
&= E\left[E\left[Y(d) \mid P_d=p, X=x\right] \mid D=d, P_d=p\right] \\
&= E\left[E\left[Y(d) \mid P_d=p\right] \mid D=d, P_d=p\right] \\
&= E\left[Y(d) \mid P_d=p\right]
\end{aligned}
$$

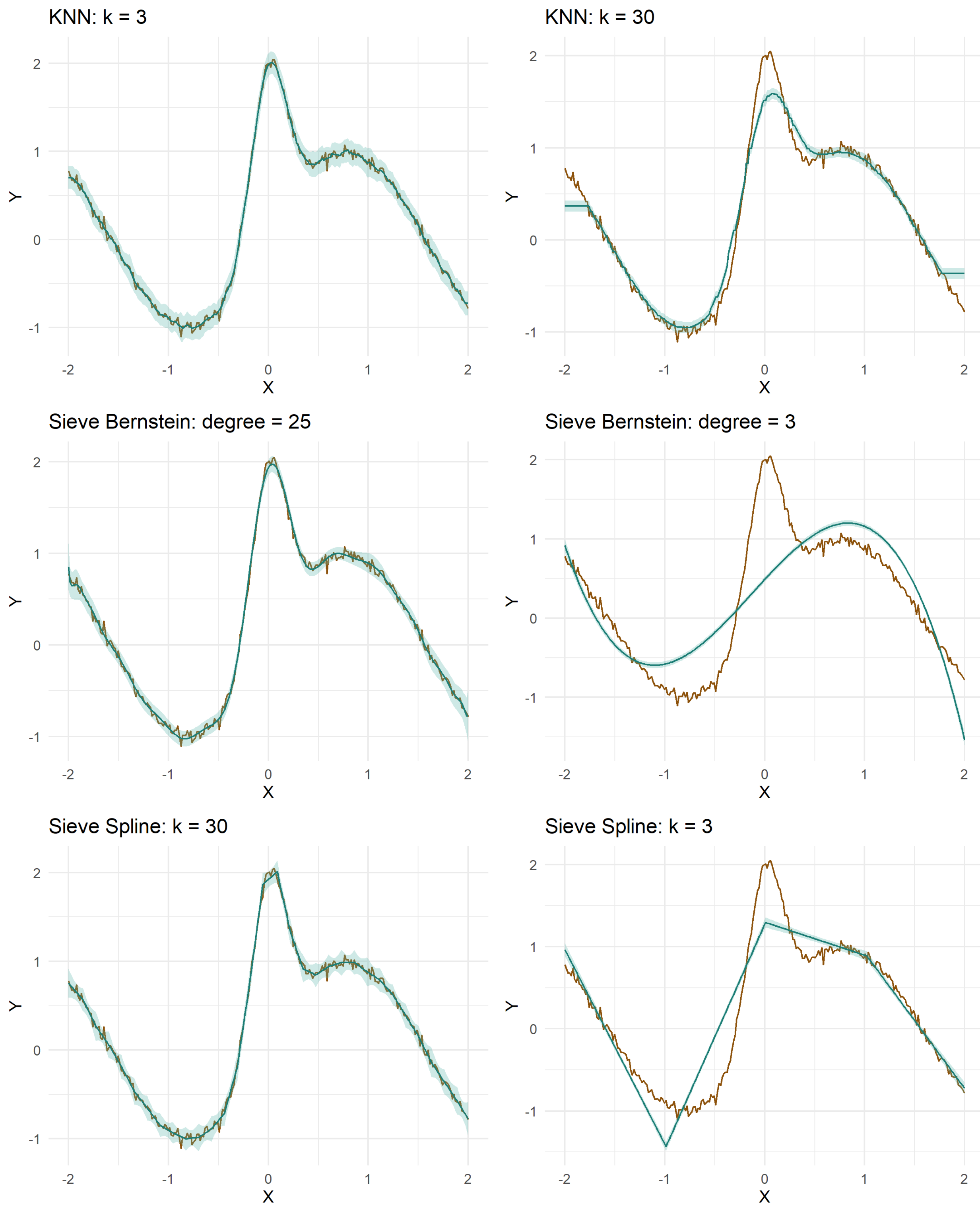
For the second part of the question, we have:

$$
\begin{aligned}
E[Y(d)] &= E[E[Y(d)|X]] \\
&= E\left[\frac{1}{P}E[Y(d)|X]P(D=d|X)\right] \\
&= E\left[\frac{1}{P}E[Y\mathbb{1}\left\{D=d\right\}|X]\right] \\
&= E\left[E\left[\frac{1}{P}Y\mathbb{1}\left\{D=d\right\}|X\right]\right] \\
&= E\left[\frac{Y\mathbb{1}\left\{D=d\right\}}{P}\right]
\end{aligned}
$$

## Problem 4

Below I plot the results of a Monte Carlo simulation with $M = 50$ samples, each with $N = 250$ observations. The graphs below show the mean estimate (green), with a band for the $\pm 1$ standard deviations, overlaid onto the true value of Y for each point (orange). For each method, I use two different tuning parameters to show the bias-variance trade-off, where the left hand side shows low bias and high variance, while the right side shows high bias and low variance.


Local Constant Regression: h = .05
Local Constant Regression: h = .3
Local Linear Regression: h = .05
Local Linear Regression: h = .3
Sieve Polynomial: degree 25
Sieve Polynomial: degree 3
Y
X
-2
-1
0
1
2


KNN: k = 3
KNN: k = 30
Sieve Bernstein: degree = 25
Sieve Bernstein: degree = 3
Sieve Spline: k = 30
Sieve Spline: k = 3
Y
X
-2
-1
0
1
2

# Problem 5

## a)

The authors include the percentage that is Protestant as a control, while the footnote states that "Protestants were more prone to vote for the Nazi Party". The treatment, which is pogroms in 1349, likely affects the percentage of protestants in that city. By controlling for this variable, you are potentially shutting off a path for the treatment effect, and using a control variable that is not pre-determined.

## b)

The replication of Panel A was straightforward, using simple OLS with clustered standard errors.

Table 1: Panel A: OLS

| term | value |
|---|---|
| pog1349 | 0.0607 |
| | (0.0224) |
| logpop25c | 0.0390 |
| | (0.0151) |
| perc_JEW25 | 0.0135 |
| | (0.0113) |
| perc_PROT25 | 0.0003 |
| | (0.0004) |
| N | 320.0000 |
| Adj. R^2 | 0.0514 |

I attempt to replicate Panel B and C, which reported ATT estimates using robust nearest neighbor estimation with the four closest matches. To replicate, I used the Mahalabonis distance method.

Table 2: Panel B: Matching

| term | value |
|---|---|
| pog1349 - Matching | 0.0722 |
| | (0.0222) |
| pog1349 - Geographic Matching | 0.0819 |
| | (0.0295) |

## c)

I estimate the propensity score using logit, and then use nearest neighbor estimation with the four closest matches on the propensity score. I then use 100 bootstraps to find the standard error. Both the ATT and ATE estimates come quite close to the matching estimates. The ATE is a bit smaller, suggesting that there is some selection occuring.

Table 3: Propensity Score

| term | value |
|---|---|
| pog1349 - Prop. Score ATT | 0.0744 |
| | (0.0172) |
| pog1349 - Prop. Score ATE | 0.0688 |
| | (0.0200) |