ECON 31720: Problem Set 1

Nadia Lucas

October 20, 2020

Problem 1

We are presented with a binary treatment variable potential outcomes framework with a constant treatment effect, $Y(1) - Y(0) = \alpha$. This implies

$$\begin{aligned} Y &= DY(1) + (1 - D)Y(0) \\ &= \alpha D + Y(0) \end{aligned}$$

We compare β and γ from running the following specifications

$$\begin{aligned} Y &= \eta_0 + \beta D + \epsilon_0 \\ Y &= \eta_1 + \gamma D + \tau X + \epsilon_1 \end{aligned}$$

a

It is not true that $|\alpha - \gamma| \leq |\alpha - \beta|$.

If we use the FWL approach, we can define $\tilde{D} = D - BLP(D \mid 1)$

$$\begin{aligned} \beta &= \frac{Cov(Y, \tilde{D})}{Var(\tilde{D})} \\ &= \frac{Cov(\alpha D + Y(0), \tilde{D})}{Var(\tilde{D})} \\ &= \alpha + \frac{Cov(Y(0), D)}{Var(D)} \end{aligned}$$

Similarly we define $\hat{D} = D - BLP(D \mid X, 1)$ and

$$\begin{aligned} \gamma &= \frac{Cov(Y, \hat{D})}{Var(\hat{D})} \\ &= \frac{Cov(\alpha D + Y(0), D - \alpha_0 - \beta_0 X)}{Var(\hat{D})} \\ &= \alpha - \beta_0 \frac{Cov(X, Y(0))}{Var(\hat{D})} + \frac{Cov(Y(0), D)}{Var(\hat{D})} \end{aligned}$$

We can combine these findings to see that

$$|\alpha - \gamma| - |\alpha - \beta| = \beta_0 \frac{Cov(X, Y(0))}{Var(\hat{D})}$$

Since we have no way of knowing whether the covariance between X and $Y(0)$ is positive or negative, we cannot say anything about the direction of the inequality.

b

If D and X are uncorrelated, then $\hat{D} = D - BLP(D|X, 1)$ does not depend on X at all and can be though of as

$$\begin{aligned}\hat{D} &= D - BLP(D|X, 1) \\ &= BLP(D|1) \\ &= \tilde{D}\end{aligned}$$

Therefore we end up with identical expressions for γ and β giving

$$|\alpha - \gamma| = |\alpha - \beta|$$

Since they are equivalent, the inequality in (a) holds.

c

If X is uncorrelated with $Y(0)$ and $Y(1)$ then we need to take a closer look at

$$\beta = \alpha + \frac{Cov(Y(0), D)}{Var(D)}$$

and

$$\gamma = \alpha + \frac{Cov(Y(0), D)}{Var(\hat{D})}$$

They almost look the same but the denominators are slightly different.

$$\begin{aligned}Var(\hat{D}) &= Var(D - \alpha_0 - \beta_0 X) \\ &= Var(D) - 2\beta_0 Cov(X, D) + \beta_0 Var(X)\end{aligned}$$

The sign and magnitude of $-2\beta_0 Cov(X, D)$ can make the inequality go one way or another. Because we don't know how X and D are related we cannot say anything about the direction of the inequality.

d

Suppose that $E[Y(0) \mid D = d, X = x] = E[Y(0) \mid X = x]$. This still would not change my answer to (a) because all this gives is that γ is no longer biased

$$\begin{aligned}\gamma &= \alpha - \beta_0 \frac{Cov(X, Y(0))}{Var(\hat{D})} + \frac{Cov(Y(0), D)}{Var(\hat{D})} \\ &= \alpha\end{aligned}$$

However, we still have a bias term from the covariance between $Y(0)$ and D when backing out β.

$$\beta = \alpha + \frac{Cov(Y(0), D)}{Var(D)}$$

Since we can't sign this term additional bias term on the β estimator, we have no way of knowing which direction the inequality is in for part a.

e

Suppose now that $E[Y(0)|X = x]$ is a linear function of X. This does not change anything becuase as we know, no matter what the relationship between $Y(0)$ and X is, the covariance between $Y(0)$ and D still give problematic bias terms when comparing β and γ.

Problem 2

a

Suppose we observe some variable Y where

$$Y = WX + (1 - W)Z$$

If $G(y) \equiv P[Y \leq y]$ denotes the distribution function for Y and $F(y) \equiv P$ the distribution for Z. I derive the following bounds for $F(y)$.

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

We know that any distribution function falls between 0 or 1 so those outer bounds will hold. In order to show the inequality holds, first decompose $G(y)$ using that $\pi \equiv P[W = 1]$ and that W is independent of both X and Z

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

Now take a look at how this translates to the bounds by plugging in this decomposition

$$\max\left\{\frac{P[X \leq y]\pi - \pi}{1 - \pi} + P[Z \leq y], 0\right\} \leq P[Z \leq y] \leq \min\left\{\frac{P[X \leq y]\pi}{1 - \pi} + P[Z \leq y], 1\right\}$$

The only uncertainty left in this expression are the terms that deal with the distribution of X since we have no information about this distribution. However, we can think about what the bounds on it are, using the property that a probability distribution falls between 0 and 1.

$$0 \leq P[X \leq y]\pi \leq \pi$$

Therefore $\frac{P[X \leq y]\pi - \pi}{1 - \pi}$ is bounded above by 0 and the left inequality holds. In that case that $P[X \leq y] = 1$ the left inequality holds with equality and thus this bound is sharp.

Similarly, $P[X \leq y]\pi$ is weakly positive and thus the right inequality also holds. If $P[X \leq y] = 0$ then the right inequality holds with equality and thus this bound is also sharp.

b

Now considering an inequality bounding expectations of these random variables we have

$$E[Y \mid Y \leq G^{-1}(1 - \pi)] \leq E[Z] \leq E[Y \mid Y \geq G^{-1}(\pi)]$$

For ease of notation going forward, define

$$l \equiv G^{-1}(1-\pi)$$
$$r \equiv G^{-1}(\pi)$$

Similar to the intuition in part a, decompose the expectation in the left bound, using the definition of Y and the independence of W and X, Z to get

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

Each of the probability terms can be rearranged using Bayes' rule

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

Now given what we know l is, the denominators can be rearranged.

$$\begin{aligned} l &= G^{-1}(1-\pi) \\ \implies G(l) &= 1-\pi \\ P[Y \leq l] &= 1-\pi \end{aligned}$$

We can also further simplify the distribution of X given what we know about the total probability

$$\begin{aligned} P[X \leq l]\pi &= P[Y \leq l] - P[Z \leq l](1-\pi) \\ &= (1-\pi) - P[Z \leq l](1-\pi) \end{aligned}$$

Combine all of this simplification to get

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

Thus we see that left inequality holds. The bound is sharp and binds if $P[W=1 \mid Y \leq l] = 0$.

Similarly we can rearrange the right inequality by using the law of total probability and Bayes' Rule. In this case, we have

$$\begin{aligned} h &= G^{-1}(\pi) \\ \implies G(h) &= \pi \\ \implies P[Y \geq h] &= 1-\pi \end{aligned}$$

$$
\begin{aligned}
E[Y \mid Y \geq h] &= E[WX + (1-W)Z \mid Y \geq h] \\
&= P[W = 1 \mid Y \geq h] E[X | X \geq h] + P[W = 0 \mid Y \geq h] E[Z \mid Z \geq h] \\
&= \frac{P[Y \geq h \mid W = 1] P[W = 1]}{P[Y \geq h]} E[X | X \geq h] + \frac{P[Y \geq h \mid W = 0] P[W = 0]}{P[Y \geq h]} E[Z | Z \geq h] \\
&= \frac{(1-\pi) - P[Z \geq h](1-\pi)}{1-\pi} E[X \mid X \geq h] + \frac{P[Z \geq h](1-\pi)}{1-\pi} E[Z | Z \geq h] \\
&= (1 - P[Z \geq h]) E[X \mid X \geq h] + P[Z \geq h] E[Z \mid Z \geq h] \\
&\geq P[Z < h] E[Z < h] + P[Z \geq h] E[Z \mid Z \geq h] \\
&= E[Z]
\end{aligned}
$$

This shows the right inequality holds. It holds with equality if $P[W = 1 \mid Y \geq h] = 0$ and thus these bounds both hold and are sharp.

c

Now suppose we evaluate the impact of a job training program on wages. Let $Y^* = DY(1) + (10D)Y(0)$. We do not observe Y^* because we don't observe those who are unemployed. So we use S to denote employment where $Y = Y^*$ if $S = 1$. Denote the average treatment effect of those who would be employed after one year in the program as:

$$\mu = E[Y(1) - Y(0) \mid S(0) = 1, S(1) = 1]$$

We show

$$
\begin{gathered}
E[Y \mid D = 1, S = 1, Y \leq \bar{G}^{-1}(1-\pi)] - E[Y \mid D = 0, S = 1] \\
\leq \mu \leq E[Y \mid D = 1, S = 1, Y \geq \bar{G}^{-1}(\pi)] - E[Y \mid D = 0, S = 1]
\end{gathered}
$$

where

$$\pi \equiv \frac{P[S = 1 \mid D = 1] - P[S = 1 \mid D = 0]}{P[S = 1 \mid D = 1]}$$

Similarly to part b we can think about

$$\bar{G}(y) = P[Y \leq y \mid D = 1, S = 1]$$

Decomposing this in a similar way to part b we can see that this is equal to the probability that we observe you as a treated data point and you otherwise would not have shown up or we observe you regardless.

$$\bar{G}(y) = \pi P[Y_1 \mid D = 1, S_0 = 0, S_1 = 1] + (1-\pi) P[Y \mid D = 1, S_0 = 1, S_1 = 1]$$

Conceptually,

$$E[Y(1) \mid S(0) = 1, S(1) = 1] \leq E[Y \mid D = 1, S = 1, Y \geq \bar{G}^{-1}(\pi)]$$

When we subtract $E[Y(0) \mid D = 0, S = 1]$ from both sides, the inequality still holds. Similarly with the left hand inequality, we only observe those in the low distribution of $\bar{G}$ when $Y \leq \bar{G}^{-1}(1-\pi)$ so this inequality also holds.

The bounds are sharp when $\pi = 0$ or $\pi = 1$

Problem 3

Observe a discretely distributed treatment variable D, outcome Y, and covariates X. Let $\{Y(d) : d \in \mathcal{D}\}$ denote potential outcomes for Y.

Assume distribution of $Y(d)$ conditional on $D = d$ and $X = x$ is the same distribution of $Y(d)$ conditional on $X = x$ for all d and x. Let $p(d, x) \equiv P[D = d, X = x]$, and let $P_d \equiv p(d, X)$. Show

$$E[Y \mid D = d, P_d = p] = E[Y(d) \mid P_d = p]$$

By Law of Iterated Expectations

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

Let $P \equiv p(D, X)$ and show

$$E[Y(d)] = E\left[\frac{Y\mathbb{1}[D = d]}{P}\right]$$

First starting with law of Iterated expectations

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

The second equality comes from the definition of the potential outcome $Y(d)$ as the occurrences of Y that come about when d is realized, normalized by the probability that d occurs in the data. The third equality comes from taking the expectation over all possible P_ds.

Problem 4

For various non-parametric regressions, I plot the bias-variance tradeoff in picking various tuning parameters. In general we see that tuning parameters that are closer to the data (e.g. tiny bandwidth or only a couple of nearest neighbors) give us noisier (more variable) results. However, smoother tuning parameters will lead to slightly more bias. When we get to higher powers of say sieves, we see this bias tradeoff clearly in the edges of the window where the measures are extremely sensitive. In general, non parametric methods that smooth the data become biased towards the mean of the data, and we can see that in the plots where peaks and valleys are less pronounced.

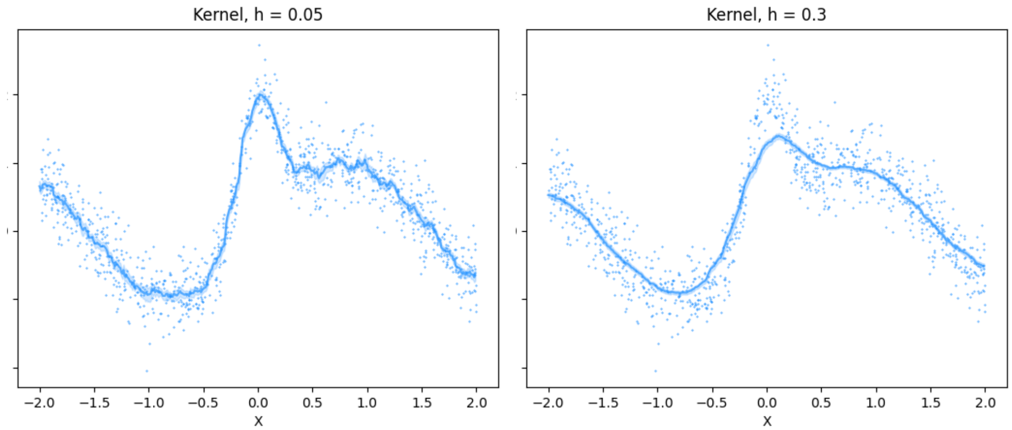

Figure 1: Local constant (kernel) regression

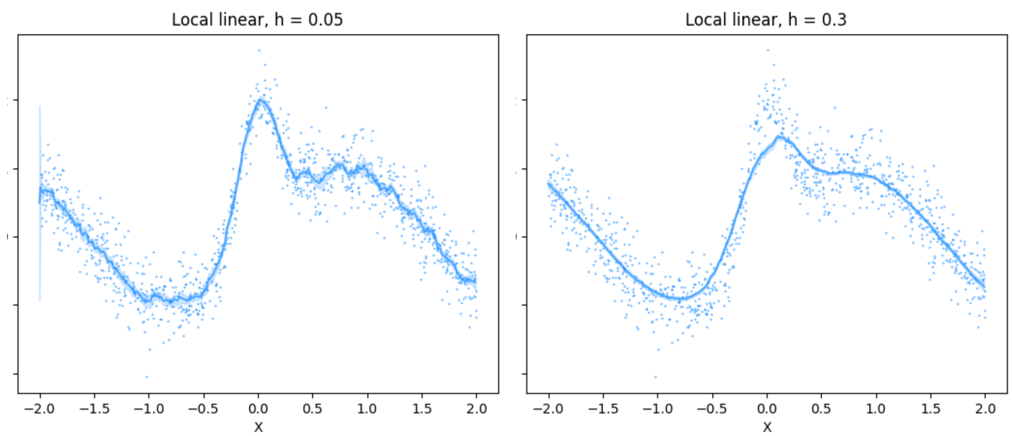

Figure 2: Local linear regression

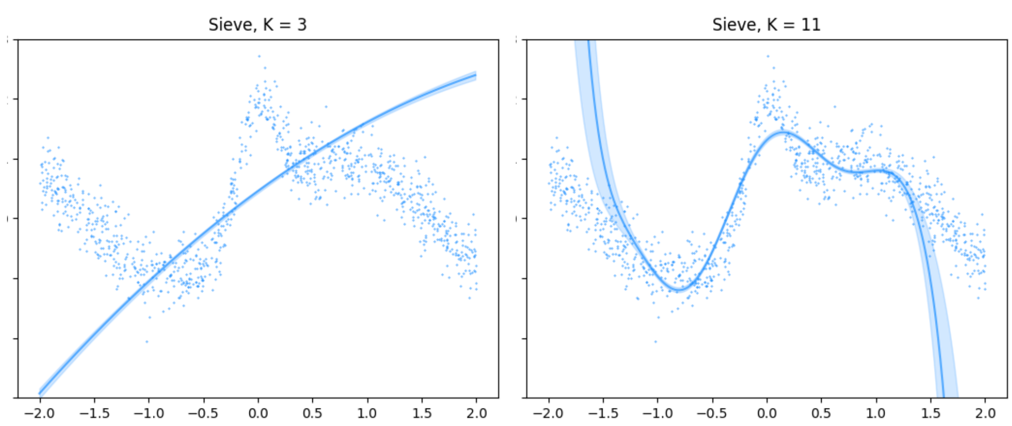

Figure 3: Local constant (kernel) regression

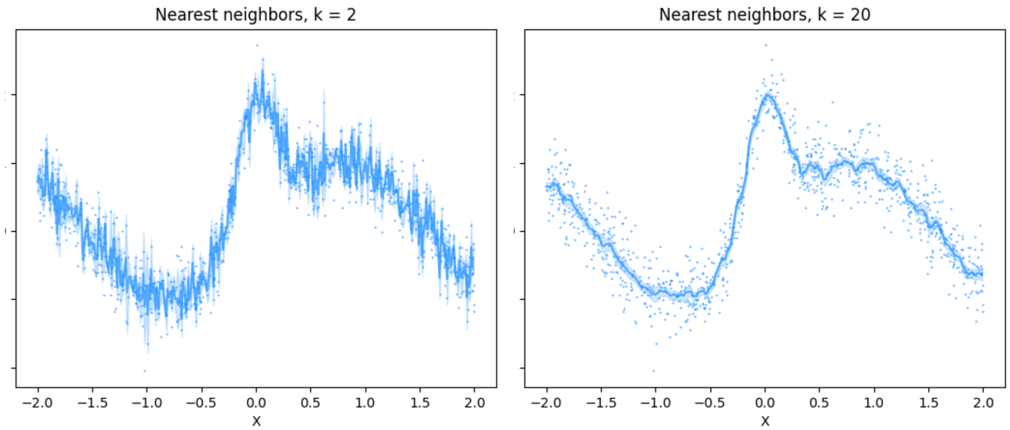

Figure 4: Local constant (kernel) regression

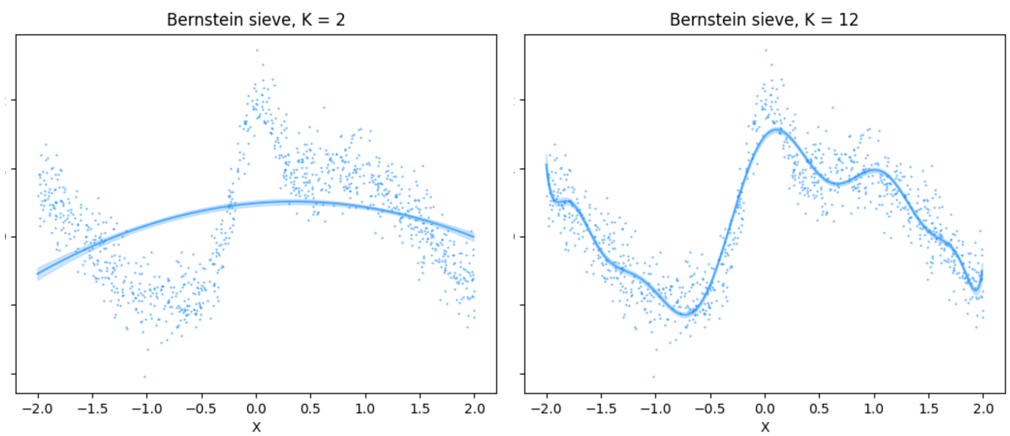

Figure 5: Local constant (kernel) regression

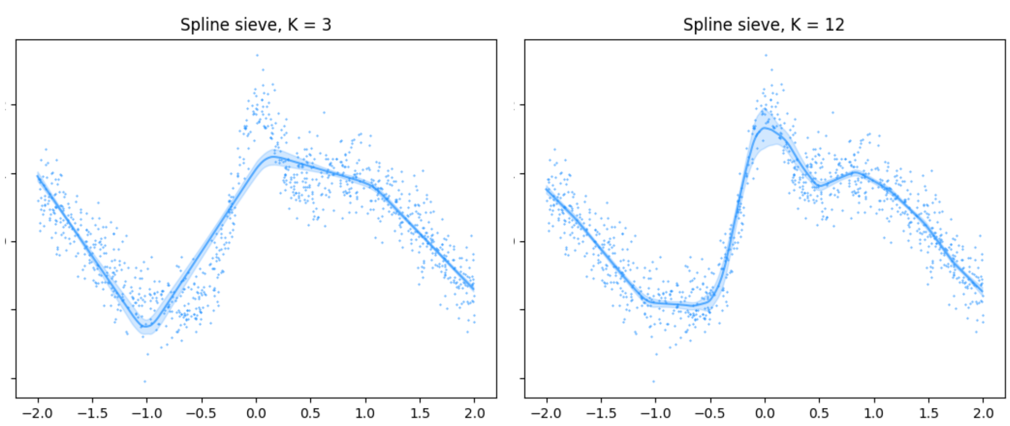

Figure 6: Local constant (kernel) regression

5

a

There seems to be a big endogeneity problem with the choice of covariates in this paper's specification. We have that percentage Jewish and percentage Protestant are controlled for. Surely the pogroms in 1349 affect the percentage of the population that is Jewish in 1933. Especially if we consider the main reasoning by which the authors argue that the pogroms during the Black Death affected pogroms today which is that there is continuity in anti-Semitism in these towns where people settle for long periods of time. If we believe that story, the pogroms in 1349 endogenously affect the percentage of the population that is Jewish in 1933. The estimate will be biased if we include Jewish population as a covariate since it picks up some of the effects of the old pogroms.

b

Table VI Replication:

Covariate	Estimate	Standard error (cluster robust)
POG (1349)	0.060703	0.02258
ln(Pop)	0.03896	0.01519
Perc Jewish	0.01351	0.01141
Perc Protestant	0.000336	0.000423
Intercept	-0.3927	0.1396

Table 1: Panel A Replication

Matching replication: The ATT estimated in Panel B: **0.07435**

The ATT estimated in Panel C: **0.0819**

c

For the propensity score matching I perform a logit estimator to get the propensity scores. Then I run on k-nearest neighbors to get both the ATT and the ATE estimators

ATT: 0.0819

ATE: 0.0562

For standard errors for both the matching and propensity score estimates, I attempted bootstrapping but something is not working (you can see the attempt in the attached code) so I have no errors to report unfortunately. There may be some bug that I am not catching since I've run out of time.