ECON 217: Section Notes

Week 10

David Sungho Park

March 10-12, 2021

Splines Regression

4. [10 points] Suppose we want to estimate the returns to education while controlling flexibly for age using a linear spline with three knots $\tau_1 < \tau_2 < \tau_3$.

   (a) What is the expression for the conditional mean of $y = income$ given education and the age spline. Include an intercept in your model and assume that education and the age spline are additively separable (i.e. no interaction effects between education and age).

Slope before knot t1

change in slope at knot t1

$$
\begin{aligned}
\mathbb{E}\left[y|educ, age\right] =& \beta_0 + \alpha_0 educ + \beta_1 age + \beta_2 (age - \tau_1)\, 1\,(age \geq \tau_1) \\
&+ \beta_3 (age - \tau_2)\, 1\,(x \geq \tau_2) + \beta_4 (age - \tau_3)\, 1\,(age \geq \tau_3)
\end{aligned}
$$

change in slope at knot t2

change in slope at knot t3

Splines Regression

(b) What are the expressions for the marginal effects with respect to age for each of the intervals defined by the knots?

$$\frac{\partial \mathbb{E}\left[y|educ, age < \tau_1\right]}{\partial age} = \beta_1$$

$$\frac{\partial \mathbb{E}\left[y|educ, \tau_1 \leq age < \tau_2\right]}{\partial age} = \beta_1 + \beta_2$$

$$\frac{\partial \mathbb{E}\left[y|educ, \tau_2 \leq age < \tau_3\right]}{\partial age} = \beta_1 + \beta_2 + \beta_3$$

$$\frac{\partial \mathbb{E}\left[y|educ, age > \tau_3\right]}{\partial age} = \beta_1 + \beta_2 + \beta_3 + \beta_4$$

Splines Regression

(c) Write STATA code to estimate the model assuming the three knots are equally spaced and that the coefficients represent the change in slope from the previous interval.

Number of **intervals**
(no. of knots + 1)

```
mkspline age 4 = age, marginal
reg income educ age1-4
```

CATE vs. Quantile regression

(f) [5 points] If we are interested in estimating the heterogeneous effect of receiving an extra year of education on expected earnings across different ethnic groups, we would use quantile regression.

i. False. We could use any method that estimates conditional average treatment effects but not quantile regression because we are interested in the effect on expected earnings rather than on different quantiles of the earnings distribution.

- **Quantile regression** is over a specific quantile of the **outcome** variable.
- **CATE** is still for the average of the outcome variable (i.e. all of the quantiles) but only over different subgroups defined by **covariates**.

Selection-on-observable assumption

(g) [5 points] The selection on observables assumption requires that the treatment is obtained through a randomized experiment.

i. False. The selection on observables assumption only requires that the treatment is independent of the potential outcomes conditional on the covariates, rather than unconditionally independent of the potential outcomes. As long as the selection into treatment is based only observable characteristics, rather than unobservable characteristics, the treatment does not need to be obtained through a randomized experiment.

$$W_i \perp (Y_i(0), Y_i(1)) | X_i$$

Conditional average treatment effect (CATE)

(h) [5 points] Conditional average treatment effects equal the difference in the expected potential outcomes.

i. False. Conditional average treatment effects equal the difference in the expected potential outcomes conditional on the covariates.

- We would like to estimate the conditional average treatment effect (CATE):

$$\tau_0(X_i) \equiv E[Y_i(1) - Y_i(0)|X_i] = \mu_1(X_i) - \mu_0(X_i)$$

(Slide 19)

$$\hat{\tau}^*_{m,b}(x) = \frac{\sum_{\{i:W_i^*=1, X_i^* \in R_b^*(x)\}} Y_i^*}{|\{i : W_i^* = 1, X_i^* \in R_b^*(x)\}|} - \frac{\sum_{\{i:W_i^*=0, X_i^* \in R_b^*(x)\}} Y_i^*}{|\{i : W_i^* = 0, X_i^* \in R_b^*(x)\}|}$$

(Slide 21)

Average treatment effect (ATE) (Slides 18-20)

- Let $W_i \in \{0,1\}$ be a binary treatment variable.
- Let $Y_i(0)$ and $Y_i(1)$ be the potential outcomes.
- The potential outcomes can be expressed as:

$$Y_i(0) = \mu_0(X_i) + \epsilon_{0i}, E[\epsilon_{0i}|X_i] = 0 \implies E[Y_i(0)|X_i] = \mu_0(X_i)$$

$$Y_i(1) = \mu_1(X_i) + \epsilon_{1i}, E[\epsilon_{1i}|X_i] = 0 \implies E[Y_i(1)|X_i] = \mu_1(X_i)$$

- We would like to estimate the conditional average treatment effect (CATE):

$$\tau_0(X_i) \equiv E[Y_i(1) - Y_i(0)|X_i] = \mu_1(X_i) - \mu_0(X_i)$$

- The relation $Y_i = W_i Y_i(1) + (1 - W_i) Y_i(0)$ gives us

$$Y_i = \tau_0(X_i) W_i + \mu_0(X_i) + \underbrace{W_i \epsilon_{1i} + (1 - W_i)\epsilon_{0i}}_{\eta_i}$$

- $W_i \perp (Y_i(0), Y_i(1))|\, X_i$ implies that $W_i \perp (\epsilon_{0i}, \epsilon_{1i})|\, X_i$ and $E[\eta_i|W_i, X_i] = 0$
- In the case where $\tau_0(X_i) = \tau_0$ does not depend on X_i and $\mu_0(X_i) = X_i'\gamma$, one can estimate the average treatment effect (ATE) by running OLS on

$$Y_i = \tau_0 W_i + X_i'\gamma + \eta_i$$

Estimating CATE with Random Forest

(Slide 21)

- The estimate of the conditional average treatment effect is the average over the estimates computed over B subsampled trees.

$$\hat{\tau}_n(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{\tau}^*_{m,b}(x)$$

$$\hat{\tau}^*_{m,b}(x) = \frac{\sum_{\{i: W_i^*=1, X_i^* \in R_b^*(x)\}} Y_i^*}{|\{i : W_i^* = 1, X_i^* \in R_b^*(x)\}|} - \frac{\sum_{\{i: W_i^*=0, X_i^* \in R_b^*(x)\}} Y_i^*}{|\{i : W_i^* = 0, X_i^* \in R_b^*(x)\}|}$$

- $R_b^*(x)$ is the region that contains x and it is computed using the propensity score algorithm of Wager and Athey (2017), which uses the covariates and the treatment variable instead of the outcome to place splits.

- $\hat{\tau}^*_{m,b}(x)$ is the difference in the average of the outcome between the treated and control observations in the region that contains x.

(Pset4 example)
- Y: test score
- W: being in a small class
- X: *fem,wh,fl,urb,age,exp,lad,deg*