17.802 – Recitation 3

Regression Recap and Experiment

Yiqing Xu

MIT

Feb 21, 2014

MHE 3.1

- Regression solves the population least square problem and is the Best Linear Predictor (BLP) of Y_i given X_i

$$\begin{aligned}\beta &= \underset{b}{\operatorname{argmin}} E[(Y_i - X_i b)^2] \\ &= E[X_i' X_i]^{-1} E[X_i' Y_i]\end{aligned}$$

- If the CEF is linear (normality, saturated), regression is it

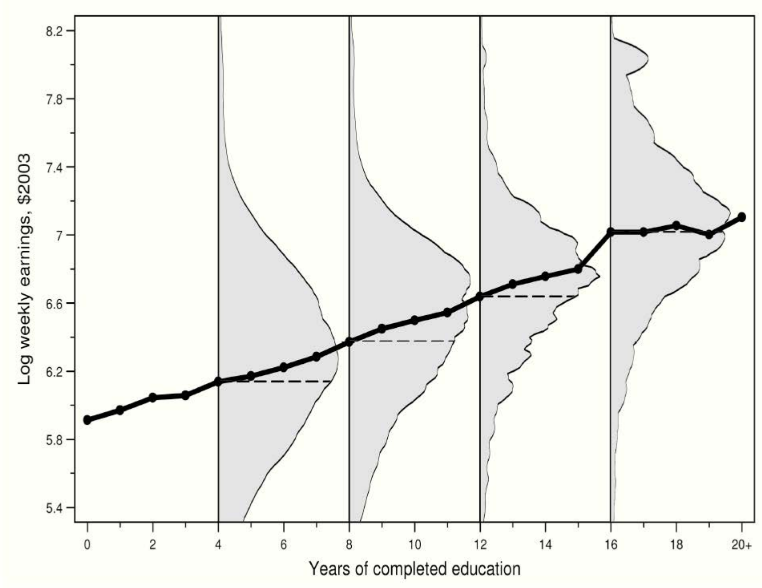
$$E[Y_i | X_i = x] = \int t f_y(t | X_i = x) dt$$

- Regression gives the best linear approximation to the CEF

$$E[(Y_i - X_i b)^2] = E[(Y_i - E(Y_i|X_i))^2] + E[(E(Y_i|X_i) - X_i b)^2]$$

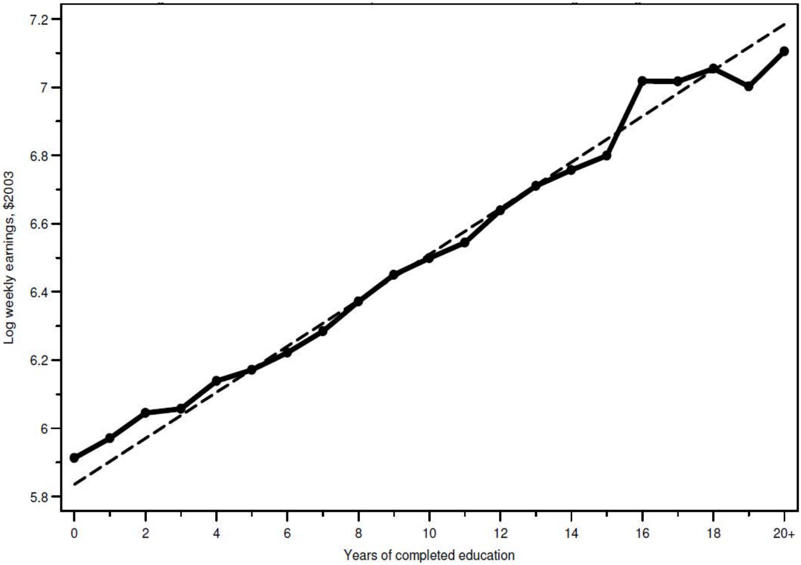
$$\beta = \underset{b}{\operatorname{argmin}} E[(E(Y_i|X_i) - X_i b)^2]$$

Figure 3.1.1



8.2
7.8
7.4
7
6.6
6.2
5.8
5.4
Log weekly earnings, $2003
0
2
4
6
8
10
12
14
16
18
20+
Years of completed education

Figure 3.1.2



Sample is limited to white men, age 40-49. Data is from Census IPUMS 1980, 5% sample.

Anatomy of Regression

- The Frisch-Waugh-Lovell (FWL) theorem:

$$\beta_k = \frac{Cov(Y_i, \tilde{x}_{ki})}{Var(\tilde{x}_{ki})}$$

$$\beta_k = \frac{Cov(Y_i, \tilde{x}_{ki})}{Var(\tilde{x}_{ki})} = \frac{Cov(\tilde{Y}_i, \tilde{x}_{ki})}{Var(\tilde{x}_{ki})}$$

$$\beta_k = \frac{Cov(Y_i, \tilde{x}_{ki})}{Var(\tilde{x}_{ki})} = \frac{Cov(\tilde{Y}_i, \tilde{x}_{ki})}{Var(\tilde{x}_{ki})} = \frac{\beta_k Cov(\tilde{x_{ki}}, \tilde{x}_{ki})}{Var(\tilde{x}_{ki})}$$

  where $\tilde{x}_{ki}$ is the residual from a regression of x_{ki} on all other covariates.

- $$Y_i = \alpha + \delta D_i + \beta x_i$$

  (1) Reg D_i on x_i and constant, get residual $\tilde{D}_i$;
  (2) Reg Y_i on $\tilde{D}_i$ and constant, get $\hat{\delta}$.
- Since under random assignment $D_i \approx \tilde{D}_i$, controlling for x_i doesn't matter much

Omitted Variable Bias: An Example from MHE

- Wages on schooling (S_i), controlling for ability (A_i)

$$Y_i = \alpha + \rho S_i + A_i'\gamma + \epsilon_i$$

- Ability is hard to measure. What if we leave it out?

$$\frac{Cov(Y_i, S_i)}{V(S_i)} = \rho + \gamma'\delta_{AS}$$

where δ_{AS} is the vector of coefficients from regressionos of the elements of A_i on S_i

- Omitted variable bias: the effect of omitted $\times$ the correlation between omitted and included (which is often of interest)

Use regression to analyse experimental data

- Why using regressions is valid?

$$
\begin{aligned}
Y_i &= D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i} \\
&= Y_{0i} + (Y_{1i} - Y_{0i}) \cdot D_i \\
&= (Y_{0i} - \bar{Y}_0) + [(Y_{1i} - \bar{Y}_1) - (Y_{0i} - \bar{Y}_0)] \cdot D_i + \bar{Y}_0 + (\bar{Y}_1 - \bar{Y}_0) D_i \\
&= \bar{Y}_0 + (\bar{Y}_1 - \bar{Y}_0) D_i + \{(Y_{i0} - \bar{Y}_0) + D_i \cdot [(Y_{i1} - \bar{Y}_1) - (Y_{i0} - \bar{Y}_0)]\} \\
&= \underbrace{\bar{Y}_0}_{\alpha} + \underbrace{(\bar{Y}_1 - \bar{Y}_0)}_{b} D_i + \underbrace{\{(Y_{i0} - \bar{Y}_0) + D_i \cdot [(Y_{i1} - \bar{Y}_1) - (Y_{i0} - \bar{Y}_0)}_{\epsilon} \\
&= \alpha + b D_i + \epsilon_i
\end{aligned}
$$

- $b = \bar{Y}_1 - \bar{Y}_0$ is ATE!
- In order to have $plim \hat{b} \rightarrow \bar{Y}_1 - \bar{Y}_0$, we need $E[\epsilon_i | D_i] = 0$.

When $E[\epsilon_i|D_i] = 0$?

$$\begin{aligned} E[\epsilon_i|D_i] &= E[(Y_{i0} - \bar{Y}_0) + D_i \cdot [(Y_{i1} - \bar{Y}_1) - (Y_{i0} - \bar{Y}_0)|D_i] \\ &= E[(Y_{i0} - \bar{Y}_0)|D_i] + D_i \cdot E[(Y_{i1} - \bar{Y}_1) - (Y_{i0} - \bar{Y}_0)|D_i] \end{aligned}$$

$$E[\epsilon_i|D_i = 0] = E[(Y_{i0} - \bar{Y}_0)|D_i = 0] = 0$$
$$E[\epsilon_i|D_i = 1] = E[(Y_{i1} - \bar{Y}_1)|D_i = 1] = 0$$

- Under random assignment:
  $E[(Y_{i0} - \bar{Y}_0)|D_i = 0] = E[Y_{i0} - \bar{Y}_0] = 0$;
  $E[(Y_{i1} - \bar{Y}_1)|D_i = 1] = E[Y_{i1} - \bar{Y}_1] = 0$

Lin (2013)

$$Y_i = \alpha + bD_i + \beta_1 \cdot (x_i - \bar{x}) + \beta_2 \cdot D_i \cdot (x_i - \bar{x}) + \epsilon_i$$

- What does it mean, intuitively?
- Why demeaning?

$$\begin{aligned} E_x[Y_i|D_i] &= \alpha + bD_i + (\beta_1 + \beta_2 \cdot D_i)E[x_i - \bar{x}|D_i] + E[\epsilon_i|D_i] \\ &= \alpha + bD_i + (\beta_1 + \beta_2 \cdot D_i)E[x_i - \bar{x}|D_i] \\ &= \alpha + bD_i \end{aligned}$$

The Interaction Estimator

- Three steps:
  1. Using the treated units, reg Y_{1i} on $x_i - \bar{x}$ and constant, obtaining $\hat{\beta}^1$ and $\hat{\alpha}^1$
  2. Using the control units, reg Y_{0i} on $x_i - \bar{x}$ and constant, obtaining $\hat{\beta}^0$ and $\hat{\alpha}^0$
  3. $\widehat{ATE} = \bar{\widehat{Y}}_1 - \bar{\widehat{Y}}_0$

- This is equivalent to regress Y_i on $(x_i - \bar{x})$ and $D_i \cdot (x_i - \bar{x})$

$$Y_i = \alpha + bD_i + \beta_1 \cdot (x_i - \bar{x}) + \beta_2 \cdot D_i \cdot (x_i - \bar{x}) + \epsilon_i$$

$$\widehat{ATE} = \hat{b}$$

- What is this doing?

$$\widehat{Y}_1 = \hat{\alpha}^1 + (x - \bar{x})\hat{\beta}^1$$
$$\widehat{Y}_0 = \hat{\alpha}^0 + (x - \bar{x})\hat{\beta}^0$$

Wrap-up

- Why regression?
- Regression anatomy
- Regression and experiment
- Covariate adjustment