7.6 Program Evaluation

1. Selection bias and Programs: consider the following example where those firms who received for job training grants by government are considered treated while those who did not are controls

   ▷ The argument is that firms apply on a first come, first serve basis. The estimated regression from this data (JTRAIN for 1988) is

$$\begin{aligned}\widehat{log(scrap)} =& 4.99 - 0.052grant - 0.455log(sales)\\ &+ 0.639^{*}log(employ)\end{aligned}$$

   ▷ *grant*: firms recieving grants have a scrap rate of 0.052 less than the ones not receiving grant, but small t prevents us from trusting the results completely.

- Selection bias: What if the firms who applied for the grants had more resources and were otherwise more proactive to seek alternative finances?

- We must be careful to include factors that might be systematically related to the binary independent variable of interest in order to provide correct inference.

- The bias induced by omiting out key control factors which explain systematic difference between the control and treated induce what we call a "Selection bias".

7.6.1 Program Evaluation and Unrestricted Regression Adjustment

1. General model for program evaluation when w is the policy or program indicator, $x_1, x_2 \ldots, x_k$ are control variables,

$$E(y|w, \mathbf{x}) = \alpha + \tau w + \gamma_1 x_1 + \cdots + \gamma_k x_k$$

   where $y = (1-w)\hat{y}(0) + w\hat{y}(1)$ is the observed average y.

2. *Controls and self selection*: The problem of participation decisions differing systematically by individual characteristics is often referred to as the self-selection problem.
   Once we control for all the variables (**co-variates**) we could possibly control for, we could capture the effect of the program.

3. **Example**: We want to capture the effect of drug consumption during teens on labor force participation. However, those who consume drugs are likely to be from the poor family, subjected to lower levels of education and less skilled by the time they enter labor force. These other characteristics interact with their ability to join labor force.

4. Regression adjusted estimate: is the estimate of $\hat{\tau}$ for the program in the presence of co-variates. This implies conditional independence of the treatment.

   - ▷ Thus an individual who took drugs conditional on all other controls will have the following predicted value

$$y(1) = \psi_1 + (\mathbf{x} - \boldsymbol{\eta})\boldsymbol{\gamma_1} + u(1)$$

   - ▷ If the same individual did not take drugs, then his predicted labor market outcome would be

$$y(0) = \psi_0 + (\mathbf{x} - \boldsymbol{\eta})\boldsymbol{\gamma_0} + u(0)$$

   Read: section 2.7 again from JW Chapter 2 to distinguish between ATE and TE (a) under $\tau = y_i(1) - y_i(0)$ i.e. constant treatment effect and (b) different linear effect functions for treatment and control i.e. $\psi_0 \neq \psi_1$

5. Average Treatment effect: without restrictions (a) and (b) i.e. identical linear effect $\psi_0 = \psi_1$, we can write the ATE as:

$$E(te_i) = (\psi_1 - \psi_0) + E\{(\mathbf{x}_i - \boldsymbol{\eta})(\boldsymbol{\gamma_1} - \boldsymbol{\gamma_0}) + u_i(1) - u_i(0)\} = \tau_i$$

where $(\mathbf{x}_i - \boldsymbol{\eta})$ has a zero mean by construction and $u_i(1)$ and $u_i(0)$ has zero mean because they are the errors obtained from conditional expectations.

Unconfoundedness: implies conditional independence where the dummy variable for treatment does not depend on individual level co-variates i.e. $\mathbf{x}_i$

$$w \text{ is independent of } [y(0), y(1)] \text{ conditional on } \mathbf{x}$$

$$\begin{aligned} E(u_i|w_i, \mathbf{x}_i) =& E(u_i(0)|w_i, \mathbf{x}_i) + w_i E\{[u_i(1) - u_i(0)]|w_i, \mathbf{x}_i\} \\ =& E(u_i(0)|\mathbf{x}_i) + E\{[u_i(1) - u_i(0)]|\mathbf{x}_i\} \end{aligned}$$

6. Directly estimating average treatment effect: by regressing y_i on $w_i, x_{i1}, x_{i2}, \ldots$ and on demeaned controls $x_{ik} - \bar{x}_{ik} \cdot w_i$. The estimate of $\hat{\tau}$ on w_i is out ATE from a *restricted regression adjustment.*

   - It is critical to demean the x_j before constructing the interactions in order to obtain the average treatment effect as the coefficient on w_i.
   - The two methods give us exact same estimated of $\hat{\tau}$

7. Consider the following example

Job Training program example: uses data JTRAIN98 and measures the effect of job training in 1997 on wages in 1998 for workers who got the training in comparison with those who didnt. We control for past earnings, education, age and maritial status.

- Simple difference of means estimate (3.74) of -2.05

$$\widehat{earn98} = 10.61 - 2.05 train$$

- Examine the regression with controls (3.75) called the restricted regression estimate of +2.41

$$\begin{aligned}\widehat{earn98} =& 4.67 + 2.41 train + 0.373 earn96 + 0.363 educ \\ & - 0.181 age + 2.48 married\end{aligned}$$

8. Now consider our estimate without any restriction

- ▷ Examine the regression with the un-restricted regression estimate of 3.11

$$\begin{aligned}\widehat{earn98} =& 5.08 + 3.11 train + 0.353 earn96 + 0.378 educ \\ & - 0.196 age + 2.76 married \\ & + 0.133 train \cdot (earn96 - ea\bar{r}n96) \\ & - 0.035 train \cdot (educ - e\bar{d}uc) \\ & + 0.58 train \cdot (age - a\bar{g}e) \\ & - 0.993 train \cdot (married - ma\bar{r}ried)\end{aligned}$$

These estimates prove to provide the most significant effect of the job training program. The t-stat on train is significant. Which model would you choose?

9. *Average Treatment Effect*: demean married and replace the final interaction term with $unmarried \cdot train$ to obtain an estimate of 3.79 as $\hat{\beta}$ of train. Coefficient on married is the same as 0.993. The ATE is 2.797.

10. Spurious effect of treatment is possible unless there are claims to causality.

7.7 Interpreting Regression Results with Discrete Dependent Variables

1. When the dependent variable follows a multi-nominal distribution, then the interpretation follows our usual multiple linear regression, i.e. for a given set of x_js we interpret the predicted value as $\hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots \hat{\beta}_k x_k$ as an estimate of the *conditional expectation function* i.e. $E(y|x_1, x_2, \ldots, x_k)$.

2. The estimates of $\hat{\beta}$ then tell us how the average of y changes over the sample when y takes many discreete values.