Problem Set 3

1.) 2SLS with a multi-level instrument

Assume the setup from lecture, but suppose we have a multilevel instrument. That is, suppose S takes on values in $\{s_0, s_1, \ldots, s_L\}$. We suppose that our treatment T is still a binary random variable. Suppose also that

$$Y(t, s_m) \overset{d}{=} Y(t, s_n) \;\forall\; m, n \in \{0, 1, \ldots, L\}.$$

Then $Y(t, s_m) = Y(t)$. The rest of the setup is identical. We define

$$\mathbb{E}[Y(0)] = \beta_0, \quad \mathbb{E}[Y(1)] = \beta_1,$$

and

$$U_0 = Y(0) - \mathbb{E}[Y(0)], \quad U_1 = Y(1) - \mathbb{E}[Y(1)].$$

$$\text{ATE} = \theta = \beta_1 - \beta_0.$$

We have also the same threshold crossing model, so

$$T = \begin{cases} 1, & V \leq g(S), \\ 0, & \text{otherwise.} \end{cases}$$

We will suppose that $g(s_m) < g(s_n)$ whenever $m < n$.

We will use a set of indicators $\mathbb{1}(S_i = s_l)$ for our instrument set. The population 2SLS coefficient β_{2SLS} is given by $\text{Cov}\left(Y, \hat{T}\right) / \text{Cov}\left(T, \hat{T}\right)$, where $\hat{T}$ is the population predicted values from the first stage.

Problem 1. *Prove that the population coefficient on $\mathbb{1}(S_i = s_l)$ from the first stage is equal to $\mathbb{P}(D = 1 | S = s_l) = g(s_l)$.*

Hint: Consider a regression of T on $\boldsymbol{S}$, where $\boldsymbol{S}$ is the $n \times L{+}1$ matrix encoding assignment of the instrument as characteristic functions. Reorder observations so that the first n_0 rows are observations with $S_i = s_0$, the next n_1 rows are observations with $S_i = s_1$, and so on. Compute $\boldsymbol{S}'\boldsymbol{S}$, and consider the plim of $\left(\frac{1}{n}\boldsymbol{S}'\boldsymbol{S}\right)^{-1}$. Similarly, consider the plim of $\frac{1}{n}\boldsymbol{S}'T$.

Therefore, the population-level 2SLS estimate is gives us the population-level IV estimate with instrument $g(S)$. Then

$$\beta_{\text{2SLS}} = \frac{\text{Cov}(Y, g(S))}{\text{Cov}(T, g(S))}.$$

Problem 2. *Prove that*

$$\text{Cov}(Y, g(S)) = \sum_{l=0}^{L} \left(\sum_{m=1}^{l} [\mathbb{E}[Y|S = s_m] - \mathbb{E}[Y|S = s_{m-1}]] \right) (g(s_l) - \mathbb{E}[g(S)])\mathbb{P}(S = s_l).$$

Hint: Use the law of total expectation. Then, introduce a telescoping sum, noting that $x_0 + \sum_{m=1}^{l} x_m - x_{m-1} = x_m$.

Problem 3. *Define* $\text{LATE}_{m-1}^{m} = \mathbb{E}[Y(1) - Y(0)|g(s_{m-1}) \leq V \leq g(s_m)]$. *Show that*

$$\text{Cov}(Y, g(S)) = \sum_{l=0}^{L} \left(\sum_{m=1}^{l} [g(s_m) - g(s_{m-1})]\text{LATE}_{m-1}^{m} \right) (g(s_l) - \mathbb{E}[g(S)])\mathbb{P}(S = s_l).$$

Hint: Follow the same steps as those used in Gary Chamberlain's notes.

Problem 4. *We will rewrite this expression. Show that*

$$\sum_{l=0}^{L} \left(\sum_{m=1}^{l} [g(s_m) - g(s_{m-1})]\text{LATE}_{m-1}^{m} \right) (g(s_l) - \mathbb{E}[g(S)])\mathbb{P}(S = s_l)$$

$$= \sum_{m=1}^{L} \text{LATE}_{m-1}^{m} \left([g(s_m) - g(s_{m-1})] \sum_{l=m}^{L} (g(s_l) - \mathbb{E}[g(S)])\mathbb{P}(S = s_l) \right).$$

Problem 5. *We now consider the denominator. Show that*

$$\text{Cov}(T, g(S)) = \sum_{m=1}^{L} (g(s_m) - g(s_{m-1})) \sum_{l=m}^{L} (g(s_l) - \mathbb{E}[g(S)])\mathbb{P}(S = s_l).$$

Hint: This should be easy if you have solved 2 and 4.

Problem 6. *Let's put the pieces together. Show that*

$$\beta_{\text{2SLS}} = \sum_{m=1}^{L} w_m \text{LATE}_{m-1}^{m},$$

where

$$w_m = \frac{(g(s_m) - g(s_{m-1})) \sum_{l=m}^{L} (g(s_l) - \mathbb{E}[g(S)])\mathbb{P}(S = s_l)}{\sum_{m=1}^{L} (g(s_m) - g(s_{m-1})) \sum_{l=m}^{L} (g(s_l) - \mathbb{E}[g(S)])\mathbb{P}(S = s_l)}.$$

Note that the weights sum to 1 by construction. Moreover, prove that w_m *is positive for all* $m \in \{1, 2, \ldots, L\}$. *Explain why the weights being positive is a desirable property.*

2.) Inverted Tests and Instrumental Variables

Problem 7. *Consider $Y = D\beta + \varepsilon$, and suppose the variables D are endogenous, but we have valid instruments Z that we can use to estimate β. We have the moment condition $\mathbb{E}[Z\epsilon] = 0$. In the spirit of IVQR (but in the least squares setting), suggest a way to construct confidence intervals for $\hat{\beta}$ based on the moment condition involving Z.*

3.) DML

In this problem, we turn to the example in Chernozhukov et al. (2024) and estimate the effect of gun ownership on homicide rate. We supply their code to construct the full set of controls. The variable "logghomr" is the outcome variable, and the variable "logfssl" is the treatment. All other variables will be used as controls.

Problem 8. *What is the uncontrolled predictive effect of gun ownership on the homicide rate? Under what identification assumptions is this quantity equal to the causal effect?*

Problem 9. *Now, we'll incorporate control variables in our linear regression. Using all of the controls, estimate the predictive effect of gun ownership on the homicide rate. Under what identification assumptions is this quantity equal to the causal effect?*

Problem 10. *Next, we'll take the DML approach to incorporating controls. Use cross-fitting to fit a neural network that estimates the expectation of the outcome and treatment conditional on our controls. Play with the hyperparameters to fit several neural networks. Show the mean square errors of the predictions of the outcome and treatment for each of these models. Also, compare these to the mean square errors of predictions of the outcome and treatment by linear models (obtained by regressing "logghomr" and "logfssl" on controls). Based on these mean square errors, which model(s) will you use to construct predicted values of "logghomr" and "logfssl?"*

Problem 11. *Given your constructed predicted values computed in 10, obtain the predictive effect of gun ownership on the homicide rate. Under what identification assumptions is this quantity equal to the causal effect? Explain the pattern in the predictive effect you observe as you move from the uncontrolled setting to the setting with controls in a linear model to the DML setting.*

Next, we'll repeat similar steps using the health care dataset we worked with in Problem Set 2.

Problem 12. *Include the square of age and the square of weightgain as controls. Show the uncontrolled predictive effect of smoking on birthweight. Then, jump straight to using* *neural networks to predict birthweight and smoking as a function of controls (noting that smoking takes on a discrete set of values). Compare the mean square errors of your models for birthweight and binary cross-entropies of your models for smoking, and construct your predicted values accordingly. Estimate the predictive effect of maternal smoking on birthweight. How does it compare to the uncontrolled predictive effect? How would you explain this discrepancy?*

References

Chernozhukov, Victor, Christian Hansen, Nathan Kallus, Martin Spindler, and Vasilis Syrgkanis. 2024. "Applied causal inference powered by ML and AI." *arXiv preprint arXiv:2403.02467*.