Ec141, Spring 2019

Professor Bryan Graham

Problem Set 4

Due: April 9th, 2019

Problem sets are due at 5PM in the GSIs mailbox. You may work in groups, but each student should turn in their own write-up (including a narrated/commented and executed Jupyter Notebook). Please use markdown boxes within your Jupyter notebook for narrative answers to the questions below.

P1: (Instrumental Variables and 'Omitted Variable Bias') Let Y_2 denote the logarithm of GDP per capita; $X = (X_1', X_2')'$ is a 4×1 vector of variables which can be divided into two parts, the first part, X_1, is a 3×1 vector which includes a constant, a dummy for whether the country in question is a former colony, C, and a dummy for whether the country in question is located in the geographic tropics, T, (i.e., between the Tropics of Cancer and Capricorn); X_2 is a 1×1 vector composed of the interaction of C and T (i.e., $X_2 = C \cdot T$); Y_1 is a measure of institutional quality (scaled to lie on the unit interval, with a 1 indicating 'good' institutions and a zero 'bad' institutions), A_i represents an index of additional, but *unobserved*, variables which are also correlated with GDP (e.g., climate, culture, and historical experiences).

[a] Consider the following regression function for the logarithm of GDP per capita:

$$\begin{aligned} \mathbb{E}\left[Y_{i2}|X_{i1}, X_{i2}, Y_{i1}, A_i\right] &= \mathbb{E}\left[Y_{i2}|X_{i1}, Y_{i1}, A_i\right] \\ &= X_{i1}'\alpha_1 + \gamma Y_{i1} + \alpha_2 A_i. \end{aligned} \tag{1}$$

In what sense in (1) restrictive? Comment upon the substantive aspects of any restriction? Briefly describe how you would test this restriction if you observed A_i. What does γ measure?

[b] Under what conditions does the coefficient on Y_{i1} in the least squares regression of Y_{i2} onto X_{i1} and Y_{i1} provide a consistent estimate of γ?

[c] The CEF of institutional quality, Y_1, given colonial status, geographical location, their interaction (i.e., $X = (X_1', X_2')' = (1, C, T, C \cdot T)'$) and A equals

$$\mathbb{E}\left[Y_{i1}|X_{i1}, X_{i2}, A_i\right] = X_{i1}'\alpha_3 + X_{i2}'\pi_1 + \alpha_4 A_i. \tag{2}$$

Is (2) restrictive in any way?

[d] Assume that

$$\begin{aligned} \mathbb{E}^*\left[A_i \mid X_{i1}, X_{i2}\right] &= \mathbb{E}^*\left[A_i \mid X_{i1}\right] \\ &= X_{i1}'\alpha_5. \end{aligned} \tag{3}$$

Comment on (3).

[e] Using (2) and (3) find λ_1 such that

$$\mathbb{E}^*\left[Y_{i1} \mid X_{i1}, X_{i2}\right] = X_{i1}'\lambda_1 + X_{i2}'\pi_1. \tag{4}$$

[f] Using (1), (2) and (3) find λ_2 and π_2 such that

$$\mathbb{E}^*\left[Y_{i2} \mid X_{i1}, X_{i2}\right] = X_{i1}'\lambda_2 + X_{i2}'\pi_2. \tag{5}$$

[g] You have available OLS estimates of $\widehat{\pi}_1$ and $\widehat{\pi}_2$; suggest an estimator for γ. Does the probability limit of your estimator of γ exist if $\widehat{\pi}_1 \xrightarrow{p} 0$?

[h] Show that π_1 and π_2 equal

$$\begin{aligned} \pi_1 &= \mathbb{E}\left[Y_1 \mid C=1, T=1\right] - \mathbb{E}\left[Y_1 \mid C=0, T=1\right] - \left\{\mathbb{E}\left[Y_1 \mid C=1, T=0\right] - \mathbb{E}\left[Y_1 \mid C=0, T=0\right]\right\} \\ \pi_2 &= \mathbb{E}\left[Y_2 \mid C=1, T=1\right] - \mathbb{E}\left[Y_2 \mid C=0, T=1\right] - \left\{\mathbb{E}\left[Y_2 \mid C=1, T=0\right] - \mathbb{E}\left[Y_2 \mid C=0, T=0\right]\right\}. \end{aligned}$$

[i] Show that, given (1), (2) and (3),

$$\psi\left(Z_i, \theta\right) = X_i\left(Y_{i2} - X_{i1}'\widetilde{\alpha}_1 - \gamma Y_{i1}\right), \tag{6}$$

has expectation zero, where $\widetilde{\alpha}_1 = \alpha_1 + \alpha_2\alpha_5$ and $\theta = (\widetilde{\alpha}_1', \gamma)'$. You may assume that A_i is mean zero. Suggest an estimator based upon the population restriction that $E\left[\psi\left(Z_i, \theta\right)\right] = 0$.

[j] Calculate $\mathbb{E}\left[Y_{i2} \mid C=c, T=t\right]$ using (1) for each of the four possible combinations of c and t. Use the resulting four equations to solve for γ. Suggest an estimator for γ based on your answer.

[k] Consider the infeasible least squares regression of Y_{i2} onto X_{i1} and $\mathbb{E}^*\left[Y_{i1} \mid X_{i1}, X_{i2}\right]$. Show that the coefficient on $\mathbb{E}^*\left[Y_{i1} \mid X_{i1}, X_{i2}\right]$ would be consistent for γ. Explain the intuition behind this estimator and suggest a feasible version of it.

P2: **(Empirical Application)** The file `colonies.out` is available on the course web page. The file includes 97 observations of the log of GDP per capita (`lgdp`), a colony dummy (`colony`), a tropics dummy (`tropics`), their interaction (`col_trop`) and a measure of institutional quality (`inst`). These variables relate in the obvious way to Y_1, Y_2, X_1 and X_2 as discussed in Problem 1.

[a] Calculate the least squares regression fit of `lgdp` onto a constant `colony, tropics` and `inst`.

[b] Calculate the conditional sample means of `lgdp` and `inst` for each of the four possible combinations of `colony` and `tropics`.

[c] Use your results in (b) to directly compute estimates of π_1, π_2 and γ.

[d] Calculate the least squares regression fit of `lgdp` onto a constant, `colony, tropics` and `col_trop`. Also calculate the least squares regression fit of `inst` onto a constant, `colony, tropics` and `col_trop`. Compare your estimates of π_1 and π_2 from these regressions to those from (c). Test the null hypotheses that $\pi_1 = 0$ and $\pi_2 = 0$. Use your results to form an estimate of γ, how does this estimate compare with the one from (c)?

[e] Compute the instrumental variables estimate of γ. Compare this estimate with those from (c) and (d). Construct a 95 percent confidence interval for your estimate of γ.

[f] Calculate the fitted values of `inst` from your least squares fit in (d). Calculate the least squares fit of `lgdp` onto a constant, `colony, tropics` and $\widehat{\texttt{inst}}$. In words, explain how this least squares fit differs from the one you calculated in (a).

[g] Comment on the economic significance of your analysis.