Jeff Grogger
Harris School

Winter 2022
University of Chicago

PP 420: Problem Set 4
(due Tuesday, March 8, at 11:59 pm via Canvas)

A. Analytical problems

Problem 1: Suppose that *x* is an endogenous regressor, and *z* is a potential instrument for it. Can you use the regression

$$y \;\; = \beta_1 + \beta_2 x + \beta_3 z + \epsilon$$

to test for the exogeneity of z? Provide a proof for whatever answer you come to.

Problem 2: Let the model be given by

$$y = X_1\beta_1 + \beta_K x_K + \varepsilon$$

where X_1 is an nx(K-1) matrix of exogenous variables and x_K is an endogenous regressor. Assume you have an nxL matrix of valid instruments, where L ≥K. Assume further that $E(\varepsilon\varepsilon'|Z) = \sigma^2 I$ and that the data are centered. Show that the asymptotic variance of $b_{2SLS,K}$ can be consistently estimated by

$$\widehat{V}(b_{2SLS,K}) = \frac{\hat{\sigma}^2}{(n-1)\hat{V}(x_K)(1-\hat{R}_K^2)R_{Kz}^2},$$

where $\hat{V}(x_K)$ is the sample variance of x_K; R_{Kz}^2 is the R^2 from a regression of x_K on Z; $\hat{R}_K^2$ is the R^2 from the regression of $\hat{x}_K$ on X_1, where $\hat{x}_K$ gives the predicted values from a regression of x_K on Z; and $\hat{\sigma}^2$ is a consistent estimator of σ^2. Identify the five factors that contribute to the precision of the 2SLS estimator and provide some intuition for them.

Problem 3: Let

$$y = X_1\beta_1 + \beta_K x_K + \varepsilon$$

where $E\left(\frac{X_1'\varepsilon}{n}\right) = 0$,

$E\left(\frac{x_K'\varepsilon}{n}\right) \neq 0$,

z is a valid instrument for x_K , and

$M_1 = I - X_1(X_1'X_1)^{-1}X_1$.

Show which of the following IV regressions yield a consistent estimate of β_K:

(a) $y = \beta_K M_1 x_K + \varepsilon_1$,
with z as the instrument;

(b) $M_1 y = \beta_K x_K + \varepsilon_1$,
with z as the instrument;
(c) $y = \beta_K x_K + \varepsilon_1$,
with $M_1 z$ as the instrument;
(d) $M_1 y = \beta_K x_K + \varepsilon_1$,
with $M_1 z$ as the instrument;
(e) $y = \beta_K M_1 x_K + \varepsilon_1$,
with $M_1 z$ as the instrument.
Explain what you have shown.

Problem 4. Show that if there are no never-takers, then the LATE (Local Average Treatment Effect) is equal to the ATN (Average Treatment Effect on the Non-treated). You may assume dichotomous treatment and instrument, and no other regressors.

B. Computational Problems

These questions all make use of pp420_data.dta.

(a) Labor economists typically view earnings as a function of experience. Since experience is the sum of past labor supply, however, it may be correlated with the current earnings disturbance. It is thus a candidate for instrumental variables.

(i) Construct experience as the running sum of employment, where the sample member is employed in a quarter if her earnings are greater than zero in that quarter.

(ii) Regress earnings on experience, higrade, and black. Correct the OLS standard errors for the panel structure of the data throughout the exercise. Based on the OLS coefficient, how much is an additional quarter's experience worth?

(iii) As an instrument for experience, use the treat variable. Treat equals one for people assigned to the treatment group and zero for those assigned to the control group. Random assignment implies that treat is uncorrelated with the earnings disturbance. Since the experiment was designed to increase employment, it should also increase experience. Estimate the first-stage regression. What is the partial F statistic for the instrument? What is the Montiel-Pflueger F and how does it compare to its critical values?

(iv) Now estimate the earnings equation from (ii) by IV, using treat as the instrument for experience. Based on the IV coefficient, how much is an additional quarter's experience worth? Test the hypothesis that the true effect of experience is zero.

(v) Now compute a Hausman test for the endogeneity of experience. What do you conclude about the effect of experience on earnings?