ECON 710A - Problem Set 2

Alex von Hafften*

2/9/2020

1. Suppose $(Y, X, Z)'$ is a vector of random variables such that $Y = \beta_0 + X\beta_1 + U$, $E[U|Z] = 2$ where $Cov(Z, X) \neq 0$ and $E[Y^2 + X^2 + Z^2] < \infty$. Additionally, let $\{(Y_i, X_i, Z_i)'\}_{i=1}^n$ be a random sample from the model with $\hat{Cov}(Z, X) \neq 0$. Recall, the definition from lecture 3 $\hat{\beta}_1^{IV} = \frac{\hat{Cov}(Z,Y)}{\hat{Cov}(Z,X)} = \frac{\frac{1}{n}\sum_{i=1}^n (Z_i - \bar{Z}_n)(Y_i - \bar{Y}_n)}{\frac{1}{n}\sum_{i=1}^n (Z_i - \bar{Z}_n)(X_i - \bar{X}_n)}$ and $\hat{\beta}_0^{IV} = \bar{Y} - \bar{X}\hat{\beta}_1^{IV}$.

(i) Does $\hat{\beta}_1^{IV} \to_p \beta_1$?

...

(ii) Does $\hat{\beta}_0^{IV} \to_p \beta_0$?

...

2. Consider the simulation model $Y = \beta_0 + X\beta_1 + U$ and $X = \pi_0 + Z\pi_1 + V$ with $E[U|Z] = E[V|Z] = 0$ where $E[Y^2 + X^2 + Z^2] < \infty$. Additionally, let $\{(Y_i, X_i, Z_i)'\}_{i=1}^n$ be a random sample from this model with $\frac{1}{n}\sum_{i=1}^n (Z_i - \bar{Z}_n)^2 > 0$.

(i) Under what conditions (on $\beta_0, \beta_1, \pi_0, \pi_1$) is Z a valid instrument for X?

...

(ii) Show that $Y = \gamma_0 + Z\gamma_1 + \varepsilon$ with $E[\varepsilon|Z] = 0$ where γ_0, γ_1, and ε are some functions of $\beta_0, \beta_1, \pi_0, \pi_1$, U, and V. In particular show that $\gamma_1 = \pi_1\beta_1$.

...

(iii) Let $\hat{\gamma}_1$ and $\hat{\pi}_1$ denote the OLS estimators of γ_1 and π_1, respectively. The ratio $\hat{\gamma}_1/\hat{\pi}_1$ is called the "indirect least squares" estimator of β_1. How does it compare to the IV estimator of β_1 that uses Z as an instrument for X?

...

(iv) Show that $Y = \delta_0 + X\delta_1 + V\delta_2 + \xi$, $Cov(X, \xi) = Cov(V, \xi) = 0$ where $\delta_0, \delta_1, \delta_2$, and ξ are some functions of $\beta_0, \beta_1, Cov(U, V), Var(V), U, V$. In particular, show that $\delta_1 = \beta_1$.

...

(v) Let $\hat{V}_i = X_i - \hat{\pi}_0 - Z_i\hat{\pi}_1$ where $\hat{\pi}_0$ is the OLS estimator of π_0. Furthermore, let $\hat{\delta}_1$ be the OLS estimator from a regression of Y_i on $(1, X_i, \hat{V}_i)$; this estimator is called the "control variable" estimator. How does it compare to the IV estimator of β_1 that uses Z as an instrument for X?

...

3. The paper "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size" by J. Angrist and W. Evans (AE98) considers labor supply responses to the number of children in the household. They consider models of the form $Y = \beta_0 + X_1\beta_1 + X_2'\beta_2 + U$ where Y is some measure

*I worked on this problem set with a study group of Michael Nattinger, Andrew Smith, and Ryan Mather. I also discussed problems with Emily Case, Sarah Bass, and Danny Edgel.

of the parents' labor supply, X_1 is a binary variable indicating "more than 2 children in the household", and X_2 is a vector of (assumed) exogenous variables that control for race, age, and whether any of the children is a boy. For the next two questions we will focus on the case where Y is a binary variable indicating whether the mother worked during the year.

(i) Provide a causal interpretation of β_1.

...

(ii) Discuss why or why not you think that X_1 could be endogenous. If you think it is, discuss the direction of the (conditional) bias in OLS relative to the causal parameter.

...

(iii) Repeat the previous two questions when Y is a binary variable indicating whether the husband worked during the year.

...

(iv) Discuss why or why not you think that the binary variable Z_1 which indicates whether the two first children are of the same sex is a valid instrument for X_1.

...

(v) Estimate the reduced form regression of X_1 on Z_1 and X_2, do the results suggest that Z_1 is relevant?

...

(vi) (Attempt to) replicate the first three rows of Table 7 columns 1, 2, 5, 7, and 8 in AE98. Interpret the empirical results in relation to your discussion of the previous questions.

...