

## Problem 1

### Problem 1

Stock and Watson, Exercises 12.4, 12.5, 12.8, 12.9, \*E12.1

- (a) (12.4) Consider TLS estimation with a single included endogenous variable and a single instrument. Then the predicted value from the first-stage regression is

$$\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i.$$

Use the definition of the sample variance and covariance to show that

$$s_{\hat{X}Y} = \hat{\pi}_1 s_{ZY} \quad \text{and} \quad s_{\hat{X}}^2 = \hat{\pi}_1^2 s_Z^2.$$

Use this result to fill in the steps of the derivation of Equation (12.4) in Appendix 12.2.

SOLUTION: We see that

$$\begin{aligned} s_{\hat{X}Y} &= \frac{1}{n} \sum (\hat{X}_i - \bar{\hat{X}})(Y_i - \bar{Y}) \\ &= \frac{1}{n} \sum \hat{X}_i (Y_i - \bar{Y}) \\ &= \frac{1}{n} \sum (\hat{\pi}_0 + \hat{\pi}_1 Z_i)(Y_i - \bar{Y}) \\ &= \hat{\pi}_0 \frac{1}{n} \sum (Y_i - \bar{Y}) + \hat{\pi}_1 \frac{1}{n} \sum Z_i (Y_i - \bar{Y}) \\ &= \hat{\pi}_1 \frac{1}{n} \sum (Z_i - \bar{Z})(Y_i - \bar{Y}) \\ &= \hat{\pi}_1 s_{ZY} \end{aligned}$$

and

$$\begin{aligned} s_{\hat{X}}^2 &= \frac{1}{n} \sum (\hat{X}_i - \bar{\hat{X}})^2 \\ &= \frac{1}{n} \sum (\hat{X}_i - \bar{\hat{X}}) \hat{X}_i \\ &= \frac{1}{n} \sum (\hat{\pi}_0 + \hat{\pi}_1 Z_i - \frac{1}{n} \sum (\hat{\pi}_0 + \hat{\pi}_1 Z_i)) (\hat{\pi}_0 + \hat{\pi}_1 Z_i) \\ &= \frac{1}{n} \sum (\hat{\pi}_1 Z_i - \hat{\pi}_1 \bar{Z}) (\hat{\pi}_0 + \hat{\pi}_1 Z_i) \\ &= \frac{1}{n} \sum (\hat{\pi}_1 Z_i - \hat{\pi}_1 \bar{Z})^2 \end{aligned}$$

$$\begin{aligned}
&= \hat{\pi}_1^2 \frac{1}{n} \sum (Z_i - \bar{Z})^2 \\
&= (\hat{\pi}_1)^2 s_Z^2
\end{aligned}$$

To see that  $\hat{\beta}_1^{2SLS} = \hat{\beta}_1^{IV}$ ,

$$\begin{aligned}
\hat{\beta}_1^{2SLS} &= \frac{s_{\hat{X}Y}}{s_{\hat{X}}^2} \\
&= \frac{\hat{\pi}_1 s_{ZY}}{(\hat{\pi}_1)^2 s_Z^2} \\
&= \frac{s_{ZY}}{\hat{\pi}_1 s_Z^2}
\end{aligned}$$

Hence, it suffice to show that the denominator is  $s_{ZX}^2$  :

$$\hat{\pi}_1 s_Z^2 = \frac{s_{ZX}}{s_Z^2} s_Z^2 = s_{ZX}$$



(b) Consider the IV regression model

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i,$$

where  $X_i$  is correlated with  $u_i$ , and  $Z_i$  is an instrument. Suppose that the first three assumptions in Key Concept 12.4 are satisfied. Which IV assumption is **not** satisfied when

a.  $Z_i$  is independent of  $(Y_i, X_i, W_i)$ ?

SOLUTION: Relevance



b.  $Z_i = W_i$ ?

SOLUTION:  $Z$  cannot affect  $Y$  thru anything other than  $X$ .



c.  $W_i = 1$  for all  $i$ ?

SOLUTION: Co-linearity



d.  $Z_i = X_i$ ?

SOLUTION: Exogeneity



(c) Consider a product market with a supply function  $Q_i^s = \beta_0 + \beta_1 P_i + u_i^s$ , a demand function  $Q_i^d = \gamma_0 + u_i^d$ , and a market equilibrium condition  $Q_i^s = Q_i^d$ , where  $u_i^s$  and  $u_i^d$  are mutually independent i.i.d. random variables, both with a mean of 0.

- (i) Show that  $P_i$  and  $u_i^s$  are correlated.

SOLUTION:

$$\begin{aligned}\text{Cov}(P, u^s) &= \text{Cov}\left(\frac{1}{\beta_1}(\gamma_0 + u^d - \beta_0 - u^s), u^s\right) \\ &= \frac{1}{\beta_1} \text{Cov}(u^d - u^s, u^s) \\ &= -\frac{1}{\beta_1} \text{Var}(u^s) \\ &\neq 0\end{aligned}$$



- (ii) Show that the OLS estimator of  $\beta_1$  is inconsistent.

SOLUTION: We know that

$$\hat{\beta}_1 = \beta_1 + \frac{\sum (P_i - \bar{P}) u_i^s}{\sum (P_i - \bar{P})^2} \xrightarrow{\mathbb{P}} \beta_1 + \frac{\text{Cov}(P, u^s)}{\text{Var}(P)}$$

yikes! We showed in (i) that the right term isn't zero.



- (iii) How would you estimate  $\beta_0, \beta_1$ , and  $\gamma_0$ ?

SOLUTION: With instrument variables! Introduce a supply shifter.




- (d) A researcher is interested in the effect of military service on human capital. He collects data from a random sample of 4000 workers aged 40 and runs the OLS regression  $Y_i = \beta_0 + \beta_1 X_i + u_i$ , where  $Y_i$  is a worker's annual earnings and  $X_i$  is a binary variable that is equal to 1 if the person served in the military and is equal to 0 otherwise.

- (i) Explain why the OLS estimates are likely to be unreliable. (Hint: Which variables are omitted from the regression? Are they correlated with military service?)

SOLUTION: There might be something like education thrown into the mix of  $U$  that is omitted, and since someone with military experience might be less likely to attend higher education, they are definitely correlated.



- (ii) During the Vietnam War, there was a draft in which priority for the draft was determined by a national lottery. (The days of the year were randomly reordered 1 through 365. Those with birth dates ordered first were drafted before those with birth dates ordered second, and so forth.) Explain how the lottery might be used as an instrument to estimate the effect of military service on earnings. (For more about this issue, see Joshua D. Angrist (1990).)

SOLUTION: The lottery,  $Z$ , can be used because it directly affects  $Z$  (relevance) and it is exogenous to  $U$ . 

## Problem 2

Suppose the first stage regression model:

$$X = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 W + V$$

where  $X$  is an endogenous variable,  $Z_1$  and  $Z_2$  are instruments, and  $W$  is an exogenous control. The reported  $R^2$  of this regression is 0.8 and the sample size is 100.

Also, consider the model:

$$X = \pi_0 + \pi_3 W + V$$

where the variables are defined as in the previous model, assuming  $\pi_1$  and  $\pi_2$  equal to zero. The reported  $R^2$  is 0.79 and the sample size is 100.

- (a) Test at the 5% significance level the joint hypothesis that  $\pi_1 = \pi_2 = 0$ . You can find the critical value using the F-table. If the desired degrees of freedom do not appear in the table, use the closest values.

SOLUTION: Yes Sir!

$$F = \frac{\frac{1}{q}(R_{ur}^2 - R_r^2)}{\frac{1}{n-K_{ur}-1}(1 - R_{ur}^2)} = \frac{\frac{1}{2}(0.01)}{\frac{1}{100-4-1}(0.2)} = 2.375$$

this implies a critical value of about 3.07



- (b) Test whether the instruments are weak using the rule of thumb.

SOLUTION:  $2.375 < 10$ , thus WEAK



- (c) Discuss why weak instruments are problematic.

SOLUTION: Because  $SE(\hat{\beta})$  blows tf up and you can't do any good asymptotic results.



### Problem 3

Describe the fundamental problem of causal inference.

SOLUTION: Ok, so imagine you're me and I'm me. There's two me! Sounds like a fundamental problem. I can just kill myself now.



## Problem 4

Consider the following expressions:

$$E[Y|D = 1] - E[Y|D = 0] = ATT + (E[Y_0|D = 1] - E[Y_0|D = 0])$$

$$E[Y|D = 1] - E[Y|D = 0] = ATU + (E[Y_1|D = 1] - E[Y_1|D = 0])$$

$$ATE = \Pr(D = 1) \cdot ATT + \Pr(D = 0) \cdot ATU$$

Discuss the necessary conditions such that the naive estimator  $E[Y|D = 1] - E[Y|D = 0]$  coincides with the ATT, ATU, and ATE.

SOLUTION: In an experiment with randomization such that  $y_0^i, y_1^i \perp D^i$ , we see that

$$SB_0 = \mathbb{E}[y_0^i | D_i = 1] - \mathbb{E}[y_0^i | D_i = 0] = \mathbb{E}[y_0^i] - \mathbb{E}[y_0^i] = 0$$

and same for  $SB_1$ .

Moreover,

$$\begin{aligned} ATE &= \mathbb{E}[y_1^i - y_0^i] \\ &= \mathbb{E}[\mathbb{E}[y_1^i - y_0^i | D^i]] \\ &= p(ATT) + (1 - p)(ATU) \\ &= p(\theta - SB_0) + (1 - p)(\theta - SB_1) \\ &= \theta \end{aligned}$$

Thus, in a randomized experiment,  $ATT = ATU = ATE = \theta$

