PS7 Practice Problems : SW 12.2

12.5

12.7

12.8

12.10

Empirical 12.1 (see R Notes)

12.2 Consider the regression model with a single regressor: $Y_i = \beta_0 + \beta_1 X_i + u_i$. Suppose the least squares assumptions in Key Concept 4.3 are satisfied.

a. Show that X_i is a valid instrument. That is, show that Key Concept 12.3 is satisfied with $Z_i = X_i$.

> **KEY CONCEPT 12.3 — The Two Conditions for Valid Instruments**
>
> A set of m instruments $Z_{1i}, \ldots, Z_{mi}$ must satisfy the following two conditions to be valid:
>
> **1. Instrument Relevance**
>
> - *In general*, let $\hat{X}^*_{1i}$ be the predicted value of X_{1i} from the population regression of X_{1i} on the instruments (Z's) and the included exogenous regressors (W's), and let "1" denote the constant regressor that takes on the value 1 for all observations. Then $(\hat{X}^*_{1i}, \ldots, \hat{X}^*_{ki}, W_{1i}, \ldots, W_{ri}, 1)$ are not perfectly multicollinear.
> - *If there is only one X*, then for the previous condition to hold, at least one Z must have a nonzero coefficient in the population regression of X on the Z's and the W's.
>
> **2. Instrument Exogeneity**
>
> The instruments are uncorrelated with the error term; that is, $\text{corr}(Z_{1i}, u_i) = 0, \ldots, \text{corr}(Z_{mi}, u_i) = 0$.

Special case : one regressor

checking whether it is a valid instrument for itself

The first stage regression is X, the variable being instrumented for on X, the candidate instrument

Relevance : at least one Z w/ nonzero coefficient in the first-stage regression :

$$X = \gamma_1 X + e$$

Obviously $\hat{\gamma}_1 = 1$ $\therefore$ relevant instrument

Exogeneity : $\text{corr}(Z, u_i) = 0$

We are given that LSAs apply $\therefore$ $\text{Corr}(X, u) = 0$

$\Rightarrow \text{Corr}(Z, u) = 0$

$\therefore$ exogenous instrument

$\therefore$ X is a valid instrument for X

b. Show that the IV regression assumptions in Key Concept 12.4 are satisfied with this choice of Z_i.

c. Show that the IV estimator constructed using $Z_i = X_i$ is identical to the OLS estimator.

> **KEY CONCEPT 12.4 — The IV Regression Assumptions**
>
> The variables and errors in the IV regression model in Key Concept 12.1 satisfy the following:
>
> 1. $E(u_i \mid W_{1i}, \ldots, W_{ri}) = 0$;
> 2. $(X_{1i}, \ldots, X_{ki}, W_{1i}, \ldots, W_{ri}, Z_{1i}, \ldots, Z_{mi}, Y_i)$ are i.i.d. draws from their joint distribution;
> 3. Large outliers are unlikely: The X's, W's, Z's, and Y have nonzero finite fourth moments; and
> 4. The two conditions for a valid instrument in Key Concept 12.3 hold.

b. 1. There are no W's since one regressor

$\therefore E[u_i \mid W_{1i}, \ldots, W_{ri}]$

$= E[u_i]$

$= 0$

2. $(X_{1i}, \ldots, X_{ki}, W_{1i}, \ldots, W_{ri}, Z_{1i}, \ldots, Z_{mi}, Y_i)$

$= (X_i, Y_i)$

These are iid draws from the LSA assumptions

3. This is also an LSA assumption

4. We've shown this holds in part a

c. 2SLS : ① regress $X_i = \pi_0 + \pi_1 Z_i + e_i$ by OLS

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

② regress $Y_i = \beta_0 + \beta_1 \hat{X}_i + u_i$ by OLS

estimate $\hat{\beta}_1$

We know OLS estimator in ② of β_1 is

$$\hat{\beta}_1^{2SLS} = \frac{\text{Cov}(Z, Y)}{\text{Cov}(Z, X)} = \frac{\text{Cov}(\hat{X}, Y)}{\text{Var}(\hat{X})}$$

What is $\hat{X}$?

In regression ①

$$X_i = \underbrace{\pi_0}_{0} + \underbrace{\pi_1}_{1} X_i + \underbrace{e_i}_{0}$$

$\Rightarrow \hat{X}_i = 0 + (1) X_i = X_i$

$$\Rightarrow \hat{\beta}_1^{2SLS} = \frac{\text{Cov}(\hat{X}, Y)}{\text{Cov}(\hat{X}, X)} = \frac{\text{Cov}(\hat{X}, Y)}{\text{Var}(\hat{X})} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \hat{\beta}_1^{OLS}$$

12.5 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)?

b. $Z_i = W_i$?

c. $W_i = 1$ for all i?

d. $Z_i = X_i$?

KEY CONCEPT 12.4

The IV Regression Assumptions

The variables and errors in the IV regression model in Key Concept 12.1 satisfy the following:

1. $E(u_i \mid W_{1i}, \ldots, W_{ri}) = 0$;
2. $(X_{1i}, \ldots, X_{ki}, W_{1i}, \ldots, W_{ri}, Z_{1i}, \ldots, Z_{mi}, Y_i)$ are i.i.d. draws from their joint distribution;
3. Large outliers are unlikely: The X's, W's, Z's, and Y have nonzero finite fourth moments; and
4. The two conditions for a valid instrument in Key Concept 12.3 hold.

a. If Z_i is independent of X_i then the first stage regression will not have a non-zero coefficient on any of the instruments $\therefore$ not relevant

b. The first stage regression will be

$X_i = \pi_0 + \pi_1 Z_i + \pi_2 W_i + e_i$

$= \pi_0 + \pi_1 Z_i + \pi_2 Z_i + e_i$

$\therefore$ multicollinearity

OR

$\pi_0 = 0$, $e_i = 0$ and

$\hat{X}_i$ is a linear combination of $W_i = Z_i$

$\Rightarrow$ not relevant

c. $W_i = 1$ $\Rightarrow$ perfectly collinear with the intercept

$\Rightarrow$ not relevant

d. $Z_i = X_i$ $\Rightarrow$ not exogenous

12.7 In an IV regression model with one regressor, X_i, and two instruments, Z_{1i} and Z_{2i}, the value of the J-statistic is $J = 18.2$.

a. Does this suggest that $E(u_i \mid Z_{1i}, Z_{2i}) \neq 0$? Explain.

b. Does this suggest that $E(u_i \mid Z_{1i}) \neq 0$? Explain.

a. How to interpret J-statistic?

This statistic is used to evaluate an overidentification test. It is distributed χ^2_{m-k} so has a 1% critical value of 6.63 if $m-k=1$. It tests the null hypothesis that in the regression

$$\hat{u}_i^{2SLS} = \delta_0 + \delta_1 Z_{1i} + \dots + \delta_m Z_{mi} + \delta_{m+1} W_{1i} + \dots + \delta_{m+r} W_{ri} + e_i,$$

$H_0: \delta_1 = \dots = \delta_m = 0$ (all instruments are exogenous)

In this case instrument exogeneity, $E[u_i \mid Z_{1i}, Z_{2i}] = 0$, is rejected

b. It only tells us about the joint test over all instruments, not the exogeneity of any specific instrument.

12.8 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^s are mutually independent i.i.d. random variables, both with a mean of 0.

a. Show that P_i and u_i^s are correlated.

b. Show that the OLS estimator of β_1 is inconsistent.

c. How would you estimate β_0, β_1, and γ_0?

a. Solving for P :

$$P = (Q - \beta_0 - u_i^s)/\beta_1$$

$$= \frac{\gamma_0 + u_i^d - \beta_0 - u_i^s}{\beta_1}$$

$$\Rightarrow \text{Cov}(P, u_i^s) = \frac{-\sigma_{u^s}^2}{\beta_1} \neq 0 \quad \therefore \text{correlated}$$

b. Follows from part a $\Rightarrow$ violation of LSAs

c. Need IV to estimate β_0, β_1

$\hat{\gamma}_0$ is just the sample mean $\bar{Q}_i$

12.10 Consider the IV regression model $Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i$, where Z_i is an instrument. Suppose data on W_i are not available and the model is estimated omitting W_i from the regression.

a. Suppose Z_i and W_i are uncorrelated. Is the IV estimator consistent?

b. Suppose Z_i and W_i are correlated. Is the IV estimator consistent?

$$\hat{\beta}_1^{2SLS} \xrightarrow{p} \frac{\text{Cov}(Z, Y)}{\text{Cov}(Z, X)}$$

$$= \frac{\text{Cov}(Z, \beta_0 + \beta_1 X + \beta_2 W + u)}{\text{Cov}(Z, X)}$$

$$= \frac{\beta_1 \text{Cov}(Z_i, X) + \beta_2 \text{Cov}(Z, W)}{\text{Cov}(Z, X)}$$

$$= \beta_1 \text{ if } \text{Cov}(Z, W) \neq 0$$