Extra Questions Applied Econometrics

Week 2: IV and LATE

If you are asked to derive something, give all intermediate steps also. Do not answer questions with a "yes" or "no" only, but carefully motivate your answer. If you are asked to conduct a statistical test, you should use a 5% significance level.

Question 1

Say you are hired by KLM airlines as a data analyst to determine the effect of advertising. You quickly realize that using observational data of past advertising may be problematic because advertising campaigns were mostly run in times of low demand. You therefore convince the board of directors to conduct an experiment using a between country design, whereby you randomly implement extra television ads in some countries and others not. The extra ads (*Adextra*, either 0 or 1) are shown for half a year, from October to March, and you want to see what the effect is on subsequent demand for flights (*Flights*, total number of seats sold measured in millions) in the given country i, in the six months following the experiment. After running the experiment and collecting the data from all $n = 105$ countries, you get the following regression output (robust standard errors in parenthesis)

$$\widehat{Flights}_i = \underset{(0.018)}{0.34} + \underset{(0.034)}{0.075}\, Adextra_i \tag{1}$$

Now answer the following:

a) Given that an additional flight-seat sold yields €87 in profits, and the extra ads have a cost of €5 million on average, calculate the extra profit KLM makes due to the campaign. Is the difference significant?

b) Your sample contains countries from 5 different continents. Describe how you could test if the effect of the ad campaign is the same in all continents.

Although knowing the effect of the campaign is interesting in itself, it might be even more informative to know the effect of each additional minute of television advertising on demand. You therefore also measure the total ad-broadcasting time from October to March in each country (*Adtime*, measured in hours). The campaign should increase the total number of ads in each country, so you could use it as an instrument for ad-time. Below are the TSLS estimates and corresponding first-stage estimates (robust standard errors in parenthesis)

$$\widehat{Flights}_i = \underset{(0.23)}{-0.17} + \underset{(0.040)}{0.094}\, Adtime_i \tag{2}$$

$$\widehat{Adtime}_i = \underset{(0.06)}{5.48} + \underset{(0.085)}{0.81}\, Adextra_i \tag{3}$$

Now answer the following:

c) Local KLM branches in the treatment group are responsible for implementing the additional ads. The experimental protocol that you created states that the extra ads should amount to 60 minutes of additional broadcasting time. Can you test whether all the local branches have followed the experimental protocol?

d) Is the instrument exogenous? In particular, do you believe that the instrument

i. is unrelated to omitted factors? (5p)

ii. has no direct effect on flights sold? (5p)

Question 2

Consider the following model

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \tag{4}$$

where X_i and u_i are independent. A regression of Y on X would give a consistent estimate of β_1. Unfortunately, however, you do not exactly observe X_i but with some measurement error. What you observe is $X_{1i} = X_i + \varepsilon_{1i}$, where the measurement error ε_{1i} is purely random and, thus, independent of X_i and u_i.

a) Show that $\text{cov}(Y_i, X_{1i}) = \beta_1 \text{var}(X_i)$ in this model.

b) Show that the OLS estimator $\hat{\beta}_1^{OLS}$ of a regression of Y_i on X_{1i} is, in general, inconsistent.

Now suppose you have a second measure of X_i, again with some measurement error, $X_{2i} = X_i + \varepsilon_{2i}$. The measurement error ε_{2i} is again purely random and, thus, independent of X_i, u_i, and ε_{1i}.

c) Show that the TSLS estimator using the first measurement X_1 as an instrument for the second X_2 will give a consistent estimate of β_1.

Question 3

Consider the standard model with heterogeneous effects β_{1i}

$$Y_i = \beta_{0i} + \beta_{1i} X_i + u_i. \tag{5}$$

Also, we have an instrument Z_i that is related to X_i through the (first stage) equation

$$X_i = \pi_{0i} + \pi_{1i} Z_i + v_i. \tag{6}$$

The heterogeneous parameters $(\beta_{0i}, \beta_{1i}, \pi_{0i}, \pi_{1i})$ are assumed to be independent of each other, the instrument Z_i, and the error terms u_i and v_i, which have $cov\,(u_i, v_i) = \gamma$. Also, we assume that Z_i is a valid instrument in the sense that $E\,(u_i|Z_i) = E\,(v_i|Z_i) = 0$.

a) Show that $cov\,(u_i, X_i) = \gamma$.

b) We know that the OLS estimator converges to $\hat{\beta}^{OLS} = \frac{s_{XY}}{s_X^2} \xrightarrow{p} \frac{cov(Y_i, X_i)}{var(X_i)}$ when $n \to \infty$. Show that $\hat{\beta}^{OLS}$ is an inconsistent estimator of the average treatment effect in this model. In deriving your answer you may assume that X_i is independent of (β_{0i}, β_{1i}).

Now assume that the heterogeneous parameters are no longer independent of each other. In particular, we have a population that consists of three groups with different parameter values:

	Group		
	Never-Taker	Compliers	Others
Fraction	50%	30%	20%
β_{1i}	$\beta_{1i}^{NT} = 0$	$\beta_{1i}^{C} = 2$	$\beta_{1i}^{D} = 6$
π_{1i}	$\pi_{1i}^{NT} = 0$	$\pi_{1i}^{C} = 1$	$\pi_{1i}^{D} = d$

where d is an unknown parameter. Recall that under these conditions the instrumental variable converges to

$$\hat{\beta}^{TSLS} = \frac{s_{ZY}}{s_{ZX}} \xrightarrow{p} \frac{cov\left(Y_i, Z_i\right)}{cov\left(X_i, Z_i\right)} = \frac{E\left(\beta_{1i} \cdot \pi_{1i}\right)}{E\left(\pi_{1i}\right)}.$$

c) Show that it is possible for the instrumental variable estimator to converge to a value that is lower than the minimum effect found in the three groups ($\beta_{1i}^{NT} = 0$) when $n \to \infty$.