Homework 5

Economics 7103

February 19, 2024

1 Python

1.1

Run the ordinary-least-squares regression of price on mpg, the car indicator variable, and a constant. Report and interpret the coefficient on miles per gallon
The coefficient on mpg from the basic OLS regression was -21.2121. This is interpreted as each additional mile per gallon is associated with a 21 dollar decrease in the value (price) of the car.

1.2

What forms of endogeneity are you concerned about when estimating the coefficient on mpg?
The main form of endogeneity I am concerned about here is that miles per gallon is correlated with unobserved characteristics like engine size, age of the car, and type of car that are also correlated with price.

1.3

To correct for this endogeneity, you would like to use instrumental variables. Report the estimated second-stage coefficients, standard errors or confidence intervals, and the first-stage F-statistic for the excluded instrument in the same table for the following procedures:
a Perform two-stage-least-squares estimation by hand using weight as the excluded instrument.
b Perform two-stage-least-squares estimation by hand using $weight^2$ as the excluded instrument.
c Perform two-stage-least-squares estimation by hand using height as the excluded instrument.

Model	3a (weight)	3b ($weight^2$)	3c (height)
First Stage F-Statistic	256.802687	257.017435	203.655981
Intercept	17627.6397	17441.2286	-264024.2000
	(1754.8654)	(1751.1209)	(746919.2687)
Car	-4676.0923	-4732.6673	-90156.3873
	(574.3666)	(573.2907)	(226687.3470)
$\hat{mpg}$	150.433236	157.061859	10165.737682
	(62.1555)	(62.0219)	(26559.8254)

Table 1: Coefficients from Python

d In words, what are the different exclusion restrictions required for parts (a)-(c)? Does this seem reasonable for these instruments?
The exclusion restriction requires that the instrumental variable only affects the dependent variable through the independent variable and not in any other way. For example, for model a we require that weight only effects car price through its impact on mpg and not in any other way. Same goes for $weight^2$ and height in models b and c respectively. I think this is relatively reasonable since we are conditioning on car type.
e Compare and contrast the estimated coefficient on mpg from parts (a)-(c). What explains the discrepancies?

Estimates for models a and b are relatively similar and are all significant. The results for model c look very different. The estimates are much larger and no longer significant. This change is clearly coming from the use of height as the excluded instrument rather than a weight variable. Thinking intuitively, height is extremely correlated with the type of car (sedan or SUV) which may be why it is performing so poorly as an instrument for mpg when we already condition on car type. It is not capturing the variation that we want to exploit.

1.4

Calculate the IV estimate using GMM with weight as the excluded instrument.Report the estimated second-stage coefficient and standard error or confidence interval for mpg. What factors account for the differences in the standard errors?

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Intercept	1.763e+04	1667.3	10.573	0.0000	1.436e+04	2.09e+04
mpg_hat	150.43	59.207	2.5408	0.0111	34.389	266.48
car	-4676.1	548.12	-8.5312	0.0000	-5750.4	-3601.8

Table 2: IVGMM results from Python

The standard errors are different because the standard errors from the by hand 2SLS are to an extent incorrect because they are computed using instruments rather than the original variables. When we use a "canned" IV package it automatically calculates the standard errors the correct way for us.

2 Stata

2.1

Use the ivregress liml command to compute the limited information maximum likelihood estimate using weight as the excluded instrument. Report your second-stage results in a nicely-formatted table using outreg2. Use heteroskedasticity-robust standard errors.

	IV LIML
MPG	150.43
	(63.05)
Car	-4676.09
	(589.70)
Montiel-Pflueger F-statistics	78.36
LIML critical value for $\tau = 5\%$	37.42
Observations	1000

Table 3: Stata estimates

2.2

Use weakivtest to estimate the Montiel-Olea-Pflueger effective F-statistic. What is the 5 percent critical value, the F-statistic, and conclusion?
The 5 percent critical value and f-statistic are reported in the above table. The F-statistic is larger than the critical value which means we reject the null hypothesis that the instrument is not significant. This implies that weight is not a weak instrument.