Homework 6 Solutions

Kelly Lifchez

February 17, 2025

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 (do not construct a table).

The coefficient on miles per gallon is -22.21, indicating that an additional mile per gallon is associated with a 22.21 USD average decrease in sales price.

1.2 What forms of endogeneity are you concerned about when estimating the coefficient on mpg?

There is likely omitted variable bias, as a multitude of factors could be correlated with mpg and price. For example, subsidies for fuel-efficient vehicles could make the sales price lower for more fuel-efficient vehicles. Also, smaller cars often cost less than bigger cars and are, on average, more fuel-efficient.

1.3 To correct for this endogeneity, you would like to use instrumental variables. Specifically, you are interested in the system of equations:

$$price_v = \beta_0 + \beta_1 mpg_v + \beta_2 car_v + e_v \tag{1}$$
$$mpg_v = \gamma_0 + \gamma_1 z_v + \gamma_2 car_v + u_v \tag{2}$$

where z_v is the value of the instrument for vehicle v and e_v and u_v are error terms. 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 (just use a regular F-statistic for this exercise rather than the robust Montiel-Olea-Pflueger F-statistic).

a) Perform two-stage least-squares estimation by hand using weight as the excluded instrument.

See Table 2 column (a).

b) Perform two-stage least-squares estimation by hand using weight^2 $as the excluded instrument.$

See Table 2 column (b).

c) Perform two-stage-least-squares estimation by hand using height as the excluded instrument.

See Table 2 column (c).

d) In words, what are the different exclusion restrictions required for parts (a)-(c)? Does this seem reasonable for these instruments?

Per the exclusion restriction, the excluded instrument has to affect price only through miles per gallon. This means that in part (a), weight would have to affect price through miles per gallon, conditional on car type, and be uncorrelated with the error terms. In part (b), weight^2 would have to affect price through miles per gallon, conditional on car type, and be uncorrelated with the error terms. In part (c), height would have to affect price through miles per gallon, conditional on car type, and be uncorrelated with the error terms. Even conditioned on car type, the size of a vehicle is likely to influence the price in more ways than just through miles per gallon, so the exclusion assumption may not be met.

1.4 Compare and contrast the estimated coefficient on mpg from parts (a)-(c). What explains the discrepancies?

The coefficients on weight and weight^2 are similar (which is unsurprising given that they are two forms of the same variable), but height is significantly larger and not statistically significant. The F-statistic is also smaller on the height variable, but still very large for an instrument with no statistically significant associations with the other covariates. All three IVs have high first-stage F-statistics, but that makes sense given how highly correlated vehicle size and miles per gallon is likely to be.

Table 1: Limited Information Maximum Likelihood Estimates using Different IVs

	(a)	(b)	(c)
IV	Weight	Weight^2	Height
MPG	150.43	157.06	10165.74
	(28.46, 272.40)	(35.35, 278.77)	(-41953.84, 62285.31)
Car	-4676.09	-4732.67	-90156.39
	(-5803.20, -3548.99)	(-5857.66, -3607.67)	(-534995.45, 354682.67)
Constant	17627.64	17441.23	-264024.20
	(14183.99, 21071.29)	(14004.92, 20877.53)	(-1729738.42, 1201690.02)
Observations	1000	1000	1000
First-Stage F-Stat	256.80	257.02	203.66

2 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?

The second-stage coefficient on mpg is 150.43. The confidence interval is [26.856, 274.01] and a standard error of 63.05. In the first part, the standard errors did not account for the two-stage process, meaning that the standard errors in question 1 are less accurate than those calculated in the GMM estimation.

3 Stata

3.1 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.

Table 2: Limited Information Maximum Likelihood Estimates using Weight as IV

	(1)
Miles per Gallon	150.43*
	(63.05)
Vehicle class	-4,676.09**
	(589.70)
Constant	17,627.64**
	(1,772.78)
Observations	1,000
R-squared	0.10

Robust standard errors in parentheses
** p<0.01, * p<0.05

3.2 Use weakivtest to estimate the Montiel-Olea-Pflueger effective F-statistic. What is the 5% critical value, the F-statistic, and conclusion?

5% critical value for TSLS: 37.42
5 % Critical Value for LIML: 37.42
F-statistic: 78.36
Conclusion: We can reject the null hypothesis that the instrument is weak.