Homework 5 Answers

Economics 7103

1   Python

1. The OLS estimate is -22.21, which implies that within a vehicle class, an additional mile per gallon *decreases* the sales price by \$22.21 on average.

2. There may be several reasons that we should not trust the OLS estimate of the coefficient on *mpg*. One likely source of endogeneity is omitted variables bias: it is likely that unobserved attributes of the vehicle are correlated with miles per gallon. Lower fuel efficiency vehicles are often luxury vehicles, so luxury features may be negatively correlated with miles per gallon (and positively correlated with the price), which would negatively bias the coefficient on miles per gallon.

3. Two-stage-least-squares estimation:

    (a) See table 1, column (a).

    (b) See table 1, column (b).

    (c) See table 1, column (c).

    (d) The exclusion restriction states that conditional on vehicle class, the excluded instrument should not be correlated with the error term *e*. Or, the excluded instrument should only influence *price* through its impact on *mpg*. It is not likely that any of these instruments would satisfy the exclusion restriction in real life. More features in a car will likely make it heavier. Taller vehicles have more cargo space and may sell for more.

    (e) The estimated coefficients and confidence intervals in (a) and (b) are similar. It is surprising to note that the estimated coefficient changes when you change the functional form of the instrument–despite the assumptions for IV continuing to hold! In part (c), the estimate goes haywire. The reason for this is that *height* is a *weak instrument* for *mpg*. The rule of thumb for instrument strength is that the first-stage partial F statistic should be approximately 8 or higher. Using a weak instrument is often *worse* than not using an instrument at all.

4. Using GMM, the estimated coefficient on mpg is 150.43 with 95% confidence interval (26.856,274.01). The GMM estimation takes place in one step and so the standard errors are correct, whereas in (a), the standard errors were not adjusted to account for the two-stage estimation. Also note that for IV estimates that one should use heteroskedasticity-robust or clustered standard errors as the situation calls for.

2   Stata

1. See table 2.

2. The 5% critical value is 37.4 for the LIML estimator. The F statistic is 78, so we can reject the null hypothesis that the instrument is weak.

	(a)	(b)	(c)
Sedan	-4676.09	-4732.67	-90156.39
	(-5803.2, -3548.99)	(-5857.66, -3607.67)	(-534995.45, 354682.67)
MPG	150.43	157.06	10165.74
	(28.46, 272.4)	(35.35, 278.77)	(-41953.84, 62285.31)
Constant	17627.64	17441.23	-264024.2
	(14183.99, 21071.29)	(14004.92, 20877.53)	(-1729738.42, 1201690.02)
Observations	1000	1000	1000
First stage F	75.464083	75.769006	0.000386

Table 1: Estimation results from the two-stage-least-squares estimation. 95% confidence intervals are over-estimates of the precision due to the two-stage estimation and should be bootstrapped in practice.

VARIABLES	(1) price
MPG	150.4**
	(63.05)
Sedan	-4,676***
	(589.7)
Constant	17,628***
	(1,773)
Observations	1,000
R-squared	0.104

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2: LIML instrumental variable estimates for Stata question 1.