Homework 5

Economics 7103

February 20, 2024

1 Python

1. OLS regression results are presented in the output Table 1. As the car becomes more fuel-efficient (able to travel more miles per gallon), its price decreases by $22 for each additional mile per gallon. This implies that higher fuel efficiency is associated with a lower cost for the car.

Dep. Variable:	price	**R-squared:**	0.193
Model:	OLS	**Adj. R-squared:**	0.191
Method:	Least Squares	**F-statistic:**	119.1
Date:	Mon, 19 Feb 2024	**Prob (F-statistic):**	4.09e-47
Time:	21:48:16	**Log-Likelihood:**	-9574.6
No. Observations:	1000	**AIC:**	1.916e+04
Df Residuals:	997	**BIC:**	1.917e+04
Df Model:	2		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
mpg	-22.2121	16.521	-1.344	0.179	-54.632	10.208
car	-3202.5704	261.823	-12.232	0.000	-3716.357	-2688.783
const	2.248e+04	490.096	45.874	0.000	2.15e+04	2.34e+04

Omnibus:	1.020	**Durbin-Watson:**	1.989
Prob(Omnibus):	0.601	**Jarque-Bera (JB):**	0.888
Skew:	0.039	**Prob(JB):**	0.642
Kurtosis:	3.123	**Cond. No.**	151.

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Table 1: OLS Regression results

2. In this model, I am concerned about the **Omitted Variable Bias**. For instance, car characteristics other than being a Sedan or SUV, are omitted from the model that might be correlated with both the price and mpg. This can lead to biased and inconsistent estimates.

3. (a-c) Manully calculated 2Stage-least-square results are presented in the output Table 2. In the first column is presented regression results from part 3a, in the second column 3b and in the third one 3c.

	Dependent variable: price		
	(1)	(2)	(3)
Miles per gallon	150.43**	157.06**	10165.74
	(62.16)	(62.02)	(26559.83)
Car	-4676.09***	-4732.67***	-90156.39
	(574.37)	(573.29)	(226687.35)
F-statistics from the 1st Stage	256.8	257.02	203.66
F-stat p-value	0.0	0.0	0.0
Observations	1000	1000	1000
R^2	0.20	0.20	0.19
Adjusted R^2	0.19	0.19	0.19
Residual Std. Error	3481.08	3480.12	3491.04
F Statistic	121.62***	121.97***	118.09***
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 2: Manually Calculated IV Results

3.d. The instrumental variables should satisfy the following requirements:

1. The instrument must clearly affect the endogenous variable, in this case, 'mpg'. F-statistics from the first stage shows how strong the relationship between the instruments and the endogenous variable is. As we see from the Table 2 the weight square has the higher F-statistic, suggesting a stronger relationship between mpg and the car's weight squared. The F-statistic is lower in the first stage, where height is used as an instrument. This suggests that height is likely correlated with the car variable, as it serves as an indicator of whether a car is a sedan or SUV. This violates the second condition - exclusion condition for height.

2. **Exclusion restriction** - an instrumental variable should be uncorrelated with any other determinants of the dependent variables. In other words, an instrument must have a unique channel for causal effect. In this case, instruments (weight, weight square, and height) should not be correlated with the variable 'car'. The only channel impacting the prices should be through 'mpg'.

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

3.e The coefficient for mpg in the third column is both insignificant and very large in magnitude, which is most likely caused by a violation of the exclusion restriction for the instrumental variable (IV) height. Coefficients when weight is included as IV in the first stage is significant. Also, using weight's quadratic form as an IV results in significant estimators with larger magnitude for the mpg coefficient.

4. IV GMM results are presented in the Table 3. The coefficient for mpg is the same as in the previous model, but the standard error is slightly larger. This difference might be a sign of a weak instrument.

	IVGMM
Miles per gallon	150.43
	(63.05)
=1 if the vehicle is sedan	-4676.09
	(589.70)

Table 3: IV estimate calculated using GMM

2 Stata

1. The limited information maximum likelihood estimates are presented in the Table 4. The results are the same as those obtained from IV-GMM.

	(1)
VARIABLES	The limited information maximum likelihood estimate
Miles per gallon	150.4**
	(63.05)
=1 if the vehicle is a 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 4: IV estimate calculated using GMM

2. **'weakivtest'** tests the null hypothesis that instruments are weak or that the estimator's Nagar bias is large relative to a benchmark for both two-stage least-squares estimation and limited-information maximum likelihood with one endogenous regressor.The Montiel-Olea-Pflueger effective F-statistic estimated using 'weakivtest' is 78.362 at 5% confidence level. At 5% confidence level, Two-Stage-Least-Squares (TSLS) is 37.418 and Limited Information Maximum Likelihood (LIML) is 37.418. Since F statistics is higher than TSLS and LIML, weakivtest rejects the null of weak instruments for TSLS or for LIML for a weak instrument threshold of $\tau = 5\%$.