ECON 7103 Homework 5

Yifan Liu (yliu3494)

Spring 2023

1 Python

---

1. OLS

Table 1 shows the ordinary-least-squares regression of *price* on *mpg*, the *car* indicator variable, and a constant.
The coefficient on mpg is not statistically significant. If we have to interpret, -22.21 means that one unit in the increase of miles per gallon is expected to decrease the sales price of the vehicle by 22.21 units, holding all other variables constant.

Dep. Variable:	price	**R-squared:**	0.193
Model:	OLS	**Adj. R-squared:**	0.191
Method:	Least Squares	**F-statistic:**	119.1
Date:	Mon, 20 Feb 2023	**Prob (F-statistic):**	4.09e-47
Time:	13:49:26	**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]**
const	2.248e+04	490.096	45.874	0.000	2.15e+04	2.34e+04
car	-3202.5704	261.823	-12.232	0.000	-3716.357	-2688.783
mpg	-22.2121	16.521	-1.344	0.179	-54.632	10.208

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 estimates

2. Endogeneity

The endogeneity might result from simultaneity. While the fuel efficiency affects the price, the price also affects the fuel efficiency in miles per gallon. In other words, the fuel efficiency and the price are determined by each other. A change in error term causes the price to change, which causes the fuel efficiency to change. Therefore, the error term and the fuel efficiency are not independent.

3. IVs

(a)(b)(c) use *weight*, *weight*^{2}, and *height* as the excluded instrument respectively. In other words, the different exclusion restrictions are:
(a) the weight of the vehicle must explain some of the variation in fuel efficiency; and the weight of the vehicle must not be correlated with the error term in the regression of interest.
(b) the square weight of the vehicle must explain some of the variation in fuel efficiency; and the square weight of the vehicle must not be correlated with the error term in the regression of interest.
(c) the height of the vehicle must explain some of the variation in fuel efficiency; and the height of the vehicle must not be correlated with the error term in the regression of interest.
I think the first two instruments are more reasonable since the weight of the vehicle seems to be highly related with the fuel efficiency, and thus affects the price. The heavier the vehicle is, the more energy it needs to get moving. Heavier vehicles have greater inertia and greater rolling resistance, which both contribute to increased fuel consumption.
The height of the vehicle can be associated with the fuel efficiency since the higher vehicles are more likely to be larger. However, some sports cars that can be very low but use more fuels, are very expensive. It violates the monotonicity assumption.

The two-stage-least-squares estimations by hand using different instrumental variables are shown in Table 2 as follows:

	weight as IV	weight2 as IV	height as IV
constant	17627.64	17441.23	-264024.21
mpg	150.43	157.06	10165.74
car	-4676.09	-4732.67	-90156.39

Table 2: Two-stage-least-squares estimations

As Table 2 shows, the estimates in (a) and (b) are similar, but they have a huge discrepancy compared with the estimates in (c). The *mpg* estimate in (b) is slightly larger than that in (a). The *height* might not be a good instrument in this case as it can violate the monotonicity assumption.

4. IVGMM

Table 3 reports the estimated second-stage coefficient and standard error for *mpg*.
Compared with 2SLS, IVGMM relaxes the assumption of instrument exogeneity and accounts for potential correlation between the instrument and the error term. If the instruments are weak or endogenous, the 2SLS standard errors may be biased and the test statistics may be invalid. In IVGMM estimation, however, the standard errors are calculated using a more general method that accounts for both clustering/robustness and potential instrument correlation with the errors. The IVGMM method estimates the variance-covariance matrix of the model parameters by minimizing a moment condition-based distance function, which is robust to weak or endogenous instruments. This explain the differences in the standard errors between 2SLS and IVGMM.

Dep. Variable:	price	**R-squared:**	0.1045
Estimator:	IV-GMM	**Adj. R-squared:**	0.1027
No. Observations:	1000	**F-statistic:**	218.28
Date:	Wed, Feb 22 2023	**P-value (F-stat)**	0.0000
Time:	21:22:42	**Distribution:**	chi2(2)
Cov. Estimator:	robust		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
one	1.763e+04	1772.8	9.9435	0.0000	1.415e+04	2.11e+04
car	-4676.1	589.70	-7.9295	0.0000	-5831.9	-3520.3
mpg	150.43	63.051	2.3859	0.0170	26.856	274.01

Endogenous: mpg
Instruments: weight
GMM Covariance
Debiased: False
Robust (Heteroskedastic)

Table 3: IV estimates using GMM with *weight* as the excluded instrument

2 Stata

1. LIML

VARIABLES	(1) Limited information maximum likelihood estimates
mpg	150.43*
	(63.05)
car	-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

Table 4: Limited information maximum likelihood estimate using *weight* as the excluded instrument

2. Weak IV test

The effective F statistics is 78.362 at 5% confidence level. The 5% critical value is 37.418. In other words, F statistics is larger than the critical value. The null hypothesis for weak instruments is rejected for the large value of the effective F.
The result table of weakivtest can be found after running the attached do. file on Stata.