

Computer Practice 2: Demand Estimation

This Practice Set is partly based on Empirical Problem Set # 2 from Glenn Ellison and Stephen Ryan (MIT). The intent of this problem set is to get you familiar with IV techniques. After finishing this problem set, you should be familiar with performing simple instrumental variables regressions using Stata. Since Stata 10, you can use command `ivregress`. For Stata 9, commands `ivreg` and `ivreg2` can be used. `ivreg2` provides, among other things, automatic output of the Sargan statistic for over-identifying restrictions.

The data set is quantity, price, cost, and demographic variables on broiler chickens over 40 years in the United States, and is called “broiler.csv”. The data is taken from Dennis Epple and Bennett McCallum’s paper: “Simultaneous Equation Econometrics: The Missing Example.” The data is in comma delimited format, and can be imported into Stata format. Inspecting the data, you will see the following column headers:

year: year
q: quantity of chicken consumption
y: per capita real disposable income
pchick: price of chicken
pbef: price of beef
pcor: price of corn
pf: price of chicken feed
cpi: consumer price index
qproda: aggregate production of chicken in pounds
pop: population of the US
meatex: exports of beef, veal, and pork in pounds

We are interested in using this data to estimate the demand curve for broiler chickens. The first step is to upload the data file to a directory in your PC and then start Stata. Once Stata is up and running, you are ready to load your first data. Your first task once you have loaded the data is to generate some variable transformations. The model that we are interested is:

$$q = \alpha + \beta p + \gamma X + u \quad (1)$$

which is the demand curve relating quantity demanded to prices and other covariates. What should enter X ? The answer is everything in our data set that we think is going to shift demand for chickens, such as the price of substitutes (*pbef*), population (*pop*), and income (*y*). In practice, economic theory is helpful in suggesting what kinds of variables might go here. We want to estimate a constant elasticity of demand model, so we are going to take logs of all of our variables. To do this, we generate new variables using the `gen` command: “`gen logq = log(q)`”, and so on, for all the variables in this data set. Recall that this allows us to interpret the coefficients as elasticities. For clarity, all variables from this point forward are assumed to be in logs.

1. Import the data as comma-separated-values, csv. Prepare a do file to execute all commands. Transform all variables in logs. Label all transformed variables according to their descriptions mentioned above.
2. Regress using OLS the following equation:

$$\log(q) = \alpha + \beta \log(pchick) + v_1 \quad (2)$$

Interpret and report your results (coefficient, robust standard errors, and R^2 are sufficient). What exactly are you recovering by running OLS on such an equation? What is the interpretation of the coefficient on price, and what do you make of its sign?

3. Regress using OLS the following equation:

$$\log(q) = \alpha + \beta \log(pchick) + \gamma_1 \log(pop) + \gamma_2 \log(y) + \gamma_3 \log(pbeef) + \gamma_4 \log(CPI) + v_2 \quad (3)$$

Are the added regressors jointly significant? What has happened to the fit of the model overall? What about the precision of the individual estimates? How do you interpret the coefficient for income? And the coefficient for price?

4. Give arguments to defend that pf and $pcor$ are good instruments.
5. Estimate the IV model using pf and $pcor$ as instruments. The elasticity of demand that Epple and McCallum report is -0.40 , with a standard deviation of 0.086 . How do your results compare?
6. Test that price is endogenous. Do the same for model (2). Comment.
7. Report both the Sargan test of over-identifying restrictions and the Hausman test for the validity of the OLS estimates. Comment. In view of all results, what would you do next?