github.com/alfurka/teaching

ECON3360 Causal Inference for Microeconometrics

Tutorial 5: Stata application of SEMs

Instructor: Julie Moschion

Problem I: the deterrent effects of smoking on income

Use the data in smoke.dta for the following questions. A model to estimate the effects of smoking on annual income (perhaps through lost work days due to illness, or productivity effects) is provided by:

$$ln(income) = \beta_0 + \beta_1 cigs + \beta_2 educ + \beta_3 age + \beta_4 age^2 + u_1 \quad (1)$$

where $cigs$ is number of cigarettes smoked per day, on average.

(1) Describe the data.

(2) Run an OLS regression for (1). How do you interpret β_1?

To reflect the fact that cigarette consumption might be jointly determined with income, a demand for cigarettes equation is:

$$cigs = \gamma_0 + \gamma_1 ln(income) + \gamma_2 educ + \gamma_3 age + \beta_4 age^2 + \gamma_5 ln(cigpric) + \gamma_6 restaurn + u_2 \quad (2)$$

(-) (-)

where $cigpric$ is the price of a pack of cigarettes (in cents) and $restaurn$ is a binary variable equal to 1 if the person lives in a state with restaurant smoking restrictions.

(3) Assuming these are exogenous to the individual, what signs would you expect for γ_5 and γ_6?

(4) Under what assumption is the income equation (1) is identified?

there must be at least one exogenous variable in supply equation that does not appear on demand equation.

(5) Use equation (2), remove the endogenous variable from the RHS and express cigs as a function of exogenous variables only. Estimate this model for $cigs$ using OLS. Are $ln(cigpric)$ and $restaurn$ significant in this equation? Perform an $F-test$.

F-stat<10 implies weak IV

(6) Now estimate the income equation (1) by 2SLS. Compare β_1 from 2SLS to that from OLS. Check the F-stat in the first stage. How could you try to increase the F-stat?

(7) Do you think that cigarette prices and restaurant smoking restrictions are exogenous in the income equation?

cig prices and restrictions could be correalted with the state-level confounders.

- It could be more difficult to put restrictions on high-smoking states.

Problem II: demand and supply for fish

Use the data in fish.dta for the following questions.

Background Graddy (1995) collected data on a market for whiting (a kind of fish) at the Fulton Fish Market in New York City. For each of 111 days, she collected observations on the price and quantity of whiting sold, along with some other variables. (So here, i indexes the market on different days.) A key additional variable she collected was a measure of offshore weather. The idea is that variation in offshore weather affects the supply of fish, but not demand. (Graddy, K., (1995), Testing for Imperfect Competition in the Fulton Fish Market", RAND Journal of Economics.) We are primarily interested in estimating the demand for fish.
Assume that the demand equation can be written, in equilibrium for each time period, as

$$ln(totqty_t) = \alpha_1 ln(avgprc_t) + \beta_{10} + \beta_{11} mon_t + \beta_{11} tue_t + \beta_{11} wed_t + \beta_{11} thur_t + u_{t1}$$

so that the demand is allowed to differ across the days of the week.

(1) Treating the price variable as endogenous, what additional information do we need to consistently estimate the demand-equation parameters?

(2) Variables $wave2_t$ and $wave3_t$ are measures of ocean wave heights over the past several days. What two assumptions do we need to make in order to use $wave2_t$ and $wave3_t$ as IVs for $ln(avgprc_t)$ in estimating the demand equation?

(3) Derive a reasonable supply equation for $ln(avgprc_t)$. Use the supply, get rid of the endogenous variable on the RHS and then estimate this equation using OLS. Are $wave2_t$ and $wave3_t$ jointly significant? What is the p-value of the test?

(4) Now estimate demand equation using 2SLS. What is the 95% confidence interval for the price elasticity of demand? Is the estimated elasticity reasonable?

(5) Given that the supply equation evidently depends on the wave variables, what two assumption would we need to make in order to estimate the price elasticity of supply?

ln(avgprct) = b1ln(totqtyt) + c1 wave2 + c3 wave3 + ut2

(6) Let's assume that we are thinking of using the day-of-the-week dummies as excluded variables (they do not affect supply) to identify the supply equation. Do these dummies affect demand? Are the day-of-the-week dummy variables jointly significant? What do you conclude about being able to estimate the supply elasticity?