KAN-CEAPV2505U
Econometric Analysis of Firm Data

July 30, 2025

Copenhagen Business School
Department of Economics
Ralf A. Wilke
Mette Franck
2025/2026, Autumn semester

Problem Set 4

Part 1: Simultaneous Equation Models

1. Use SMOKE.DTA for this exercise.

   (a) A model to estimate the effects of smoking on annual income (perhaps through lost work days due to illness, or productivity effects) is

   $$log(income) = \beta_0 + \beta_1 cigs + \beta_2 educ + \beta_3 age + \beta_4 age^2 + u,$$

   where *cigs* is number of cigarettes smoked per day, on average. How do you interpret β_1?

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

$$cigs = \gamma_0 + \gamma_1 log(income) + \gamma_2 educ + \gamma_3 age + \gamma_4 age^2 + \gamma_5 log(cigprice) + \gamma_6 restaurn + u,$$

where *cigprice* is the price of a pack of cigarettes (in cents), and *restaurn* is a binary variable equal to unity if the person lives in a state with restaurant smoking restrictions. Assuming these are exogenous to the individual, what signs would you expect for γ_5 and γ_6?

(c) Under what assumptions is the income equation from part (a) identified?

(d) Estimate the reduced form for *cigs*. (Recall that this entails regressing *cigs* on all exogenous variables.) Are log(*cigprice*) and *restaurn* significant in the reduced form?

(e) Now, estimate the income equation by 2SLS. Discuss how the estimate of β_1 compares with the OLS estimate.

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

2. Use the data airfare.DTA, but only for the year 1997.

(a) A simple demand function for airline seats on routes is

$$log(passen) = \beta_{10} + \alpha_1 log(fare) + \beta_{11} log(dist) + \beta_{12}[log(dist)]^2 + u,$$

where *passen* is the average number of passengers per day, $fare$ is the average airfare, and *dist* is the route distance (in miles). If this is a demand function, what should be the sign of α_1?

(b) Estimate the equation from part (a) by OLS. What is the estimated price elasticity?

(c) Consider the variable *concen*, which is a measure of market concentration. (The share of business accounted for by the largest carrier.) Explain what we must assume to treat *concen* as exogenous in the demand equation.

(d) Now assume that *concen* is exogenous to the demand equation. Estimate the reduced form for $log(fare)$ and confirm that *concen* has a positive partial effect on $log(fare)$.

(e) Estimate the demand equation using IV. Now, what is the estimated price elasticity of demand? How does it compare with the OLS estimate?

(f) Using the IV estimates, describe how demand for seats depends on route distance.

Part 2: Pooled Cross Section Models

3. Use the data KIELMC.dta for this exercise. The data contain pooled cross sections of house prices in North Andover, Massachusetts in 1978 and 1981. A garbage incinerator was build from 1981 and we are interested in how the (future) incinerator affects house prices. Note: It was not known in 1978 that an incinerator will be build.

   (a) The variable *dist* is the distance from each home to the incinerator site, in feet. Consider the model:

   $$log(price) = \beta_0 + \delta_0 y81 + \beta_1 log(dist) + \delta_1 y81 * log(dist) + u.$$

   If building the incinerator reduces the value of homes closer to the site, what is the sign of δ_1? What does it mean if $\beta_1 > 0$?

   (b) Estimate the model from part (a) and report the results in the usual form. Interpret the coefficient on $y81 * log(dist)$ What do you conclude?

   (c) Add *age*, age^2, *rooms*, *baths*, log(*intst*), log(*land*), and log(*area*) to the equation. Now, what do you conclude about the effect of the incinerator on housing values?

These problems are taken from the Wooldridge (2020) textbook.