BA-BMECV2502U Econometrics

September 9, 2025

Copenhagen Business School
Department of Economics
Ralf A. Wilke
2025/2026, efterårsemester

Problem Set 5

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: Maximum Likelihood Estimation

3. If $f(\mathbf{y}|\mathbf{x};\boldsymbol{\theta})$ is a correctly specified model for the density of $\mathbf{y}_i$ given $\mathbf{x}_i$, does $\boldsymbol{\theta}_0$ solve $\max_{\theta\in\Theta} E[f(y_i|\mathbf{x}_i;\boldsymbol{\theta})]$?

4. Suppose that for a random sample, $y_i|\mathbf{x}_i \sim N(m(\mathbf{x}_i,\boldsymbol{\beta}_0),\sigma_0^2)$, where $m(\mathbf{x},\boldsymbol{\beta})$ is a function of the K-vector of explanatory variables $\mathbf{x}$ and the $P\times 1$ parameter vector $\boldsymbol{\beta}$. Recall that $E(y_i|\mathbf{x}_i) = m(\mathbf{x}_i,\boldsymbol{\beta}_0)$ and $\text{var}(y_i|\mathbf{x}_i) = \sigma_0^2$.

   (a) Write down the conditional log-likelihood function for observation i. Show that the CMLE of $\boldsymbol{\beta}_0$, $\hat{\boldsymbol{\beta}}$, solves the problem $\min_\beta \sum_{i=1}^N [y_i - m(\mathbf{x}_i,\boldsymbol{\beta})]^2$. In other words, the CMLE for $\boldsymbol{\beta}_0$ is the so called nonlinear least squares estimator.

   (b) Let $\boldsymbol{\theta} = (\boldsymbol{\beta}',\sigma^2)'$ denote the $(P+1)\times 1$ vector of parameters. Find the score of the log-likelihood for a generic i. Show directly that $E[\mathbf{s}_i(\boldsymbol{\theta}_0)|\mathbf{x}_i] = \mathbf{0}$. What features of the normal distribution do you need to be correctly specified in order to show that the conditional expectation of the score is zero?

   (c) Use the first-order condition to find $\hat{\sigma}^2$ in terms of $\hat{\boldsymbol{\beta}}$.

   (d) Find the Hessian of the log-likelihood function with respect to $\boldsymbol{\theta}$.

   (e) Show directly that $-E[\mathbf{H}_i(\boldsymbol{\theta}_0|\mathbf{x}_i)] = E[\mathbf{s}_i(\boldsymbol{\theta}_0)\mathbf{s}_i(\boldsymbol{\theta})_0'|\mathbf{x}_i]$.

   (f) Propose an estimated asymptotic variance of $\hat{\boldsymbol{\beta}}$, and explain how to obtain the asymptotic standard errors.

These problems are taken from the Wooldridge (2010) and Wooldridge (2020) textbooks.