IIA-3 Econometrics: Supervision 6

Emre Usenmez

Lent Term 2025

Topics Covered

Faculty Qs:

Supplementary Qs: Endogeneity, measurement errors; simultaneous equations;

Related Reading:

Dougherty, *Introduction to Econometrics*, 5th ed, OUP
>   Chapter 14: Introduction to Panel Data Models

Wooldridge J M (2021) *Introductory Econometrics: A Modern Approach*, 7th ed,
>   Chapter 13: Pooling Cross Sections across Time: Simple Panel Data Methods

Gujarati, D N and Porter, D (2009) *Basic Econometrics*, 7th International ed, McGraw-Hill
>   Chapter 16: Panel Data Regression Models

Gujarati, D (2022) *Essentials of Econometrics*, 5th ed, Sage
>   Chapter 12: Panel Data Regression Models

Stock, J H and Watson M W (2020) *Introduction to Econometrics.* 4th Global ed, Pearson
>   Chapter 10: Regression with Panel Data

FACULTY QUESTIONS

QUESTION 1

SUPPLEMENTARY QUESTIONS

These questions are intended to guide the students through the procedures for estimation using Panel Data sets. A large part of the problem with this topic is keeing in mind the basic structure of a panel data set, there is little more theory than already covered with omitted variable bias.

Question A deals with pooled cross-sections in order to make the comparison with panel data sets in questions B and C.

QUESTION A

(1) Explain the difference between independently pooled cross section data sets and panel data sets. Is either heteroskedasticity or serial correlation likely to be more of a problem for pooled cross section estimates? Explain why.

Answer: Let's first define the two types of data sets.

Indy-pooled X-section: When we sample randomly from a population at different points in time we obtain an *independently pooled cross section*. These data sets consist of independently sampled observations which, among other things, ensures error terms across different observations would not be correlated.

One reason for using independently pooled cross sections is to increase the sample size, which would result in more precise estimators as well as in test statistics with more power. As long as the relationship between at least some of the explanatory variables and the response variable remains constant over time.

Because the data is pooled from different time periods, the populations may have different distributions in different time periods. To accomodate for this, the intercept in the model is allowed to differ across time periods. This is done by incorporating dummy variables for all but one time periods, where earliest time period is usually chosen as the base year. We can also use a time period dummy variable to check for structural changes as we did in the Michaelmas term, by interacting that dummy with a key explanatory variable of interest.

Panel Data: A *panel data*, or *longitudinal data* are different from independently pooled cross section in that the *same* unit of observation in a cross-sectional sample are surveyed across time. As a result, we cannot assume that the observations are independently distributed across time.

Given these definitions, both types of data sets can have both heteroskedasticity and serial correlation.

---

(2)

(a) Using data in the "houseprices" worksheet, and noting the definitions of each variable, estimate the following equation for 1981:

$$log(rprice)_i = \beta_0 + \beta_1 \ nearinc_i + u_i \tag{1}$$

Given that the building of the incinerator was completed before 1981, interpret your results. Do your results imply that the building of the incinerator causes house prices to fall?

Answer:

(b) Estimate equation (1) using data for 1978 (at which time the incinerator had not even been planned) and calculate the difference-in-differences estimator.