

ANALYSIS OF PANEL DATA

An Introduction

datascience@berkeley

Unobserved Effect Models and Pooled OLS and First-Difference Models

Panel Data and Unobserved Effect

- In the previous example, we have a cross section of cities observed in two different years.
- In general, the observational units can be individuals, companies, schools, countries, and so on.
- However, the two naïve OLS regression models estimated above are likely suffer from omitted variable problems.
- One can argue to include more observable explanatory variables in the regression model, such as education level, age distribution, gender distribution, and so on.
- There are a few alternative ways to utilize information available in a panel dataset in order to both deal with unobserved variables and capture the dynamic that would not have been possible using cross-section data.
- There are two types of unobserved variable (or unobserved effect): (1) those that are time invariant, and (2) those that are time varying.
- Note that this and the next two lectures are conceptually more abstract, and the mathematical notations used are more involved. Students are reminded to read the assigned chapters.

Simple Formulation With Unobserved Effect

Let i denote the cross-sectional unit and t the time period. A very simple formulation that includes a single observed explanatory variable and an unobserved variable is

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it} + (a_i + \epsilon_{it})$$

where $t = 1, 2$

$d2_t$ is a dummy variable equal to 0 when $t = 1$ and 1 when $t = 2$.

ϵ_{it} is the idiosyncratic error, which, in the current context, is time-varying.

a_i captures all unobserved, time-invariant variables that affect y_{it} . In econometrics, it is often called **unobserved effect**, **fixed effect**, or even **unobserved heterogeneity**. As such, the model above can be called an **unobserved effect model** or a **fixed effect model**.

Berkeley

SCHOOL OF
INFORMATION