ANALYSIS OF PANEL DATA

An Introduction

datascience@berkeley

Distributed Lag Models

Distributed Lag of Clear-Up Rate on Crime Rate

- Eide (1994) uses a panel data from police districts in Norway to estimate a *distributed lag* model for crime rate.
- In this example, we use only a single explanatory variable, but in a distributed lag framework.
- The explanatory varible is "*clear-up percentage (clrpc)*", the percentage of crime that leads to conviction.
- The crime rate data are collected on year 1972 to 1978. In this example, we follow Eide and use two lags. The intuition is that the past clear-up rate may have a deterrent effect on current crime rate. From a policy perspective, one could think of it as "current clear-up rate may have a deterrent effect on future crime rate."

$$log(crime_{it}) = \beta_0 + \delta_0 d78_t + \beta_1 clrpc_{i,t-1} + \beta_2 clrpc_{i,t-2} + a_i + u_{it}$$

The Estimated Model

$$\widehat{\Delta \log(crime)} = .086 - .0040\ \Delta clrprc_{-1} - .0132\ \Delta clrprc_{-2}$$
$$(.064) \quad (.0047) \qquad\qquad (.0052)$$
$$n = 53,\ R^2 = .193,\ \bar{R}^2 = .161.$$

- The second lag is statistically significant and is negative, implying that a higher clear-up rate two years ago would deter crime rate in the current year. Specifically, a 10% increase in $clrpc$ two years ago would lead to an estimated 13.2% decline in crime rate in the current year (in the time periods in the dataset).
- It is very important to remember that this model is estimated using data from 1972 to 1978. The country might be very different now, and the same estimated effect may not be applicable to the current environement. As a data scientist, you should always keep in mind the purpose of the building and estimating a model.

Berkeley SCHOOL OF INFORMATION