ANALYSIS OF PANEL DATA

An Introduction

datascience@berkeley

Using OLS Regression Model on Panel Data

Structure of the Data

- We need to pay attention to the structure of the data.
- This (**crime2**) is a panel dataset in which each of the 46 cities was observed two times (1982 and 1987).
- We cannot simply treat these 46 "once repeated" observations as 92 independent observations.

OLS Regression Revisit

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}.$$

$$\underset{n \times (k+1)}{\mathbf{X}} \equiv \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & & & & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}.$$

$\mathbf{X}$ is $n \times (k+1)$ and $\boldsymbol{\beta}$ is $(k+1) \times 1$, $\mathbf{X}\boldsymbol{\beta}$ is $n \times 1$.

Assumptions:

1. Linearity (in parameters)
2. $\mathbf{X}$ has rank $(k+1)$
3. $\mathrm{E}(\mathbf{u}|\mathbf{X}) = \mathbf{0}$
4. $\mathrm{Var}(\mathbf{u}|\mathbf{X}) = \sigma^2\mathbf{I}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix.
5. $\mathbf{u} \sim \mathrm{Normal}(\mathbf{0}, \sigma^2\mathbf{I}_n)$

Potential Violations of Underlying Assumptions

- The independence of the independent and identically distribution (iid) is violated due to the repeated observations.
- In fact, the Durbin–Watson test confirms the violation of the independence assumption.

```
        Durbin-Watson test

data:  crmrte ~ unem
DW = 1.2074, p-value = 3.681e-05
alternative hypothesis: true autocorrelation is greater than 0
```

- When this assumption is violated, OLS standard errors and test statistics are not valid. Statistical inference becomes unreliable.

Berkeley SCHOOL OF INFORMATION