Data Processing – Big Data Analysis

Jung PARK, PhD
Research Fellow in Data Science

First uploaded 2019 Nov 12
Last modified 2019 Nov 12

Table of contents
: Big Data Analysis

The structure of economic data

- Cross-sectional data

  : multiple samples observed at a single timing

  Ex. several families are surveyed during different weeks within a year (ignoring any minor timing differences in collecting the data)

- Time series data

  : single (multiple) sample observed at a multiple timing (the main focus is the time variation of a single sample)

  Ex. stock prices and GDP changes

- Pooled cross sections

  : different multiple samples taken at multiple timing

  Ex. housing prices in 1993 and 1996

- Panel or longitudinal data

  : same multiple samples observed at multiple timing,

  Ex. the same set of counties in the US on tax rates for the years 1980, 1985, and 1990

What is Big Data

- Definition using 3V aspects
  - Volume: such big that one computer cannot store and handle
  - Variety: in different formats including texts, numbers, images, etc
  - Velocity: speed of data creation, in real-time or nearly real-time

- In academic research, replicability is critical. Thus, the aspect of Velocity becomes less important.

References

Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach*. Nelson Education.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, *90*(10), 60-68.