Short introduction to multisource statistics

Ton de Waal

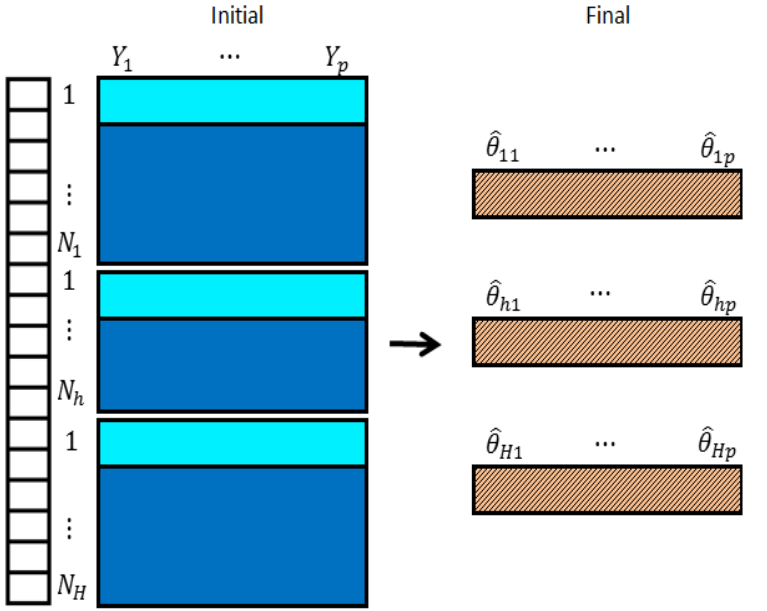
Summer school 2023

The bad news

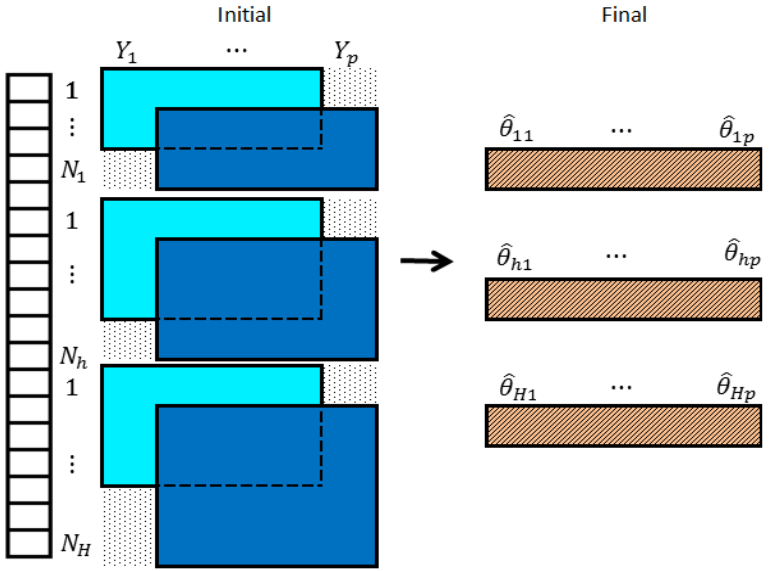
- Producing multisource statistics is much more complicated than single source statistics
  - There are many different situations
  - When we observe same target variable in different datasets, we are likely to see measurement errors
  - We may want to combine estimates from different datasets in order to improve those estimates
  - When we want to use nonprobability sample, we may have to use other information in order to correct for selectivity
  - We may have to correct for linkage errors
  - …

Basic Data Configurations

Combining non-overlapping microdata without coverage problems (BDC 1)

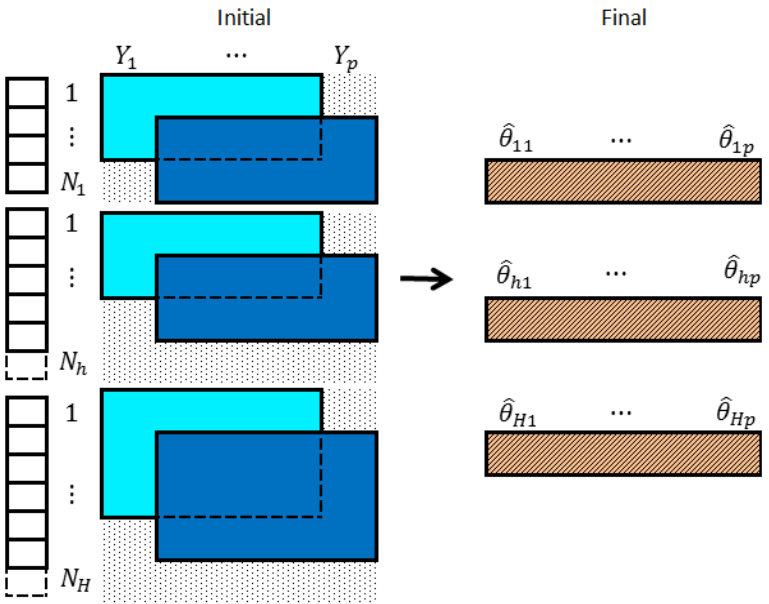


Combining overlapping microdata without coverage problems (BDC 2)

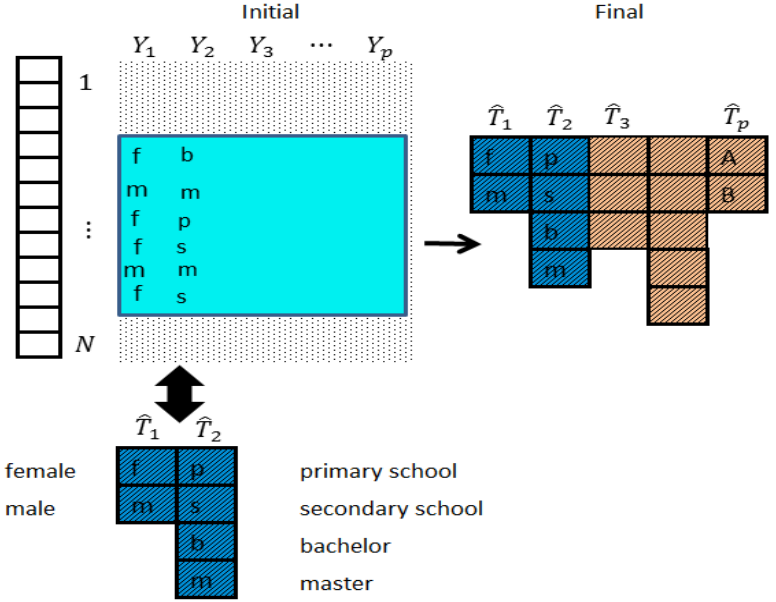


Basic Data Configurations

Combining overlapping microdata with undercoverage (BDC 3)

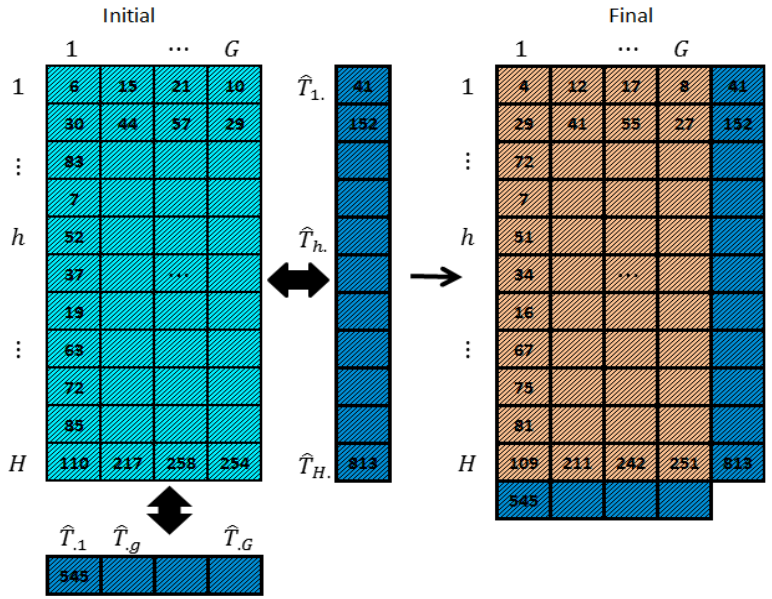


Combining microdata with a macro-data source (BDC 4)

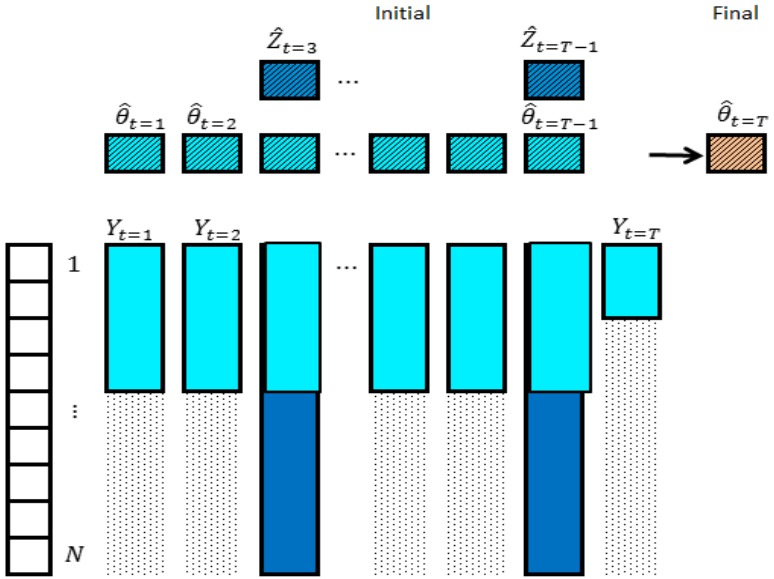


Basic Data Configurations

Combining macro-data sources (BDC 5)



Combining longitudinal data sources (BDC 6)



Contents

- Morning session
  - Correcting for measurement error (theory)
  - Correcting for selection error (theory)
  - Combining estimates from nonprobability sample and probability sample (theory)
- Afternoon session
  - Correcting for measurement error (application): Danila Filipponi & Roberta Varriale
  - Correcting for selection error (application): Natalie Shlomo
  - Combining estimates from nonprobability sample and probability sample (application): Jacco Daalmans & Sander Scholtus

Reference

- De Waal, T., A. van Delden & Sander Scholtus (2020), Multi-source Statistics: Basic Situations and Methods. *International Statistical Review* **88**(1), 203–228, doi:10.1111/insr.12352.