Incomplete Data Analysis for Economic Statistics

Incomplete data are ubiquitous in social sciences; as a consequence, available data are inefficient and often biased. This dissertation deals with the problem of missing data in official economic statistics. Chapters 2 to 6 are the body of the dissertation, all of which are based on the peer-reviewed articles written by the author. Below is the synopsis of each chapter.

Chapter 2 builds on the practices of the United Nations Economic Commission for Europe (UNECE). This chapter assesses deterministic single imputation, stochastic single imputation, and multiple imputation. Chapter 2 reveals that, in the current practices of imputation among the national statistical institutes, ratio imputation is often used for economic data and hot deck imputation for household data, both of which are deterministic single imputation. Furthermore, Chapter 2 demonstrates that multiple imputation is suited for public-use microdata.

As is made clear in Chapter 2, ratio imputation is often used to treat missing values in official economic statistics. However, there are three competing estimators in the literature: Ordinary least squares; ratio of means; and mean of ratios. Chapter 3 unifies ratio imputation models under the framework of weighted least squares and proposes a novel estimation strategy for selecting a ratio imputation model based on the magnitude of heteroskedasticity. The results in the Monte Carlo simulation give a strong support for the proposed method.

Shifting gears from single imputation to multiple imputation, Chapter 4 compares, from a new perspective, the three computational algorithms for multiple imputation: Data Augmentation (DA), Fully Conditional Specification (FCS), and Expectation-Maximization with Bootstrapping (EMB). In the literature, little is known about the relative superiority between the MCMC algorithms (DA, FCS) and the non-MCMC algorithm (EMB), where MCMC stands for Markov chain Monte Carlo. Based on simulation experiments, Chapter 4 contends that EMB is a confidence proper multiple imputation algorithm without between-imputation iterations, meaning that EMB is more user-friendly than DA and FCS.

Based on these findings, Chapter 5 proposes a novel application of the EMB algorithm to ratio imputation in order to create multiple ratio imputation, the new multiple imputation version of ratio imputation. Chapter 5 presents the mechanism of multiple ratio imputation and assesses the performance compared to traditional imputation methods using Monte Carlo simulation.

Furthermore, Chapter 6 outlines a concrete code in the R statistical environment to execute multiple ratio imputation by the EMB algorithm and provides brand-new software *MrImputation* implemented in R.

Chapter 7 summarizes the central findings and indicates the possible future courses of research. Combining all of these findings, this dissertation will be an important addition to the literature of missing data in particular and official statistics in general.