Theme: Exploiting new technologies and new data sources

Title: Diagnosing the Imputation of Missing Values in Official Economic Statistics via Multiple Imputation: Unveiling the Invisible Missing Values

Name of author: Masayoshi Takahashi
Full affiliation: National Statistics Center, Research Division
Complete address of corresponding author: 19-1 Wakamatsu-cho, Shinjuku-ku, Tokyo, 162-8668, Japan
Tel: +81-3-5273-1368
E-mail: mtakahashi@nstac.go.jp

Abstract:

Survey data almost inevitably have missing values. If missingness is at random (MAR), imputation can ameliorate the bias in the dataset. However, imputed values are merely estimates; thus, diagnosis is necessary. The paradox in imputation is that the true values are always missing, which makes it impossible to compare the imputed values with the truth. One solution to this conundrum is to indirectly assess the validity of missingness assumptions, solely based on observed data.

This paper proposes to use multiple imputation as a diagnostic tool. Based on the size of variability of imputed values by multiple imputation, we examine the stability and confidence of imputation models. Also, this paper evaluates the validity of missingness assumptions by comparing the kernel densities between observed and imputed values, and graphically showing the structure of missingness in official economic statistics. Finally, the fit of various imputation models is diagnosed by the overimputation technique, where observed values are temporarily deleted and multiply imputed to form confidence intervals of imputation models.

Based on these results, we propose a guideline to diagnose the accuracy of imputation in official economic statistics. As a concrete example, this paper utilizes the 2012 Japanese Economic Census for Business Activity dataset.

Keywords: missing data, diagnostics, multiple imputation, official economic survey, expectation-maximization, bootstrap, R