

さまざまな頑健比率代入法の比較

東京外国語大学 高橋 将宜

調査票によりデータを収集する公的経済統計では、すべてのデータを収集できることは稀であり、欠測が発生することが多い。公的経済統計の集計対象は、対数正規分布に従うと想定される変数の合計値（または平均値）である。そこで、諸外国では、 $Y_i = \beta X_i + \varepsilon_i$ で表される比率代入法（ratio imputation）によって、経済統計の欠測値を処理していることが明らかとなっている（高橋, 2017）。ここで、 $\varepsilon_i \sim N(0, X_i^{2\theta} \sigma^2)$ である。

比率代入法におけるパラメータ β の推定方法には3種類あるが、不均一分散の度合いを表す θ に応じてモデルを選択する方法が提案されている（Takahashi et al., 2017）。また、ブートストラップに期待値最大化法を適用した EMB アルゴリズムによって適切な多重代入法（multiple imputation）を構築できることも分かっており（Takahashi, 2017c）、この EMB アルゴリズムを活用して、比率代入法を多重代入化する方法も提案されている（Takahashi, 2017a; Takahashi, 2017b）。

一方、公的経済統計では、中小企業から大企業までさまざまな規模の企業を調査の対象としている。よって、欠測値を処理する際、外れ値の影響を考慮する必要がある。そこで、国連欧州経済委員会（UNECE: United Nations Economic Commission for Europe）の統計的データエディティングに関するワークセッション（Work Session on Statistical Data Editing）に参加していた国家統計機関における処理方法について調査した。約 43%の国において頑健（robust）な比率代入法を用いて外れ値に対処していたが、具体的な手法は中央値、トリム平均値、ウィンザー化平均値（Winsorized mean）、繰り返し加重最小二乗法（iteratively reweighted least squares）など、統一的な見解は得られていない。また、これら伝統的な頑健手法は、 Y_i における外れ値に対してのみ頑健であり、 X_i における外れ値に対して頑健ではない。

本研究では、クックの距離とトリム平均値の考え方を応用した新たな頑健比率代入法を提案する。また、外れ値の個数を自動的に決定する手法も考案する。この新たな手法は、 X_i における外れ値に対しても、 Y_i における外れ値に対しても頑健である。モンテカルロ・シミュレーションにおいて、上述の伝統的な手法と同等かそれ以上の性能を持つことが示されている。結果の詳細は当日報告する。

参考文献

- [1] 高橋将宜. (2017). 「諸外国の公的統計における欠測値の対処法：集計値ベースと公開型マイクロデータの代入法」、『統計学』第 112 号, pp.65-83.
- [2] Takahashi, M., Iwasaki, M., and Tsubaki, H. (2017). Imputing the mean of a heteroskedastic log-normal missing variable: A unified approach to ratio imputation. *Statistical Journal of the IAOS*, 33(3), in press.
- [3] Takahashi, M. (2017a). Multiple ratio imputation by the EMB algorithm: Theory and simulation. *Journal of Modern Applied Statistical Methods*, 16(1), pp.630-656.
- [4] Takahashi, M. (2017b). Implementing multiple ratio imputation by the EMB algorithm (R). *Journal of Modern Applied Statistical Methods*, 16(1), pp.657-673.
- [5] Takahashi, M. (2017c). Statistical inference in missing data by MCMC and non-MCMC multiple imputation algorithms: Assessing the effects of between-imputation iterations. *Data Science Journal*, in press.