

Estimation of gross flows under nonresponse and complex sampling

Guilherme Anthony Pinheiro Jacob(*)
National School of Statistical Sciences (E-NCE/IBGE)
Rio de Janeiro, Brazil
guilhermejacob91@gmail.com

Pedro Luis do Nascimento Silva
National School of Statistical Sciences (ENCE/IBGE)
Rio de Janeiro, Brazil
pedronsilva@gmail.com

Gross flows provide a much richer description of changes in the distributions of categorical variables over time when compared to estimates of changes in the marginals. However, non-response can introduce bias in gross flows estimates. Stasny (1987) developed a two-stage model connecting the observed counts (after non-response) with the underlying, non-observable Markov process in the population, modeling the non-response and the changes in distribution at the same time. Gutiérrez (2014) extended this approach to complex survey designs using the Maximum Pseudo-Likelihood method (Binder 1983; Molina and Skinner 1992). This approach produced interesting results, indicating that ignoring non-response introduces bias in the estimates. In order to simplify the process of estimating gross flows with non-response under complex survey designs, Jacob (2021) developed an R (R Core Team 2020) package called `surf` that produces: (1) estimates of the model parameters; (2) their standard-errors; and (3) a Rao-Scott adjusted Pearson χ^2 test (Rao and Scott 1981). The package is also easily integrated with the existing survey package (Lumley 2004). This article presents an analysis of the gross labour flows estimated using the `surf` package using real data from Brazil's Continuous National Household Sample Survey (PNADC). Limitations and possible extensions are also discussed.