Impacts of item nonresponse and imputation on variance estimation of labor income in the Continuous National Household Survey

Elizabeth Belo Hypolito[1] and Denise Britz do Nascimento Silva[1]

Studies on nonsampling errors, although extremely important for the evaluation and improvement of survey quality, tend to be less frequent than those about sampling errors. The Continuous National Household Survey (CNHS) variance estimates are calculated using a naive estimator that treats item nonresponse as true values and no information about the item nonresponse is published. The present study provides variance estimates of the average labor income produced by CNHS, separating the sampling and nonsampling components. The results of the survey current imputation procedure, the nearest neighbor method, is compared with those that would be obtained if the random hot deck imputation was used, identifying the most efficient method in relation to the precision of the estimates. The data are from the third quarter of 2013, one of the periods with the highest imputation rates in the entire CNHS series, and the first of 2018, a period after the stabilization of the survey process started in 2012. The method chosen for estimating the variance was the one proposed by Beaumont, Bissonnette and Bocci (2010), with some adjustments implemented in R software to suit the particularities of the Brazilian survey. The results showed that, in 2013, for the nearest neighbor method, the nonresponse variance estimate was considerably high for some indicators. When the imputation was performed by the random hot deck method, the nonresponse component was lower. For the period of 2018, when the imputation rates were very low, the percentages of the variance due to nonresponse were also low. Even though, these results indicate that variance estimates obtained when considering the imputed data as true and, therefore, the sample design as the only source of variability, underestimates the total variance. In addition, they reinforce the need for studies on nonsampling errors in Brazilian surveys for improving the process.

Keywords: item nonresponse, imputation, total variance, nonresponse variance.

[1] National School of Statistical Sciences (ENCE/IBGE), Brazil.