

Dealing with Non-ignorable Sampling and Non-response in Statistical Matching

Daniela Marella¹, Danny Pfeffermann²

Abstract

Statistical matching deals with the problem of how to combine information collected in different samples taken from the same target population, but on partly different survey variables. In this article, we consider the use of statistical matching when the samples are drawn by informative sampling designs and are subject to not missing at random nonresponse. The problem with ignoring the sampling process and nonresponse can result in severely biased estimators and a misrepresentative fused dataset. The proposed methodology employs the empirical likelihood approach.

Keywords: Empirical likelihood, Informative sampling, IPF algorithm, Matching uncertainty, NMAR nonresponse.

¹ Sapienza University of Rome, email: daniela.marella@uniroma1.it.

² University of Southampton, Hebrew University of Jerusalem and former National Statistician and Director General of Israel's Central Bureau of Statistics, email: d.pfeffermann@soton.ac.uk.