

# On the evaluation of matching noise produced by nonparametric imputation techniques

Pier Luigi Conti

*Dipartimento di Statistica, Probabilità e Statistiche Applicate  
Università di Roma "La Sapienza"*

Mauro Scanu

*Dipartimento per la Produzione Statistica e il Coordinamento Tecnico Scientifico  
Istituto Nazionale di Statistica*

## Abstract

Statistical matching attempts at producing a unique, synthetic data file, where variables observed in different sample surveys are jointly recorded. Such file is appropriate for any statistical analysis when the joint probability distribution of the variables of interest in the population coincides with the probability distribution of the same variables in the synthetic data file, or at least when these two distributions are “very close”. The “distance” between these distributions is called matching noise. In this paper, statistical matching methods based on hot-deck imputation procedures are investigated as a possible cause of matching noise. Two examples when data are generated from uniform and normal distributions are discussed.

*Keywords.* Data fusion, hot-deck imputation procedures, conditional independence assumption