



PHILIPPINE STATISTICAL ASSOCIATION, INC.

2/F Philippine Social Science Center
Commonwealth Avenue, Diliman, Quezon City
Telephone (632) 920-6513 Telefax (632) 456-1928
Email: psa.sec@gmail.com; psai.sec@gmail.com
PSA website : www.philstat.org.ph

THE PHILIPPINE STATISTICIAN 2006, Volume 55, Nos. 1-2

Empirical Evaluation of Statistical Matching

Zita VJ Albacea, and Glenn Melvin P. Gironella, Salve Kristina T. Mogul and Jezaniah Kira S. Tena

Abstract

This paper aims to empirically evaluate statistical matching as a method to create panel data sets. Two methods of statistical matching, namely: constrained and unconstrained statistical matching were studied. The properties of the distribution of distances measured using a distance function were examined. In particular, the 1997 Family Income and Expenditure Survey (FIES) and 1998 Annual Poverty Indicators Survey (APIS) data of Southern Tagalog Region in the Philippines were statistically matched after removing the exact matched records. To identify the “best” form of the distance function to use in matching, four statistically matched files (STM_A, STM_B, STM_C_UN, STM_C_WE) using different forms of a distance function were obtained. Descriptive statistics, histograms, and box and whisker plots of distances measured show the closeness of the values of the matching variables in the performed statistical matching. The cumulative percentage distributions show that at some point almost 95% of the data are already matched having small calculated-distances. The results indicate that the form of the distance function used in generating STM_A is the “best” since the distribution of the distances measured in STM_A clusters near zero with 94.95% of the records that were matched having calculated distances at most equal to one. Also, this form of the distance function resulted to a unit less measure of distances. With the distance function used in STM_A, unconstrained and constrained statistical matching procedures were applied to generate the unconstrained (UNC_STM) and constrained statistical matched data sets (CON_STM), respectively. The statistics of test variables in the exact matched data set (EXM) are found to be similar with those of the statistically matched data sets. However, the average computed distances, which indicate the degree of similarity of the matched records of the UNC_STM data set is found significantly lower than that of the CON_STM data set at 1% level of significance.

Keywords: *Unconstrained and constrained statistical matchings, distance function*