

Towards Better Edit And Imputation - The Euredit Project

John Charlton

Office for National Statistics, 1 Drummond Gate, London SW1V 2QQ, UK; e-mail:
john.charlton@ons.gov.uk

Keywords: Statistical data editing; imputation; evaluation of methods; neural networks; SVM; outlier robust methods; traditional methods;

Abstract

Advances in statistical and computer science have created opportunities for the application of newer technologies such as neural networks and outlier-robust methods for edit and imputation. The EUREDIT project, now completed, set out to facilitate informed choices about methods to use, through evaluating alternatives systematically using real data. This paper describes the work of the project, which was funded under EU Fifth Framework research programme - a large multi-national collaboration (see <http://www.cs.york.ac.uk/euredit>), involving twelve partners from seven countries. The project had the following objectives:

- To establish a standard collection of data sets for evaluation purposes
- To develop a methodological evaluation framework and develop criteria
- To establish a baseline by evaluating currently used methods.
- To develop and evaluate a selected range of new techniques.
- To evaluate different methods and establish best methods for different situations.
- To disseminate the best methods via a software CD and publications.

The statistical framework (Chambers, 2001) and operational characteristics developed by the project formed the basis of evaluations of existing and new methods developed during the project. These could also be used outside the Euredit project to evaluate other methods, including new techniques developed in the future. An overview of the findings will be presented.