

# New ways of specifying data edits

George Petrakos,

*Department of Public Administration, Panteion University and Liaison Systems SA, Athens, Greece*

Claudio Conversano

*Department of Economics, University of Cassino*

Gregory Farmakis,

*Liaison Systems SA, Athens, Greece*

Francesco Mola

*Department of Economics, University of Cagliari*

Roberta Siciliano

*Department of Mathematics and Statistics, University of Naples Federico II*

Photis Stavropoulos,

*Liaison Systems SA, Athens, Greece*

## Summary

Data editing is the process by which data collected in some way (a statistical survey for example) are examined for errors and are corrected with the help of software. Edits, the logical conditions that should be satisfied by the data, are specified by subject matter experts with a procedure which could be tedious and could lead to mistakes with practical implications. In order to render the process of edit specification more efficient we provide a new step - the definition of the so called abstract data model of a survey – which describes the structure of the phenomenon studied in a survey. The existence of this model enables experts to identify all combinations of variables which should be checked by edits and to avoid the definition of conflicting edits. Furthermore, we introduce an automatic data validation strategy – TREEVAL – that consists of a fast tree growing to derive automatically the functional form of edits and of a statistical criterion to clean up the incoming data. The TREEVAL strategy is cast within a Total Quality Management framework. The application of the proposed methodologies is demonstrated with the help of a real life application.

*Keywords:* data quality, data editing, object-oriented analysis, CART, fast partitioning algorithm, TREEVAL, Total Quality Management.

-