

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Prague, Czech Republic, 14-17 October 1997)

Item 3 of the provisional agenda

A SELECTIVE EDITING METHOD BASED ON ARIMA MODELS

(Abstract)

Submitted by the National Institute of Statistics , Spain¹

¹ Prepared by Pedro Revilla Novella

1. Time Series Analysis and ARIMA models are not very much used in official statistical agencies, with the exception of the seasonal adjustment. Nevertheless, short term indicators are in fact observed time series, and for this reason Time Series Analysis could be the theoretic framework of short term indicators methodology (together with sampling, statistical indices, etc). An example where ARIMA models can be used as a useful tool in statistical agencies is data editing.

2. The procedure presented in this paper makes use of the Total Quality Management principles and of Selective editing philosophy. ARIMA with Intervention Analysis models are the basic tool of the procedure. Following the Selective editing philosophy, the Procedure tries to solve two questions: 1) to characterize and to detect outliers in the macrodata. 2) to define the influential microdata. In short term surveys and indicators, the most useful information for data editing is the past data of the same population. For example, to edit a monthly short term indicator in time t (x_t), monthly and annual rates are often used for data editing. This method can be improved using the one step ahead ARIMA forecast for x_t (\hat{x}_t), and relating the observed data x_t with that ARIMA forecast, \hat{x}_t . If the observed value is far from the forecast, maybe the data have some kind of error. A confidence interval for x_t can be constructed when ARIMA forecast is used. For example, the 95% confidence interval is $P[\hat{x}_t - 1,96 \sigma < x_t < \hat{x}_t + 1,96 \sigma] = 0,95$ (σ is the forecast error standard deviation).

3. The method based on ARIMA forecast has at least two advantages as compared to monthly and annual rates. First, the monthly and annual rates use just one value of past data (x_{t-1} , or x_{t-12}). Instead, the ARIMA forecast use, in a optimum way, the whole past of the series ($x_{t-1}, x_{t-2} \dots$). Second, the ARIMA forecast allows the construction of statistical confidence intervals. This is very useful because the different subpopulations to edit (sectors, products, etc.) show a very different behaviour and variability.

4. The Macroediting procedure is being used in the Spanish National Statistical Institute to elaborate the Industrial Production and Price Indices. Savings in cost and in time of dissemination are being achieving.