BANCA D'ITALIA

Supplements to the Statistical Bulletin

Methodological notes and statistical information

Selective Editing to Improve the Quality of Banking Statistics

New series

Volume XIV Number 29 - 24 May 2004

SUPPLEMENTS TO THE STATISTICAL BULLETIN

Monetary Financial Institutions: Banks and Money Market Funds (monthly**)**

The Financial Market (monthly)

The Public Finances (monthly**)**

Balance of Payments (monthly**)**

Monetary and Credit Aggregates of the Euro Area: the Italian Components (monthly**)**

Financial Accounts (quarterly**)**

Payment System (half yearly**)**

Public Finance Statistics in the European Union (annual**)**

Methodological Notes and Statistical Information (irregular**)**

All the supplements are available on the Bank of Italy's site (www.bancaditalia.it).

Requests for classifications concerning data contained in this publication can be sent by e-mail to statistiche@bancaditalia.it

SELECTIVE EDITING TO IMPROVE THE QUALITY OF BANKING STATISTICS

P. Battipaglia, G. Bruno, F. Farabullini, C. Mahlknecht[1]

Summary

The balance sheet statistics of Monetary Financial Institutions' (MFIs) are produced on a monthly basis and are used by the European Central Bank (ECB) to compile the Eurosystem's monetary and credit aggregates.

In this paper we consider Italian banking statistics that cover approximately 90% of the total assets of the Italian MFI sector. Commercial banks' reports are received and checked for gross errors by a first-level filtering that provides on-line quality control. The editing workload is related to the number of banks (about 830) times the number of statistical aggregates that have to be sent to the ECB monthly. The time available for processing and editing is extremely short.

Here we propose a simple and efficient procedure to reduce the editing costs of both the central bank and the commercial banks.

1. INTRODUCTION

Quality control is one of the most costly activities in the production of statistical data for economic analysis. The cost is obviously greater when dealing with data sets with many high-frequency time series.

In these circumstances it is quite difficult to establish a satisfactory trade-off between timeliness and end-users' quality requirements.

Here the subject is addressed by considering the monthly production of the banking statistics that the Bank of Italy collects and provides to the European Central Bank (ECB). The quality/timeliness trade-off is mission critical because only one working day is available for controlling and assembling aggregate data. This situation calls for a data quality control procedure that satisfies efficiency, reliability and flexibility requirements. Moreover the procedure should feature high-level statistical and automation performances.

---

[1] *Bank of Italy, Research Department. The views expressed here do not involve the responsibility of the Bank of Italy.*

To this end we explore and test techniques combining time series Autoregressive Integrated Moving Average (ARIMA) modeling with a selective editing strategy. The latter allows the outliers flagged by the ARIMA models to be ranked in order to increase the efficiency of the data quality control activity. The ranking is built by estimating the impact of the editing of the outliers on the aggregate statistics under examination.

The application is carried out on the 17 series composing the “M3 counterparts” aggregate. These series refer to the loans granted by Italian banks to the money-holding sector and to banks’ own securities portfolios.

The paper is divided into five sections. In the next section we review the main statistical methodologies for data quality control. In the third section the banking statistics application is presented. The fourth section describes the proposed strategy while the last section presents the main results and outlines possible further research.

2. IDENTIFICATION TECHNIQUES AND TREATMENT OF OUTLIERS

The identification of outliers is a crucial phase of data quality control and ends with the confirmation of the datum or its replacement with an estimated or revised datum.

The literature contains many different proposals. Basically, there are two alternative approaches: the parametric approach based on a model generating the observations and on its parameters and the non-parametric approach based on the empirical analysis of the sample data without any model assumption.

Adopting the hypothesized probability distribution, under the parametric approach automatic routines can be written to pin down the extreme values depending upon preset significance thresholds and, eventually, to replace them with estimated values (see, for example, Alwan and Roberts, 1988).

Non-parametric techniques range from graphical representation to distance analysis and Artificial Intelligence systems. These methodologies are more flexible and better able to deal with nonlinear relationships (see Landenna and Marasini, 1990; Cheng and Titterington, 1994). However, they do not allow a systematic representation that can be used to interpret and forecast the behaviour and the relationships among the variables.

In data validation procedures, the identification of outliers is followed by a checking and possibly a correction step. These two phases often imply recontacting the reporting agents that provided the data. The whole set of these activities is usually referred to as the editing phase. The statistical agencies have to strike a balance between given data quality standards and timeliness in releasing the information. When dealing with a huge amount of high-frequency data, it is very costly to reconcile the two competing goals.

The need to increase efficiency in the allocation of resources for data quality control has fostered the development of selective editing strategies in the institutions responsible for the production of official statistics. The basic idea of these strategies is to define the main goals of each survey in terms of the accuracy of some output statistics. The adoption of a score function makes it possible to estimate the gain in the precision of the output statistics achievable through editing. The ranking generated by the score function provides a means for choosing to edit the data that have the largest impact on the final statistics. This methodology makes it possible to focus human and computing resources on the most important errors.

The efficiency gains obtainable by using these techniques are greater when the input data are generated by skewed distributions or when the weights used to combine the reported data are highly asymmetric.

The functional form of the score function depends on the final statistics of interest. In the case of the growth rate of aggregate **A**, given by the sum of **N** components, we have the following expression:

$$\Delta A_t = \frac{\sum_{i=1}^{N} c_{i,t}}{\sum_{i=1}^{N} c_{i,t-1}} - 1 \qquad (1)$$

The score can be computed using the linear term of the Taylor series expansion. For the i-th component the result is:

$$S_{i,t} = \frac{\partial(\Delta A_t)}{\partial c_i} = \left(\frac{1}{\sum_{j=1}^{N} c_{j,t-1}} \right) \left(\Delta c_{i,t} - \Delta c_{i,t-1} \cdot \frac{\sum_{j=1}^{N} c_{j,t}}{\sum_{j=1}^{N} c_{j,t-1}} \right) \qquad (2)$$

where: $\Delta c_{i,t}, \Delta c_{i,t-1}$

are the differences between the observed and the expected values (i.e. the forecast of a model) and identify the outliers for the i-th component in the time periods t and t-1.

If, at the time of the control of the data in time period t, the previous period values can be deemed unaffected by errors, the functional form of the score function is simplified because the aggregation function collapses to a sum and the score component becomes:

$$S'_{i,t} = \frac{\Delta c_{i,t}}{A_{t-1}} \qquad (3)$$

In the next section we show the operational context that suggested experimenting with selective editing for the quality control of money statistics. The selective editing technique is applied following a top-down approach. We start from the highest aggregation level and go down to the components. Here the process is halted at the level of component series making up a significant aggregate, but it could be pushed further to the level of each reporting agent.

3. BANKING STATISTICS QUALITY CONTROL

In this paper we used the data commercial banks send to the Bank of Italy on a monthly basis for the compilation of monetary and credit aggregates. The individual balance sheet data of about 800 banks are used to create about 350 series to be sent to the ECB, which prepares the euro-area statistics.

The data transmitted by the banks are first checked for formal correctness and internal consistency. When these two steps have produced satisfactory results, the information processing phase starts at an aggregate level. This last step, which often involves recontacting banks, has to be completed in just one working day. Under these circumstances, the search for and experimentation with efficient quality control techniques is a critical issue.

As a first test we chose the macro-aggregate of the counterparts of M3, because it has a relatively simple structure and because some of its components are subject to reporting errors. The aggregate (see Banca d'Italia, 2000) comprises 17 series referring to bank loans to the money-holding sector and banks' securities portfolios (private-sector bonds and shares). For each series and each bank the data of the first report were compared with those currently present in the Bank of Italy's database, for which the validation process can be presumed to have been completed. This was achieved by using the archives in which the individual corrections made over time by banks are stored. In this way a simulation was carried out that permitted the evaluation of the performance of an editing process that starts from the first data received from each bank, selects those to be verified, replaces them where necessary and recalculates the final indicators. Comparison of the pre- and post-editing data also makes it possible to measure the correlation between the outliers detected by the model and the errors recognized by the banks.

The data analyzed here refer to the second half of 2002, a period for which the revision of the data has presumably been completed.

4. THE PROPOSED STRATEGY

For a long time the adoption of time series modeling in the context of the data control activity was considered too costly in terms of the time and resources needed to estimate all the parameters. The widespread use of these models has been popularized by the availability of the TRAMO/SEATS software (see Gomez and Maravall, 1996), which carries out automatic identification and ARIMA model estimation reliably and with optimal time performances.[2] The TERROR module ("TRAMO for errors": see Caporello and Maravall, 2003), is a TRAMO-SEATS extension for the identification of outliers. The criterion used to pin down the outliers is the standard forecasting error: the values outside the confidence interval derived from the Student t-statistic are considered outliers. The Student t-statistic is computed as the

[2] *In the tests carried out on an aggregate made up of 17 monthly series available for an interval of 8 years, the one-step-ahead estimation procedure with a rolling horizon repeated for 7 periods took about 750 milliseconds with a UNIX RISC6000 S85 system. Times of the same order of magnitude are also obtained with a PC with RAM of 512 Mbyte and a clock rate of 1.8 GHz. This result strengthens the impression that the performance of the application is not a constraint, especially considering that in practice it would be necessary to produce a one-step-ahead forecast for each series.*

ratio between the forecasting error (estimated value – reported value) and its standard deviation.

An earlier application of this software to the monthly banking series (Farabullini, 2002) suggested a change to the criterion used to detect outliers in order to take account of their impact on the aggregate series reported to the ECB. Selective editing techniques make it possible to tackle this need in a formal and systematic way by linking the quality controls directly to the expected impact on the value of the final statistics in question.

The tests described here were carried out in a Speakeasy environment using the Modeleasy+ extensions for the interface with the TERROR software and the supplementary calculation algorithms. For each of the component series, the TERROR software was applied to the banks' first report to produce, on the basis of the ARIMA models, the one-step-ahead forecasts for the period t. This exercise was repeated for 7 months so as to be able to simulate for the same number of periods the quality control of the latest datum available of a series. Taking the monthly change in the aggregate corresponding to the "M3 counterparts" aggregate as the output statistic, the calculation was made for each period and for each component of the score function by exploiting the distances between the values observed and the values forecast by the models. The components were then ranked on the basis of the absolute values of the score hence simulating a check on the data oriented according to the priorities identified in this way. The performance of the score function was examined for the (2) and (3) formulations and similar results obtained in both cases. The next section contains a description of the output of the (2) formulation.

5. MAIN RESULTS AND FURTHER RESEARCH

Use of the ARIMA methodology with automatic identification of the model produces extremely variable outcomes in terms of goodness of fit: the relative error of one-step-ahead forecasts, calculated for each time series as an average over time, ranges from a minimum of 0.70 to a maximum of 81 per cent; excluding these two extreme values, the relative error of the remaining 15 series is, on average, 10 per cent. It should be noted that the ability of the forecasting model to identify and rank the outliers is more important than its accuracy in assessing their magnitude. Values forecasted by the model do not replace the true data but are used as a benchmark to build a ranking for further checking. The variability of the forecasting error of the model is partially due to the breaks in some of the 17 component series.[3] Adjustment series are sometimes available; however, in the context of this exercise it is more useful to assess the performances of the procedure in non-optimal circumstances.

The outcomes are briefly reported in the figures attached. Each line matches the values which the monthly growth rate that the counterparts of money would assume in an editing procedure guided by a score function: at the beginning, the growth rate is computed using the first data set provided by the reporting agents; in the following points, the growth rate is recalculated replacing, separately for each of the 17 component series, the value of the first data transmission with an adjusted or estimated value. The three lines correspond to three different editing procedures guided by three score functions; these functions are always obtained using formula (2) but with different criteria for computing the benchmark. In the case of process (a)

[3] *For example, in this version of the paper breaks due to changes in the reporting population or in the classification of customers, which imply a new make-up of the money holding sector, have not been fixed.*

the priority for checking each component depends on the distance between the real data and the forecasts of the ARIMA models estimated using the TRAMO-SEATS package; line (b) reports the "checking path" built on the outliers identified by TERROR on the basis of the standardized errors of the model; case (c) reconstructs - ex-post – the most efficient process, centered on the revisions effectively provided by the reporting agents. Figures are normalized by the final value of the growth rate which includes all the revisions sent by intermediaries; thus, at the end of the quality control procedure, the final value of the three lines is equal to 1. In absolute value, the impact of the checks on the growth rate of an aggregate anticipated by the score varies, in the period observed, between 0.1 and 3 per cent, depending on the component. If the model forecasts correctly, this difference can be interpreted as the efficiency gain obtained by applying selective editing: giving priority to the checks with the highest impacts, it is possible to stabilize the aggregate on the correct value very quickly. A brief measure of the reliability of the ranking based on ARIMA models has been obtained by comparing these outcomes with the optimal sequence built up ex-post using the real revisions sent by reporting agents. Each chart reports the relative Spearman rank correlation index.

The evidence provided by both the charts and the rank correlation index shows that major gains of efficiency in quality control data procedures can be achieved when it is possible to anticipate the size of the revisions and to measure their impact on the final statistics. In all the months for which the simulation was implemented, the speed of convergence of the estimated values towards the definitive values is higher for process (a) (which uses the score function) than for process (b) (which uses the relative errors of the model). The high level of the rank correlation between the process guided by the model and that built up ex-post on the basis of the real revisions transmitted by reporting agents provides an additional measure of the soundness of the strategy adopted. In the presence of resource constraints, the selective editing procedure can either improve the quality of the output by a given number of checks or allow the number of checks to be reduced with no significant information losses regarding the final statistics.

Starting from this initial approach, it is possible both to build more complex score functions and to extend the methodology to the elementary series provided by reporting agents. For example, a different a priori error can be inserted in the score for aggregates or for intermediaries. It is also possible to collapse different scores for different final statistics into a global score which takes account of the needs of accuracy even in the case of statistics that differ in terms of subject matter, aggregation level or function form. The probability of errors in the observed values is connected with the size of the relative error forecast, which represents the distance between the last reported value and the 'normal' trend of the series. In order to check new data, it seems useful to combine the score approach with the traditional method based on the t statistic for assessing the significance of forecasting errors. This combination can be turned into an index of the following type:

$$\text{Werror}(t) = |\text{score}(t)| + \text{alpha} * |\text{relerr}(t)|$$

where relerr(t) is the t statistic given by the ratio between the forecasting error for each observation at time t and the standard deviation of the forecasting errors of the complete time series.

This synthetic index can be interpreted as a sign of deviance corrected by the impact on the final statistic or as an index of the impact of the outliers corrected by a measure of the irregularity. The weights of the two addends should be determined on the basis of experience and special events might require the threshold of just one or both factors to be raised. In principle, the weight to be assigned to the relerr(t) parameter should represent an adjustment of the ranking obtained using the score function in order to take accurate account of the outliers which are relatively small but have a major impact on the final statistic and the quality of each time series.

In order to carry out a parallel analysis of the two factors, the output of the quality control procedure could be organized to permit different options in the last step of the TERROR application. In particular, after the data have been classified by means of TERROR as normal, possible error and likely error, on the basis of the t statistic and the level of sensitivity defined by the user, the additional options could provide - alternatively – for the printing of:

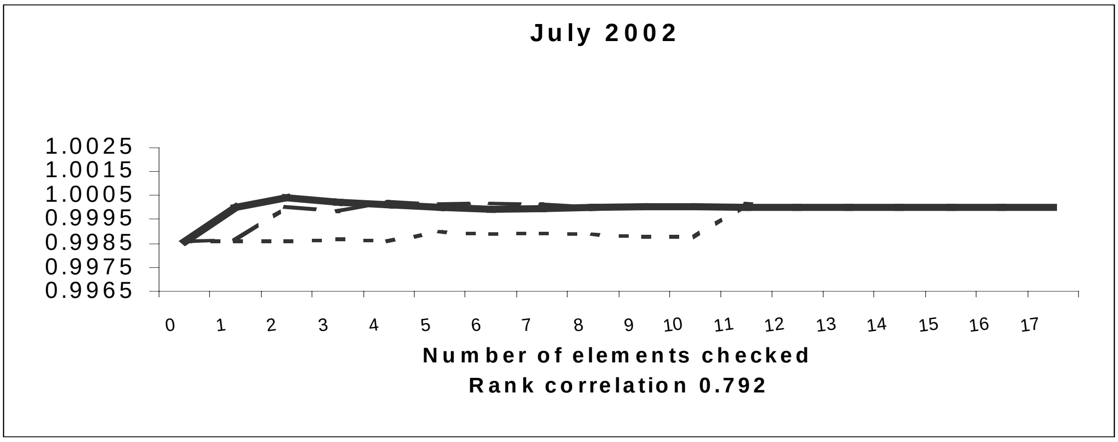
- all ‘possible’ or ‘likely’ errors – for the selected threshold of sensitivity – in decreasing order of the absolute value of the score;
- all ‘likely’ errors – for the selected threshold of sensitivity – in decreasing order of the absolute value of the score;
- data with an absolute score value > x (for example x=0.01 if the score refers to a ratio), regardless of the size of the relative error.

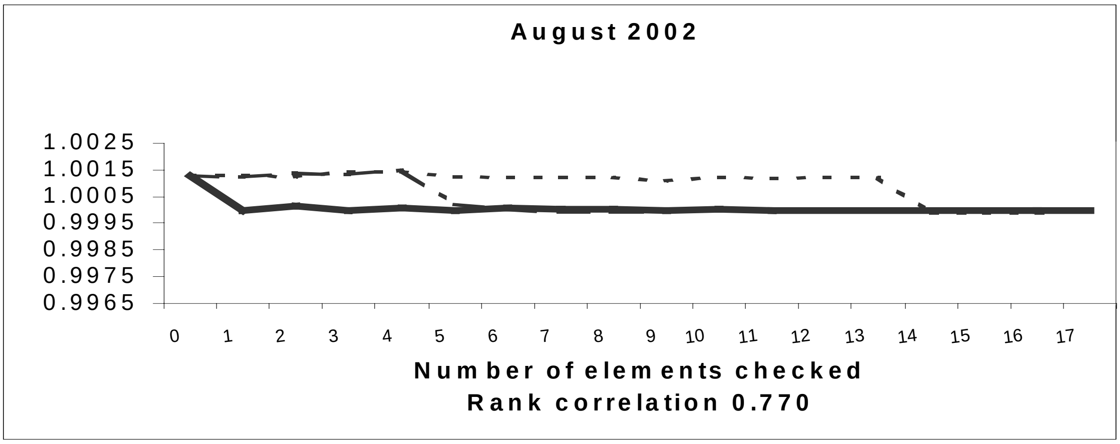
The number of suspected errors printed could easily be increased or decreased by changing the level of sensitivity and the degree of “abnormality” to be shown.

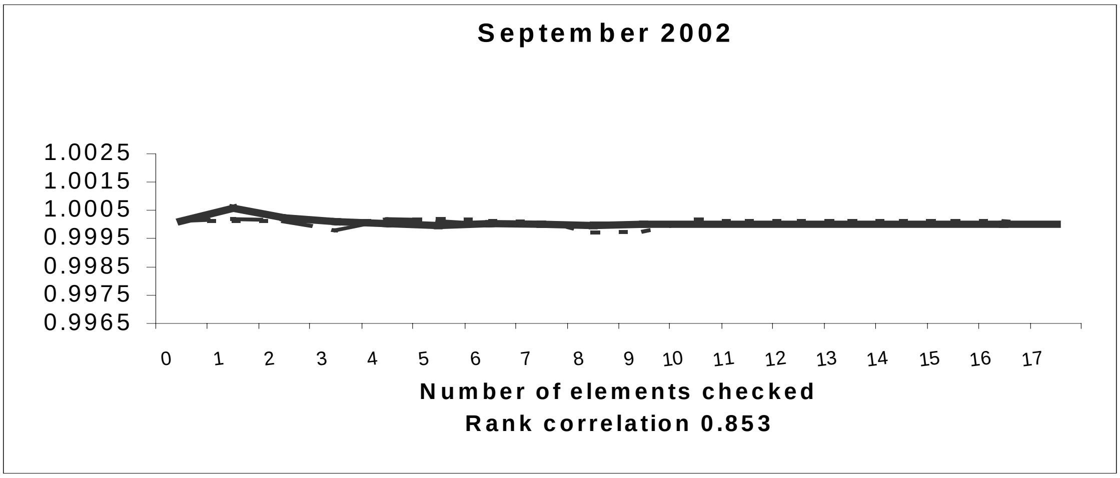
Further tests on the performance of the score function and the extension of the methodology to the data provided by each reporting agent are the natural goals of future research.

References

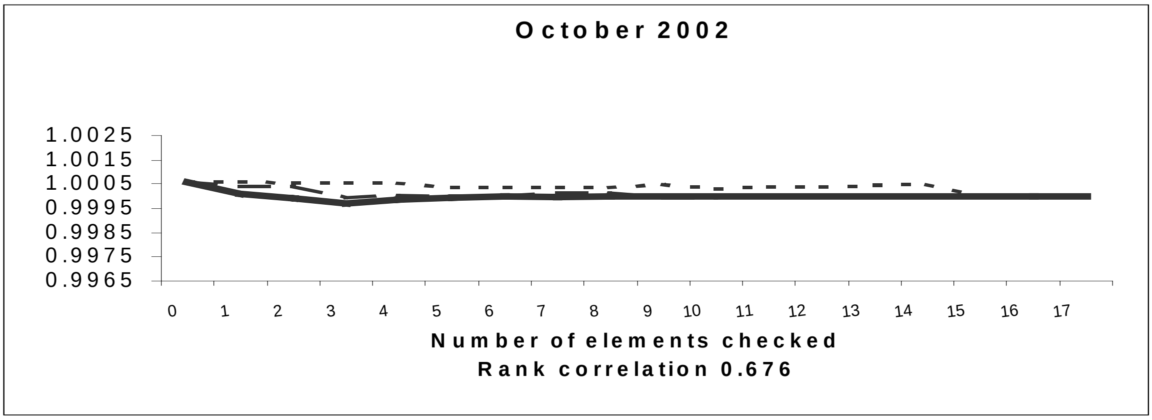
1. Alwan, L.C. e Roberts, H.V.: “Time Series Modeling for Statistical Process Control”, *Journal of Business and Economic Statistics*, No. 6, pp. 87-95, 1998.
2. Banca d'Italia: *Supplementi al Bollettino Statistico: note metodologiche e informazioni statistiche*, No. 33, 2000.
3. Barnett V. and Lewis T.: *Outliers in Statistical Data*, Wiley and Sons, 1998.
4. Caporello, G. and Maravall, A.: *A Tool for Quality Control of Time Series Data – Program TERROR*, 2003.
5. Cheng, B. and Titterington, D.M.: “Neural networks: a review from a statistical perspective”, *Statistical Science*, No.1, pp.1-54, 1994.
6. Delgado, A. and Prat, A.: “Modeling time series using a Hybrid System: Neural Networks and Genetic Algorithms”, in Bellacicco A. and Lauro N. (eds) *Reti Neurali e Statistica*, Franco Angeli, 1997.
7. Farabullini F.: “I modelli ARIMA per la stima e il controllo delle statistiche delle istituzioni finanziarie e monetarie”, *Bancaria*, No. 2/2002, pp. 36-48, 2002.
8. Gomez, V. and Maravall, A.: “Programs TRAMO (Time series Regression with Arima noise, Missing observations, and Outliers) and SEATS (Signal Extraction in Arima Time Series). Instructions for the User”, Working Paper 9628, Servicio de Estudios, Banco de Espana, 1996.
9. Landenna, G. and Marasini, D.: *Metodi Statistici Non Parametrici*, Il Mulino, Bologna, 1990.
10. Latouche M. and Berthelot J-M.: “Use of a Score Function to Prioritize and Limit Recontacts in Editing Business Survey”, *Journal of Official Statistics*, Vol. 8, No. 3, 1992.
11. Lawrence D. and McKenzie R.: “The General Application of Significance Editing”, *Journal of Official Statistics*, Vol. 16, No. 3, 2000.
12. Luzi O. and Pallara A.: “Combining Macroediting and Selective Editing to Detect Influential Observations in Cross-Sectional Survey Data”, *Statistical Data Editing*, ISTAT, No. 6, 2000.

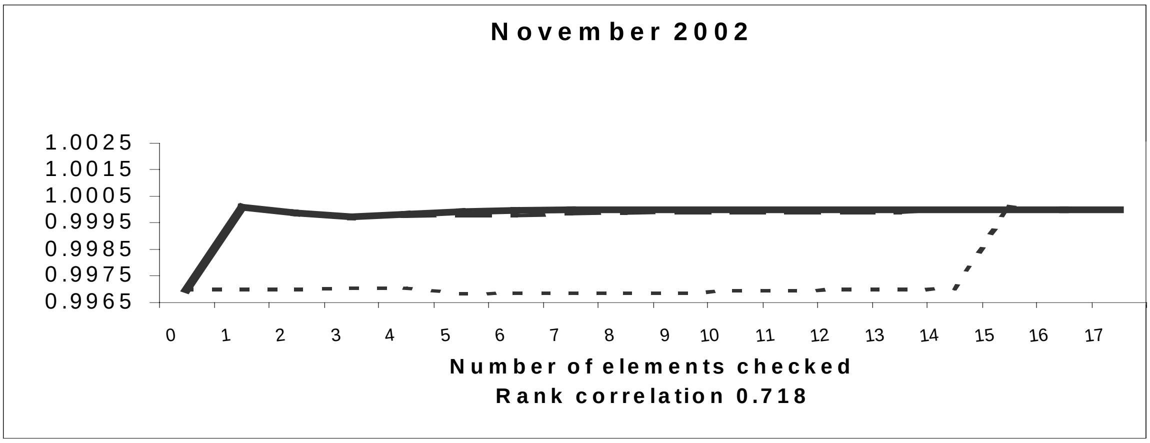


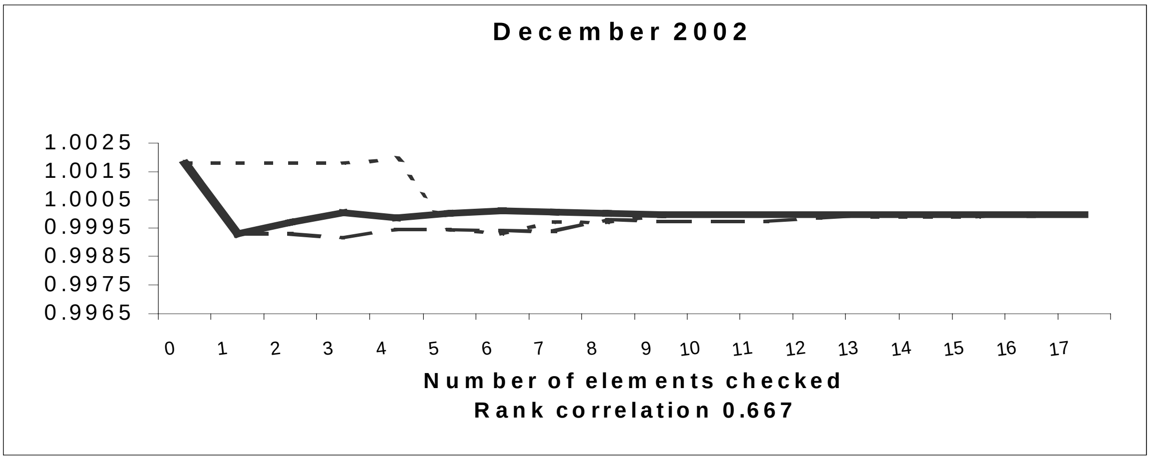




Legend:
Priority of checks depending on:
__.__.__._ (a) score
………… (b) forecast relative errors
________ (c) revisions provided by reporting agents







Legend:
Priority of checks depending on:
__.__.__._ (a) score
………… (b) forecast relative errors
________ (c) revisions provided by reporting agents