Catalogue no. 11-522-XIE

## Statistics Canada International Symposium Series - Proceedings

# Symposium 2003: Challenges in Survey Taking for the Next Decade

2003


Statistics Canada Statistique Canada

Canada

Proceedings of Statistics Canada Symposium 2003
Challenges in Survey Taking for the Next Decade

# REPORTING ON THE QUALITY OF SEASONAL ADJUSTMENT PROCESS AT THE STATISTICAL OFFICE FOR THE EUROPEAN COMMISSION

Jean-Marc Museux and Nina Jukic[1]

## ABSTRACT

Seasonally adjusted (SA) figures are the most used source of information for users of official statistics, most notably analysts and researchers. Eurostat is providing European Union and EU-Zone with economic seasonally adjusted series. The difficulty to reach convergence between Member States' and Eurostat's SA policy makes the need to provide users with meta data on the SA process even more striking. Eurostat has recently undertaken to report internally on the quality of SA process. The paper underlines the structure and the quality measures used. Four different quality measures can be envisaged: a) quality measures for research and evaluation criteria that can be used for software comparisons for example; b) quality measures for practitioners of SA, who need a set of indicators to tune their SA and to decide how SA should be set up for an individual series; c) quality measures for monitoring the SA process and production, guidelines, etc.; d) quality measures for "end users", who are looking at published data and are not interested in diagnostics. As far as possible, quality measures should be methods independent. These quality measures are compiled in multi layers quality report.

KEYWORDS: Quality Report; Seasonal Adjustment.

## 1. INTRODUCTION

### 1.1. Eurostat's compilation of seasonally adjusted figures

Seasonally adjusted (SA) figures are the most used source of information for users of official statistics and economic statistics, most notably analysts and researchers. Eurostat, the Statistical Office for the European Commission, is providing European Union (EU) and Euro-Zone with economic SA data. EU data are compiled from national data. The difficulty to reach convergence between Member States' (MS) and Eurostat's SA policies makes the need to provide users with metadata on the SA process even more striking. Eurostat has recently undertaken to document the quality on the SA process on a systematic basis. This paper presents the proposal that was supported by the management committee of Eurostat.

### 1.2. Background

Many EU bodies (National Statistical Institutes, supra-national organisations) have defined rules and guidelines for ensuring the quality of the output of the SA process. Two interesting examples at supra-national level are the European Central Bank's (ECB) quality report (QR) (ECB, 2000) and the proposal made by C. Planas at the Eurostat quality working group on SA (Planas, 2001). However, both of them verged too much on one specific method, the non-parametric X12-Arima (X12) method or the model based Tramo-Seats (TS) approach, respectively. Given these

[1] Jean-Marc Museux, Statistical Office for the European Commission, Unit B5-Research, BECH A2/157, L-2920, Luxembourg, Jean-Marc.Museux@cec.eu.int; Nina Jukic, Statistical office of the Republic of Slovenia, Vozarski pot 12, P.O.B 3570, SI-1001 Ljubljana, Slovenia, Nina.Jukic@gov.si
The views expressed by the authors do not necessarily represent the official position of Eurostat

two methods will continue to coexist in the future among EU institutions, it was not possible to find a consensus on the design of a standard QR for EU institutions.

In 2002, Ladiray *et al* (Ladiray, Museux, 2002) came with an idea of defining a method largely independent QR. One of the conclusions was: despite there was a possibility to define statistics relevant for both of the methods, there was still need for further methodological development.

In the meantime difficulties raised by lack of convergence of MS' and Eurostat's policies which has led Eurostat to define a unilaterally SA policy for the EU-indicators. Eurostat will use the direct adjustment based on the raw data and will report on the discrepancy of the direct SA policy and the mixed indirect approach based on national data. The need to document its own SA process is thus more than ever striking. Furthermore the decision was taken by the management to implement a pragmatic QR based on the output of the two main current softwares. Although Eurostat is actually mainly using TS for the production of the data, the QR has to be, as much as possible, method independent, to be in line with the Eurostat's recommendation for the use of both, TS and X12, methods equally.

# 2. THE STRUCTURE OF THE QR

## 2.1. The QR, what for

The objectives of the QR are two-fold.

First, to gather information on the method used and the methodological choices made. The basic idea is to ensure the reproduction of the output ensuring de facto transparency of the procedure.

Second, to assess the quality of the output according to recognised quality criteria. The quality of the output depends obviously on the quality of the input. The QR should thus also provide information to allow benchmarking output against input.

## 2.2. The QR, for who

The QR can have different types of users. One could distinguish between:

(1) "End users" of SA data (e.g. analysts), who want to gain some evidence of the quality of the output.
(2) Producers of SA data, who want to control the process of SA to ensure its quality and to be able to report on it to the "end users".
(3) Producers of SA data, who want to design the best procedure for SA (for instance the choice of the filter.)
(4) Researchers, who want to assess and evaluate new methods for SA.

Some of the measures of the quality of SA are common for different categories of the users but at the same time not all of them are equally detailed.

This QR is designed mainly for the first two types of users. Because it should remain close to the users, emphasis should be put on statistics familiar for the "end users".

## 2.3. Multi-level

The multiplicity of the usage and of the facets of the QR inevitably implies that information should be structured in a way to improve its readability. Ladiray (Ladiray, Museux, 2002) envisaged three levels of the QR:

*Score board*

The score board is a general management tool. It reports on the process of batch series as a whole. It should present general statistics. It does not describe the quality of individual SA series, but rather an average performance of the batch runs. Possibly, it produces a list of potential difficult series defined according to some of the quality measures used in the QR for individual time series.

Different implementations of the score board already exist in TSW, X12 and Demetra. (Afterwards we will highlight some quality measures that should be reported in the score board). Its implementation will not be further discussed here.

*Summary*

The summary should be limited to the most important measures for each theme of the QR. The summary level should be regularly updated for the most important and difficult series. It should be available for the users of the SA.

*Detailed information*

The detailed information should be compiled at regular intervals, say, once a year when the parameterisation is revised. It should contain all relevant information for a detailed picture of different themes of the QR and to base judgement for improving the procedure.

## 2.4. Method independent

As far as possible the provided measures should be method independent and easily interpretable by the users. For instance, the statistics in the output of the non-parametric X12 are well known for the users. Ladiray showed that most of them can be computed for TS as well and that some of them are relevant to qualify the SA output. On the opposite, X12 can be translated in a model based perspective (TS) and model based statistical tests can be derived to measure the quality of the SA process. In this domain, till recently, the model based measures were scarcely available and difficult to interpret for the “end users”. Some theoretical advances are still required. However, the first results (Maravall, 2002) are encouraging.

This QR has to be pragmatic and will not wait until a general consensus is reached in the SA community. Thus, rather than developing new measures, it borrows what we think are the most relevant statistics of the two approaches, using existing routines and making heavy use of cross computation using both of the routines. The practical implementation of it will be discussed at the end of the paper.

## 2.5. Quality of SA: key requirements

A general consensus seems to have emerged on the key aspect of the quality of SA process.

A good SA process should:

(1) Not leave any residual seasonality and the effects that it has corrected (trading day (TD), …) in the raw data. The idempotency of non-linear SA process has to be assessed: the procedure applied to the SA series should leave the signal unchanged.
(2) Not lead to abnormal revisions in the SA figures with respect to the characteristics of the series.
(3) Be transparent. The underlying choices should be documented.

The quality measures should be accompanied with technical measures which indicate that the method used could adequately adjust the series.

# 3. OVERVIEW OF THE QR BY THEME

## 3.1. Introduction

The QR is structured according to the themes. Each theme is summarised by a few statistics and graphs in the summary part. It is further detailed in the detailed part.

## 3.2. Description of the SA procedure

In this part of the QR the parameterisation should be defined, in order to be able to reproduce the same output. It mainly refers to the metadata.

The summary provides general statement about the SA strategy:

- ✓ ***The SA method used (X12 or TS)***
- ✓ ***Sample dates and number of observations***
- ✓ ***Seasonal factors' last update***
- ✓ ***Sample dates for computation of calendar effects***
- ✓ ***Last update of the model***

Specification of the aggregation method used and the components, if relevant:

- ✓ ***Direct/indirect method with components***

Specification of the main options linked to the procedure itself (mainly method independent)

- ✓ ***Transformation of the data (log, box-cox)***
- ✓ ***Decomposition model (multiplicative, additive, log additive)***
- ✓ ***Adjustment to annual totals or any post adjustment coming from the benchmark of monthly and annually data***
- ✓ ***The list and the type (Additive, Temporary Change, Level Shift) of outliers detected with possible explanation, when available***
- ✓ ***The list and the type of regressors used (TD, Easter, specific holidays, others)***

The detailed part mainly contains the full specification of the procedure, which is method dependent, with:

- ✓ ***The input string for the program***
- ✓ ***Meta information on the way the program operates (option for automation)***
- ✓ ***The detailed description of the regressors***
- ✓ ***The time span used to estimate the models***

## 3.3. Description of the input

The quality of the SA procedure and its output is mainly dependent on the state of the input. The SA process may be affected by the strong revision or missingness of the raw data. Too high level of volatility of the time series can also lead to difficulties in the identification of the components.

The quality of the output has to be benchmarked against the quality of the input.

The summary QR proposes a graph of the time series, to allow the user to catch intuitively the nature of the difficulties. The graph is accompanied with some rough descriptive measures:

- ✓ ***Mean and standard deviation of the series***
- ✓ ***Standard deviation relative to trend computed as a symmetric Henderson MA (13 terms)***

- ✓ ***MCD statistics which is a measure of the volatility of the series, as implemented in X12***

At the same time it should allow the user to catch the nature of the autocorrelation pattern of the series. Therefore, the document reports:

- ✓ ***The significant peaks of the autocorrelogram of the raw series***
- ✓ ***The Box-Pearce statistics for seasonality and its level of significance***
- ✓ ***F-test for seasonality (X12 implementation)***

This information would ease the user to catch the presence of non-stationarity and seasonality.

If the raw data suffer from frequent revisions, the SA output is likely to suffer from the same problem. Revisions in the raw data are measured with:

- ✓ ***The percentage changes defined on the last three years of the series***
- ✓ ***The percentage changes of the last revisions of the most recent data***

This theme is further documented into the detailed part by means of graphical information:

- ✓ ***The autocorrelogram of the raw, the first differenced and the first seasonal differenced series***

## 3.4. Adequacy of the procedure

This part of the report provides information which indicates that the method has been adequately applied. This part concerns mostly producers of SA data and "end users" who are familiar with SA procedures. It is likely to be method dependent, however, the summary part focuses on common issues. The detailed part is more specific to the method.

*Adequacy of the ARIMA modelling:*

Although less crucial in X12, the use of ARIMA modelling is a common quality feature of both procedures. In the summary report its quality is characterized by:

- Statistics on residuals of the model:
- ✓ ***Normality test***
- ✓ ***Ljung-Box Q-statistics***
- ✓ ***Pierce Q-statistics***

- Fitting and forecasting performance of the model:
- ✓ ***In sample forecast error (last 2 years)***
- ✓ ***Percentage of outliers***
- ✓ ***Out of sample percentage error (2 last years)***

In the detailed part, the information is complemented with the graphical information about the series of residuals of the model, their correlogram and the coefficients of the ARIMA model as a result of an automated identification procedure with their corresponding t-statistics and level of significance. Global quality of fit statistics such as BIC and AICC can be also found there.

*No residual seasonality and calendar effect*

The performance of procedure is measured in the light of its ability to remove all the desired components from the raw data (actually, the seasonal component and the related calendar effects). The approach developed here implements the idempotency principle and provides an in depth analysis of the irregular component. The irregular component should present neither residual calendar effect nor seasonality. Special care is taken in order to take into account spurious autocorrelation patterns of the estimator of the irregular component (Maravall, 2003).

In the summary part, one finds:

- ✓ ***An indicator of significant positive autocorrelation at seasonal lags (the presence of negative autocorrelation being generally a spurious effect of the estimation of the irregular component)***
- ✓ ***t-statistics for the autocorrelation at lag 12 and the level of significance***
- ✓ ***F-test for seasonality (X12 implementation)***
- ✓ ***Indicator of significant calendar effect in the irregular component***

In the detailed part, one finds the frequency analysis of the irregular component as provided by X12 procedure:

- ✓ ***The power spectrum of the irregular component***
- ✓ ***The number of significant seasonal peaks***
- ✓ ***The number of significant TD peaks This information requires specific X12 run on the integrated residuals***

Further details on the regARIMA modelling of the irregular component can also be found.

- ✓ ***The list of t-statstics related to the TD/Easter/holidays coefficients estimated in an independent run on the irregular with full specification***

*Stability of the procedure over time*

X12 provides extensive study of the stability of the procedure. The sliding span statistics is derived from running SA with different spans of data from the same time series, with the same SA parameter settings. Estimates of seasonal factors of the observations common to different spans are then compared.

In the detailed part, the analysis is complemented by measures of stability of the current model specification over the last year with:

- ✓ ***The history of automatically identified ARIMA models over the last year***
- ✓ ***The position of the concurrent model parameters estimates within 95% confidence interval from the last update of the model***

*Direct versus indirect adjustment*

The debate on the direct versus indirect SA is not closed yet (Ladiray, Mazzi, 2002). Rather than assessing the use of one approach (which is actually done everywhere else throughout this report) this section highlights the possible difference between them, drawing the attention of the user towards careful inspection of the different adjustments.
The summary part focuses on the deviation between the two approaches looking at the SA series and the deduced growth rate series:

- ✓ ***Average absolute percentage deviation between directly and indirectly SA series***
- ✓ ***Highest absolute percentage deviation between directly and indirectly SA series***
- ✓ ***Level of concordance of the sign of the growth rates between direct and indirect SA series (percentage and list of dates of divergences)***

The detailed part further investigates the concordance of growth rates with:

The level of concordance of growth rates for: SA series

- ✓ ***Direct SA series and SA components***
- ✓ ***Indirect SA series and SA components***

The following descriptive statistics of SA series:

- ✓ ***Mean of differences in growth rates***
- ✓ ***Minimum of differences in growth rates***

- ✓ ***Maximum of differences in growth rates***
- ✓ ***Range of differences in growth rates***
- ✓ ***Variance of differences in growth rates***
- ✓ ***Bar chart of growth rates***

*M-Statistics*

In the detailed part of the QR M-statistics are produced. For X12 they are critical for the assessing the procedure. For TS they are produced for the purpose of characterising and benchmarking the TS output with X12 run. They are not used as quality measure per se.

- ✓ ***M1***
- ✓ ***M2***
- ✓ ***M3***
- ✓ ***M4***
- ✓ ***M5***
- ✓ ***M6 ( for X12 only)***
- ✓ ***M7***
- ✓ ***M8***
- ✓ ***M9***
- ✓ ***M10***
- ✓ ***M11***
- ✓ ***Q statistics***

## 3.5. Characterisation rather than quality measure of the output

*The “Smoothness” of the output*

The degree of smoothness of time series should not be taken into account in the context of SA. However, the SA procedure that provides more constant signals of upward or downward tendency is prefered. Three measures, borrowed from business cycle analysis (Gomez, Maravall, 1999), are reported here:

- ✓ ***Mar(S)***
- ✓ ***Mar1(TC)***
- ✓ ***Mar2(TC)***

*The moving seasonality*

The model based approach handles this feature without particular issue. X12 is less robust regarding this aspect and a specific quality criteria,

- ✓ ***M7***

is introduced to check the presence of seasonality.

*The level bias*

Unless permanent variation in seasonal components are present for every year, the original and the SA series, as well as the trend-cycle component are likely to have equal mean value. The respect of the accounting constraint of equality between these annual sums is mainly relevant for what concerns the dissemination of data. It should not be considered as a theoretical quality criteria. Therefore, in this perspective, it is worth to evaluate the size of these discrepancies with

- ✓ ***Relative annual difference of original and SA series.***

*The revision analysis*

Revision analysis is probably the central issue for the "end users" of SA data. The revisions in the SA can occur due to the change in the raw data, but they are also a consequence of the SA process because of the natural need to take into account all the information available at a given moment.
In the summary report, focus is put first on the actual revisions of the released SA data, including both revisions in the raw data and revisions linked to the SA process:

- ✓ ***The average expected percentage absolute revisions in the released SA data***

The component of the SA procedure is singled out in the second statistics:

- ✓ ***Average expected percent absolute revisions in the SA data as implemented in X12***

The revisions, as computed with X12, are the observed difference between the concurrent estimate of the SA series and the final estimate whereas the model is being fixed.

The detailed part contains in addition:

- ✓ ***The revisions and the average absolute percentage revisions in the growth rates of the SA series, concurrent trend, month-to-month changes, percentage changes in trend, concurrent and predicted seasonal factors, as defined in X12***
- ✓ ***The theoretical rate of convergence of the revisions error corresponding to the model, as it is defined in TS, complements the analysis***

## 4. CONCLUSION

This attempt to define a QR independent of the method was meant to be a pragmatic exercise with heavy methodological implication. It was drawn by the need to define a unified framework to report on and document the SA process. It will be an essential tool for providing metadata to the users. It clearly expresses the need for further methodological development, for defining common measures of the quality for both, X12 and TS and further integration of the programs in a unified quality framework.

A pilot implementation of the QR is provided in annex. Its compilation requires heavy manual implementation (cross run of X12 and TS, retrieval of selected output). The compilation clearly requires an automatic tool. A SAS implementation could be developed easily and will be useful for testing the QR in the production. However such an implementation is relatively heavy in contrast with its conceptual simplicity. It clearly claims for a modular implementation of the SA routines with a well structured output and re-usable module for allowing the development of tailorable quality report.

This report focuses on the SA process which is crucial at the ESS level. However, it is just one aspect of the report on the quality of short term indicators. In this broader perspective, it is certainly too detailed. The report on SA aspects should focus only on the few key aspects relevant to the users such as revisions and metadata on the method used.

## REFERENCES

ECB (2000), "Seasonal Adjustment of Monetary Aggregates and HICP for the Euro Area", Working Paper, European Central Bank.

Gomez, V., Maravall, A. (1999), "Application of Tramo-Seats: Example 2 and 4", unpublished manuscript, Bunbesbank Seminar.

Ladiray, D., Mazzi, G.-L. (2002), “Seasonal Adjustment of European Aggregates: Direct versus Indirect Approach”, paper presented at the Seasonal Adjustment Seminar, European Central Bank.

Ladiray, D., Museux, J.M. (2002), “Quality Report for Seasonal Adjustment: Some Ideas”, paper presented at the 5th meeting of Informal working group on Seasonal Adjustment, EUROSTAT.

Maravall, A. (2002), “Diagnostics in TRAMO-SEATS and X12-ARIMA: model based versus non-parametric approaches”, paper presented at the Seasonal Adjustment Seminar, European Central Bank.

Maravall, A. (2003), “Notes on Programs TRAMO and SEATS, Part III, Signal Extraction in ARIMA Times Series”.

Planas, Ch. (2001), “Proposal for Quality Report for Seasonal Adjustment”, paper presented at the 4th Meeting of the Assessment of Quality in Statistics Working Group, EUROSTAT.