ISSN 1725-4825

2006 EDITION

WORKING PAPERS AND STUDIES

Statistics Canada Experience Towards New Standards for Time Series Processing

Conference on seasonality, seasonal adjustment and their implications for short-term analysis and forecasting

Luxembourg, 10-12 May 2006

EUROPEAN COMMISSION

EURO INDICATORS

eurostat

THEME
General
and regional
statistics

Europe Direct is a service to help you find answers to your questions about the European Union

Freephone number (*):

00 800 6 7 8 9 10 11

(*) Certain mobile telephone operators do not allow access to 00-800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server (http://europa.eu).

Luxembourg: Office for Official Publications of the European Communities, 2006

ISBN 92-79-03417-0
ISSN 1725-4825
Catalogue number: KS-DT-06-018-EN-N

STATISTICS CANADA EXPERIENCE TOWARDS NEW STANDARDS FOR TIME SERIES PROCESSING

By Susie Fortier[1] and Benoît Quenneville[2]

Statistics Canada, Time Series Research and Analysis Centre

The X-11-ARIMA seasonal adjustment method and software package is widely used in Statistics Canada and in its survey programs. X-11-ARIMA was developed under the leadership of Dr. E.B. Dagum at Statistics Canada in the 1980s. Since that time, a number of improvements to the methodology, combined with new needs by client divisions, have created numerous situations where computations must be performed outside the system before the raw series are entered into X-11-ARIMA. Following seasonal adjustment by X-11-ARIMA, many series also go through a benchmarking and/or a raking/reconciliation procedure to ensure that the adjusted series respect various aggregation constraints. In parallel, the U.S. Census Bureau has developed X-12-ARIMA, which is gradually replacing X-11-ARIMA in many agencies as the preferred method of seasonal adjustment. X-12-ARIMA includes several new modeling features which would eliminate the need for much of the pre-processing now required, since these functions can be handled within X-12-ARIMA itself. Furthermore it retains all the features that are available in X-11-ARIMA. This paper gives an overview of converting to X-12-ARIMA and updating the benchmarking and reconciliation methodologies and software at Statistics Canada. We illustrate how the production system for the International Travel Survey (ITS) has been updated to serve as a prototype. The tasks for the transition of the seasonal adjustment production systems from X-11-ARIMA to X-12-ARIMA, including updating benchmarking and reconciliation, are discussed.

Keywords: Seasonal Adjustment, Raking, Benchmarking.

1 Background information

At Statistics Canada, the X-11-ARIMA method from Dagum (1981) and its successors, X-11-ARIMA/88 from Dagum (1988) and X-11-ARIMA version 2000 have been the official seasonal adjustment methods since 1980. Since that time, a number of improvements to the method, combined with new needs by client divisions, have created numerous situations where computations must be performed outside the system before the raw series are entered into X-11-ARIMA. For example, the International Travel Survey (ITS) produces monthly data at different levels on the number of travelers to and from Canada. Recent world events such as terrorists' attacks and SARS outbreaks had a large impact on this industry. To properly seasonally adjust such series, the impact of the disturbing

[1]Susie Fortier, 150 Tunney's Pasture Drive, R.H. Coats Building, 17th floor, Ottawa, Ontario, Canada, K1A 0T6, tel. +1-613-951-4751, e-mail Susie.Fortier@statcan.ca. Paper presented at the EUROSTAT conference on Seasonality, Seasonal Adjustment and Their Implications for Short-Term Analysis and Forecasting, Luxembourg, May 2006.

[2]Benoit.Quenneville@statcan.ca .

events had to be estimated outside X-11-ARIMA, then input as temporary adjustments. Considerations for moving holidays such as the Chinese New Year also had to be estimated outside the regular package. Other examples of outside estimation for prior adjustments include reference-week adjustments for the Labor Force Survey and level-shift adjustments for survey redesign.

After seasonal adjustments, many series at Statistics Canada go through a raking procedure to ensure that the seasonally adjusted (SA) series respect various aggregation constraints. They could also be benchmarked to annual totals from an external source (say an annual survey on the same target population) before or after being SA. Many different versions of the software required to do these types of adjustments were developed throughout the years. Each version was tailor-made to survey programs to meet their specific needs or to incorporate the most current methodological development of the time. The numerous versions of the software, as well as the complex skills required to maintain them, made the whole system extremely hard to manage.

X-12-ARIMA developed by Findley, Monsell, Bell, Otto, and Chen (1998) at the US Census Bureau includes all the capabilities of the latest version of X-11-ARIMA and new improvements such as new modeling capabilities and new diagnostics. It uses linear regression models with ARIMA residuals (regARIMA models) to pre-adjust and forecast time series before their seasonal adjustment. These new modeling capabilities are of particular interest for Statistics Canada as they do solve most of the outside estimation issues of prior adjustments required with X-11-ARIMA. Chhab Alperin and Quenneville (2004) present issues related to the adoption of X-12-ARIMA as the official seasonal adjustment method at Statistics Canada and an initial transition strategy. This strategy can now be summarized in 8 points:

1. Build a prototype for a SA system;
2. Review and update the raking methodology and software;
3. Review and update the benchmarking methodology and software;
4. Review teaching material;
5. Redefine guidelines for seasonal adjustment and trend-cycle estimation of individual series and for seasonal adjustment of a coherent system of series;
6. Develop an Expert System to assist in the selection of seasonal adjustment options;
7. Review the mandate of Statistics Canada Time Series Research and Analytical Centre (TSRAC);
8. Deploy the new SA system to the various divisions at Statistics Canada;

Each point is discussed below.

2 Prototype Description

We needed a survey with a sufficient number of series to cover most of the options of the SA program. As mentioned above, ITS already had some issues with estimation of prior adjustments (outlier effect such as SARS, moving holiday effect such as Chinese New Year). It also publishes data according to many breakdowns (country of origin, mode of transportation, duration of stay...) so we had some options to SA the many aggregates (indirect adjustment or direct adjustment with reconciliation). As it was also going through a redesign and was reviewing most of its system, it was the perfect choice to use to build Statistics Canada's new SA prototype.

A simplified sketch of the prototype is provided in Figure 1. We used the divide-and-conquer approach and created many self-contained ©SAS modules to achieve our goal. The foundation of the system, also referred to as the critical path, is composed of the following three modules : SA with X-12-ARIMA, Benchmarking module, and Raking module. The core of the system, in the context of SA, is X-12-ARIMA. We run the US Census Bureau program through an X command in SAS. It should be noted that ©SAS-ETS does include a PROC X12 which implements the X-12-ARIMA methodology. However it does not include all of the regARIMA modeling capabilities yet, one of the major advantages offered in X-12-ARIMA. When it does, TSRAC intents to investigate the possibility of using it in the production systems. We still expect to use the US Census Bureau version for development, including initial selection of SA options. One advantage of switching to SAS in the future is the uniformity of systems. Most of the survey methodology in Statistics Canada is implemented in SAS.

The X12 module of the prototype comes with two sub-modules: the pre-X12 and post-X12 modules. These modules mainly validate the inputs and verify the outputs. For example, the pre-X12 step will verify that all the required input files (raw data) actually exist and that they are in the right input format. The post-X12 step will look through log and error files created by X-12-ARIMA and will warn the user in the case of potential problems. The pre-X12 also permits pre-treatment that cannot be done in X-12-ARIMA. Although we would like to reduce usage of external pre-treatment to a minimum, experience indicates that there is a need to include it in the prototype as a safeguard. Adding a constant to all data points of a raw series before seasonal adjustment is an example of such pre-treatment. This allows the use of the multiplicative decomposition model on series with some zero values as described by Chen and Durk (2005). The constant is then removed from the SA data in the post-X12. To minimize the use of external adjustments, we ask our colleagues of the US Census Bureau to include new options directly in X-12-ARIMA. Issues raised by Statistics Canada are always answered in a timely and efficient matter by the X-12-ARIMA developers' team led by Brian Monsell. The particular example above is now available in X-12-ARIMA version 0.3 (argument `constant` of spec `TRANSFORM`).

The raking and benchmarking modules will be described in their own section. The other three modules are not part of the critical path and they are yet to be fully developed and implemented. The trend-cycle is often used as auxiliary information on SA data or *in lieu* of SA in the case of very irregular series. Some survey programs publish trend information, mostly as graphs. Because of possible raking and revision strategy, the trend-cycle can not be computed with the initial SA data in the X-12-ARIMA step. It

has to be computed in a later step, by the trend module on the *published* SA data. We plan to use X-12-ARIMA with trend mode, mimicking the two steps method of Dagum (1996) already in use at Statistics Canada in the X-11-ARIMA environment.

The module of Graphs and Analytical Tools will be mainly used by subject-matter analysts. Right now, analysis of SA series is done independently in the various subject matter divisions, with support by the TSRAC staff if required. This means that each subject matter division develops its own analytical tools. We want to investigate what common analysis is done by the various experts and the possibility of developing a generalized tool that would include common statistics such as growth rates, basic overlay plots, and automatic identification of most important period-to-period changes. The short-term goal is to help the subject matter experts prepare the analytical documentation accompanying the official release of the data. In the long run, generalized systems are expected to have lower maintenance cost.

The last production module not included in the critical path is the variance estimation module. Often seen as the Achilles' heal of the non-parametric SA methods, variance estimation of SA data with X-11/X-12 has been researched by many, as reviewed and developed in Bell and Kramer (1999). We intend to put the results of research by **?** on the topic as well as variance consideration in cases where other time series processes like raking and benchmarking are applied.

3 Raking

As mentioned in the introduction, many series at Statistics Canada go through a raking procedure to ensure that the seasonally adjusted (SA) series respect various aggregation constraints. This reconciliation process can be seen as an alternative to using indirect adjustment when an additivity relationship must be preserved. There is a dire need to standardize and fully document the many raking programs in production at Statistics Canada. The recent publication of **?** on Benchmarking and Reconciliation goes a long way to document the most up-to-date methodological results on the topic. In the prototype, we implement a simplified two-steps approach presented by **?**:

- Step 1: benchmark SA data to raw annual totals while preserving month-to-month changes as much as possible. As an example, this can be obtained from table `D 11.A` in X-12-ARIMA version 0.3 using the new `FORCE` spec, as described in **?**.
- Step 2: reconciliation using a specific simplified case of the **?** regression-based method.

The second step is done through weighted regression under constraints. In many cases, it is possible to obtain the results with the usual regression procedures such as PROC REG in SAS. For the cases where the off-the-shelf procedure can not compute the required solution, we are writing new specifications to have a raking program developed by computer specialists from the System Development Division (SDD) of Statistics Canada. Using SDD expertise is expected to make program maintenance easier. The prototype

currently includes a raking module with the use of SAS PROC REG. The **?** regression-based method and the two-steps approach give a solution in real numbers. A controlled-rounding procedure is implemented in the module to get integer numbers, if required. The module also computes indirect seasonal adjustment and implements revision strategy.

4 Benchmarking

Details on the benchmarking methodology are given in **?**. This module was not originally included in the initial transition strategy because SA series with yearly totals forced to match those of the raw series are available from table `D 11.A` in X-11-ARIMA, X-12-ARIMA version 0.2 and with an updated methodology in X-12-ARIMA version 0.3. Discussions with experts from Statistics Canada's System of National Account (SNA) justified the need to review the methodology and associated software to benchmark raw or SA data to annual totals from an external source. As with the other modules build in the SA prototype, applying benchmarking in a standardized and generalized way greatly simplifies maintenance issues. Benchmarking procedures can be applied on SA data, say to force annual totals to match those of the raw data (a requirement in a complex set of series like those of the SNA and in the two-step raking approach described above) or it can be applied before SA to correct possible coverage issues. The Quarterly Service Industry (QSI) project illustrates the latter case. Quarterly data is obtained from a survey for a few known complex and large contributors. Data for the remainder of the population is obtained from calendarized (see **?**) administrative data. This last data source is deemed to give a good estimate of the quarter-to-quarter change but the quality of the overall level estimation is not as precise. Benchmarking to annual totals provided by the annual Unified Enterprises Survey (UES) is applied to adjust the level.

5 Teaching

Teaching material related to seasonal adjustment was revised to take into consideration methodological improvements available in X-12-ARIMA. It was also updated to ©PowerPoint presentations. Each chapter of the course was dry-run in front of TSRAC complete staff to verify content, to train new staff members and to introduce the new capabilities of X-12-ARIMA to experienced members. Pilot versions of the course were given in the fall of 2005, in both English and French, and yet again in English in the spring 2006 for an outside agency. The course content includes methodology description, technical syntax information, detailed examples and hands-on exercises. It can be adapted to match the expertise of the participants.

Training material has yet to be developed for both the raking and the benchmarking modules. In both cases, we plan to develop a 1-day course where the methodology will be introduced and the related software will be demonstrated.

6 Expert system

TSRAC expects a large turn around in staff, that has already started, as many experienced staff members are retiring. Reviewing our overall SA approach allows us to consolidate the knowledge base of our soon-to-be retired staff. Each individual has his own approach and intuitive method to select the best SA options, method developed over the years from experience. One of the goals of the TSRAC is to document these different approaches, highlight their similarities, and investigate their particularities. Extensive interviews with experienced staff members were conducted in 2005 to document the various methods. We investigated the overlapping steps and got to a few standardized approaches varying from the very detailed and meticulous to the very intuitive. These approaches will go through a simulation study to measure the different SA options they yield. After final reevaluation of the different methods, we plan to look at the possibility and ease of automation of these steps. We are interested in both a more efficient way of analysing time series as well as in increased accuracy. We would like to develop an expert system for the selection of X-12-ARIMA SA options, the basis of which would take into consideration what is already available with automatic options available in X-12-ARIMA as well as the specific experience of Statistics Canada's analysts.

Reviewing how we select SA options shed some light on the different diagnostics and quality measures *actually* used by the analysts. Many are available from the diagnostics introduced in X-11-ARIMA/88 such as the quality control statistics (**?**) and the new X-12-ARIMA diagnostics such as the revision history and sliding spans (**?**). We have already looked at the tools to summarize those statistics like the US Census Bureau's Review Sheet and Statistics Canada's Summary program. Our expert system for automatic selection of SA options would make use of such tools.

7 Guidelines

With the use of new tools to do seasonal adjustment, there is a need to review and redefine current guidelines for seasonal adjustment and trend-cycle estimation and to update Statistics Canada documentation on quality guidelines for seasonal adjustment (**?**). These guidelines are currently written for the X-11-ARIMA method. They need to be updated to reflect the use of X-12-ARIMA and guidelines have to be established for other processes such as raking and benchmarking.

8 Mandate

The transition to X-12-ARIMA creates a change in the mandate of the TSRAC as the Centre will not be responsible for software coding and maintenance in the future. These responsibilities are transferred to SDD, following the model of other generalized systems in Statistics Canada. System development is a science by itself and although some TSRAC staff members will always be required to have good programming skills, to write *test*

program for example, it will not be their primary assets and they will not write or maintain production system. Time series specialists can then focus on time series issues.

The prototype set-up gives clear responsibilities to all parties. The SDD involvement includes the development of SAS-based generalized components (modules in the prototype). TSRAC provides the methodology specifications and test programs, if available. SDD is also responsible for the development of a generalized processor for production jobs. This processor is inspired from Statistics Canada Edit and Imputation system called *Banff*. It provides the link between the different modules and is basically a driver using meta-data to satisfy the particularities of a specific survey while having to support only one computer system. SDD will also develop a graphical interface for some of the modules, to be used in training and to provide a user-friendly introduction to the software. Finally, SDD expertise could be required to create specific extraction and exportation programs. These programs are the link between the prototype and the survey database.

Responsibilities of Subject-Matter (SM) divisions are similar to what they were in the X-11-ARIMA context. They include identification of series of interest for potential SA, identification of aggregation rules, and some review of X-12-ARIMA specs (i.e. confirm identified outliers or types of calendar effect). SM divisions are also responsible for the extraction and exportation steps, but they can also use SDD services. Finally, SM divisions are responsible for the actual transition including testing the prototype in their own environment. The official transition is done in collaboration with TSRAC.

9 Deployment

The transition to X-12-ARIMA and the fact that X-11-ARIMA will not be supported in the future was presented to survey programs through high management. The first survey to actually switch-over to the new system is ITS, used to build the prototype. SA data done with X-12-ARIMA and the new time series prototype were first published in March 2006 for reference periods up to January 2006. Personalized presentations for other affected programs, including presentation of the prototype, are done as required. On these presentations, TSRAC presents a template that lists the implementation tasks for updating the time series processing. This list of tasks is a management tool used to plan the transition to the new system. It includes the 7 keys steps for a successful transition, with detailed tasks related to implementation. It also delegates the tasks amongst the various groups of employees (TSRAC, SM division, SDD) and states tentative completions dates.

The 7 steps are:

1. Planning
2. Initial analysis work (including selection of SA options and identification of aggregation constraints)
3. System set-up (related to the prototype itself)

4. Development (optional step, only if the prototype does not meet all the needs of the project)
5. Testing
6. Maintenance (including plans for annual review of SA options)
7. Documentation

TSRAC established a deployment schedules of the new system for all the survey programs that directly use its services. The final decision to switch over to the new system belongs to the survey program management. However, the fact that the X-11-ARIMA system will not be supported after 2007 and the need for many surveys to update their computing environment from a mainframe environment to a PC one goes a long way to favour a timely transition. Other advantages of the transition include the opportunity to review the needs of SA. Some survey programs will use the transition to increase the number of series they produce by including new break-downs in their analysis. The timing of the actual switch-over for each program is selected to minimize disturbance in the continuous publication. It is either done in parallel with a complete system or survey redesign or during the regular process of SA historical revision. If needed, results obtained under X-11-ARIMA can be very closely reproduced under X-12-ARIMA and the new time series prototype. Finally, plans for new SA systems have been done directly in X-12-ARIMA since 2005.

10 Summary

This paper gave an overview of how Statistics Canada is building its new standard for time series processing. Our approach is to use standard modules with approved and up-to-date methodologies to reduce maintenance cost and improve the overall quality. This particular setting gives a highlight of how some of our historical and recent research projects integrate as well as an outline of our future developments.

References

Bell, W., and M. Kramer (1999): "Toward Variances for X-11 Seasonal Adjustments," *Survey Methodology*, 25, 13–29.

Chen, Z.-G., and P. Durk (2005): "A Therapy for Ill-seasonality in X-11 Seasonal Adjustment," Working Paper BSMD-2005-009E.

Chhab Alperin, N., and B. Quenneville (2004): "Changing to X-12-ARIMA at Statistics Canada," *Proceedings of Business and Economic Statistics Section of the American Statistical Association*, CD-ROM.

Dagum, E. B. (1981): *The X-11-ARIMA Seasonal Adjustment Method*. Statistics Canada, Ottawa, Canada, Catalogue 12-564E.

——— (1988): *The X-11-ARIMA/88 Seasonal Adjustment Method*. Statistics Canada, Ottawa, Canada, Methodology Branch Working Paper.

——— (1996): "A New Method to Reduce Unwanted Ripples and Revisions on Trend-Cycle Estimates from X-11-ARIMA," *Survey Methodology*, 22, 77–83.

Findley, D. F., B. C. Monsell, W. R. Bell, M. C. Otto, and B. C. Chen (1998): "New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program," *Journal of Business and Economic Statistics*, 16, 127–177.

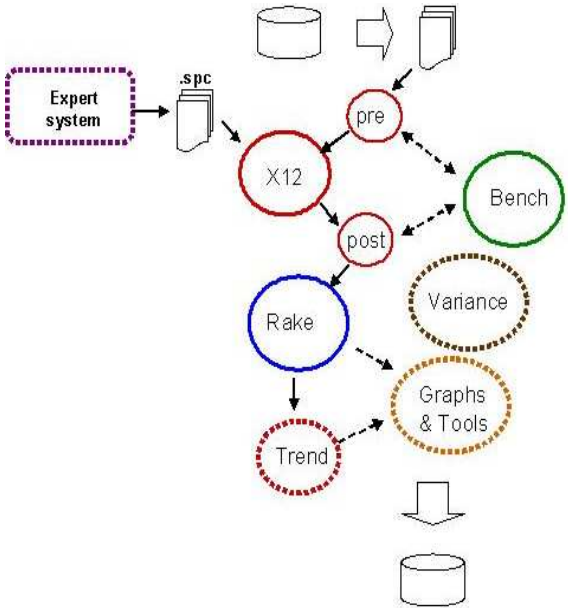

Figure 1: Sketch of the Time Series Prototype System and Expert System for selection of seasonal adjustment options.