Catalogue no. 11-522-x
ISSN: 1709-8211

Proceedings of Statistics Canada Symposium 2021
Adopting Data Science in Official Statistics to Meet Society's Emerging Needs

Statistics Canada's Seasonal Adjustment Dashboard

by François Verret

Release date: October 15, 2021

Statistics Canada Statistique Canada

Canada

Proceedings of Statistics Canada Symposium 2021
Adopting Data Science in Official Statistics to Meet Society's Emerging Needs

Statistics Canada's Seasonal Adjustment Dashboard

François Verret[1]

Abstract

Seasonal adjustment of time series at Statistics Canada is performed using the X-12-ARIMA method. For most statistical programs performing seasonal adjustment, subject matter experts (SMEs) are responsible for managing the program and for verification, analysis and dissemination of the data, while methodologists from the Time Series Research and Analysis Center (TSRAC) are responsible for developing and maintaining the seasonal adjustment process and for providing support on seasonal adjustment to SMEs. A visual summary report called the seasonal adjustment dashboard has been developed in R Shiny by the TSRAC to build capacity to interpret seasonally adjusted data and to reduce the resources needed to support seasonal adjustment. It is currently being made available internally to assist SMEs to interpret and explain seasonally adjusted results. The summary report includes graphs of the series across time, as well as summaries of individual seasonal and calendar effects and patterns. Additionally, key seasonal adjustment diagnostics are presented and the net effect of seasonal adjustment is decomposed into its various components. This paper gives a visual representation of the seasonal adjustment process, while demonstrating the dashboard and its interactive functionality.

Key Words: Time Series; X-12-ARIMA; Summary Report; R Shiny.

1. Introduction

A variety of statistical programs of Statistics Canada collect monthly or quarterly data to form time series (a sequence of observations collected at regular time intervals). Published time series cover a wide range of subjects, which include consumer prices, sales (e.g. retail and wholesale), employment, gross domestic product and international trade. The focus of analysis of those sub-annual time series is usually the movement from one period to the next, and in particular the latest observed period to period movement. Those time series more often than not show repeating patterns that can be attributed to calendar and seasonal effects (see green curve of Figure 1-1 which presents the last five years of data of a monthly time series). Those effects repeat, are predictable and are thus not particularly interesting analytically. Consequently, they are removed from the series and the result is the seasonally adjusted series (red curve of Figure 1-1). The seasonally adjusted series can in turn be decomposed into the trend-cycle (black curve of Figure 1-1) and the irregular components (difference between the red and black curves).

Figure 1-1
Last five years of data of an unadjusted series (green) and of the corresponding seasonally adjusted series (red) and trend-cycle (black)

[1]François Verret, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1Y 0T6, francois.verret@statcan.gc.ca

The trend-cycle component can be viewed as the average of the statistical process, while the irregular component can be viewed as the residual. A time series is most often decomposed either additively ($Y = S + C + I$, where Y is the original time series, S combines the calendar and seasonal effects, C is the trend-cycle component and I is the irregular components) or multiplicatively ($Y = S \times C \times I$). Seasonal adjustment (SA) at Statistics Canada is performed using the X-12-ARIMA method (Findley et al., 1998) according to published quality guidelines (Statistics Canada, 2009). For each survey or statistical program, every reference period SA is applied in SAS using a Time Series Processing System (TSPS; Ferland & Fortier, 2009). The TSPS makes use of the X12 SAS procedure to perform SA. It generates a detailed X-12-ARIMA "listing" for each series (i.e. a several pages long text file describing step by step and in detail the result of applying the X-12-ARIMA method to the series), as well as a multitude of SAS files that can be used to diagnose SA.

Matrix management is used at Statistics Canada to optimize organizational effectiveness (Statistics Canada, 2016). This type of management is used for the SA process in particular. Subject matter experts (SMEs) are responsible for the statistical program/project. They coordinate the activities of the various groups involved (collection, information technology, statistical methods, dissemination), execute the processing of the different steps and verify, analyse and disseminate the statistical outputs. Methodologists of the Time Series Research and Analysis Center (TSRAC) are responsible for developing and maintaining the processing environment for SA. The TSRAC updates the SA options regularly, performs periodic review of diagnostics and provides ongoing support for analysis and interpretation of results. To help the overall SA process an interactive dashboard was created by the TSRAC. Section 2 of this paper outlines the objectives and general approach of the dashboard, Section 3 presents the content and main features of the dashboard, while explaining the SA process for a particular monthly series, and a conclusion is given in Section 4 along with a discussion of the next steps of development.

2. Seasonal adjustment dashboard - Objectives and Approach

The first objective of the dashboard is to build capacity to help better interpret seasonally adjusted data. It was designed:

- to help SMEs understand the concepts and avoid them treating the SA process as a black box;
- as a training tool for SA courses;
- as a tool for briefing senior management.

The second objective is to increase the independence of SMEs and the efficiency of the process, by reducing SMEs reliance on TSRAC. The dashboard is an automated tool, intended to respond to most common questions. Efficiency gains are also expected by avoiding the effort to gather and summarize the relevant information from TSPS listings and SAS outputs.

The dashboard consists of one automated page by series. That page is designed to summarize the SA process in a way that is as intuitive as possible, with supporting numeric details. While being intuitive, over-simplification and approximations are avoided. The dashboard intends to respond to the most common questions for a specific series. Those questions typically take the following form:

- Is the seasonal adjustment process working as it should?
- Is the seasonal pattern changing?
- Why is the effect of seasonal adjustment different this year than last year?
- Do we see an effect of <*insert event here*> in our estimation?

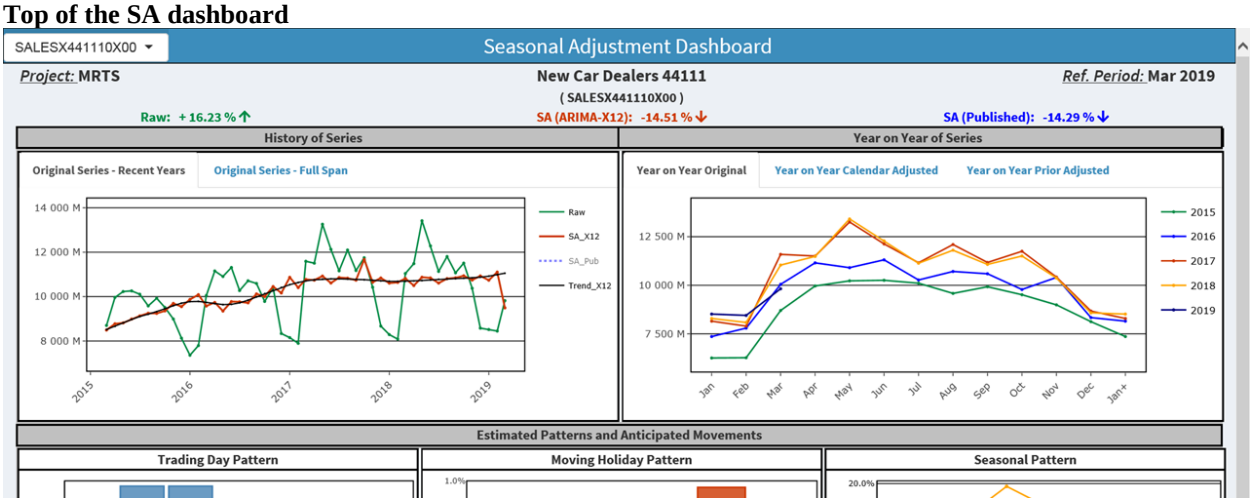
The programming platform used to produce the SA dashboard has evolved a lot over the years of development. The first version of the dashboard was an Excel spreadsheet where the input data coming from the TSPS were copy-pasted manually. Visual basic macros were added later on to help in reformatting the input data. A version of the dashboard was then programmed in Power BI, but customizing, content locking and interactivity issues were encountered. Similar issues were foreseen when evaluating the use of SAS-JMP. The final dashboard product was programmed using the R Shiny platform because of its many advantages. For one, the inputs are directly drawn from the TSPS SAS outputs. Furthermore, the output of the R Shiny dashboard is an html page, which means no special software

knowledge is required by the user for using the dashboard. Finally, the R Shiny platform offers many interesting interactive features, such as hovering, zooming, drop-down menus and the use of tabulations to overlay similar graphics.

3. Seasonal adjustment dashboard - Content

The dashboard content and functionalities will be demonstrated in this Section using the New Car Dealers sales series of Monthly Retail Trade Survey (MRTS), for the reference month of March 2019 (note that the series' value in March 2019 does not correspond to the disseminated value to show some features of the dashboard). Figure 3-1 presents the top of the dashboard page for that series. One can choose which MRTS series' page to view from the top-left menu. Below that menu information about the statistical program, the series and the reference period of interest is given. A summary of the February 2019 to March 2019 movement in the raw/unadjusted series is also given in the header (in green text), along with the movement of the seasonally adjusted series resulting from the application of the X-12-ARIMA method (in red text), as well as the seasonally adjusted movement that will be disseminated (in blue text). There might be differences between the latter two movements when a reconciliation/raking process is applied to the various seasonally adjusted series of the statistical program (Quenneville & Fortier, 2011). Raking is done so that component series sum to an aggregate series when they are seasonally adjusted independently.

Figure 3-1
Top of the SA dashboard



The first line of graphics of the dashboard presents the "History of the Series" panel on the left and the "Year on Year of Series" on the right. The graphic on the left is the one presented in Figure 1-1. By default the "Original Series – Recent Years" tabulation is selected, but one can choose to see the full span of the series by selecting the "Original Series – Full Span" tabulation. The year on year graphic presented on the right corresponds to the green curve of the left graph plotted on top of itself year by year. The year on year graphics are produced to make the sub-annual patterns more evident. By default the "Year on Year Original" tabulation is selected. The other two tabulations are shown on Figure 3-3 and will be discussed in the following paragraphs.

The second line of graphics (see Figure 3-2, which presents the middle of the dashboard page) displays each of the calendar and seasonal effects that need to be removed from the original series to obtain the seasonally adjusted series. To help understand the calendar effects, calendar of the reference period (March 2019) and of the previous reference period (February 2019) are presented on the next line. The first calendar effect presented is the trading day effect. It is the effect produced by the composition of the month in terms of each of the seven days of the week. The first half of X-12-ARIMA is the fitting of a reg-ARIMA model to the series. The vertical bar chart indicates that the model estimated that fewer sales are made on Mondays and Sundays and more sales are made on Tuesdays and Wednesdays. February 2019 is composed of an equal amount of each of the seven days of the week, which one could assume should lead to a 0 trading day effect. However, the trading day effect also accounts for the varying length of February due to leap years and is slightly negative (-0.88%) because February 2019 has 28 days, which is less than the 28 ¼ average

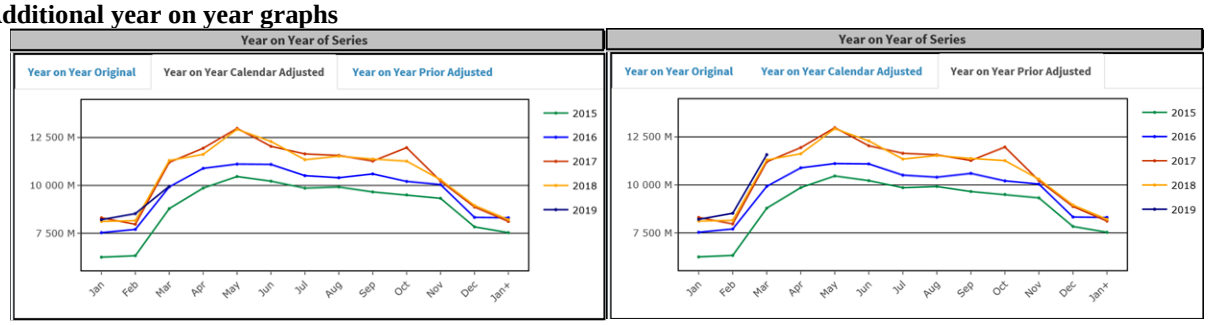
length for a month of February. March 2019 has an extra Friday, Saturday and Sunday, as shown in pale blue on the second calendar and on the vertical bar chart, resulting in an expected decrease of 2.25% in sales. So globally we are expecting a decrease in the unadjusted series from February to March 2019 due to the trading day effect.

The second calendar effect is the moving holiday effect. New car dealers are typically closed on Easter and since Easter can either fall in March or April, the month having Easter is expected to have less sales than usual and vice-versa. In 2019 Easter was on April 21. We therefore expect more sales than usual in March, almost a percent more according to the reg-ARIMA model estimate. Globally we thus expect the unadjusted series to go up by 0.89% due to the moving holiday effect. We can see the result of removing those two effects by comparing the Year on Year Calendar Adjusted graph (second tabulation of the top-right panel, see first graph of Figure 3-3) with the Year on Year Original graph (Figure 3-1, top-right panel). The calendar adjusted graph shows a more regular pattern because the "noise" attributed to trading day and moving holiday effects has been removed from the series.

Figure 3-2
Middle of the SA dashboard

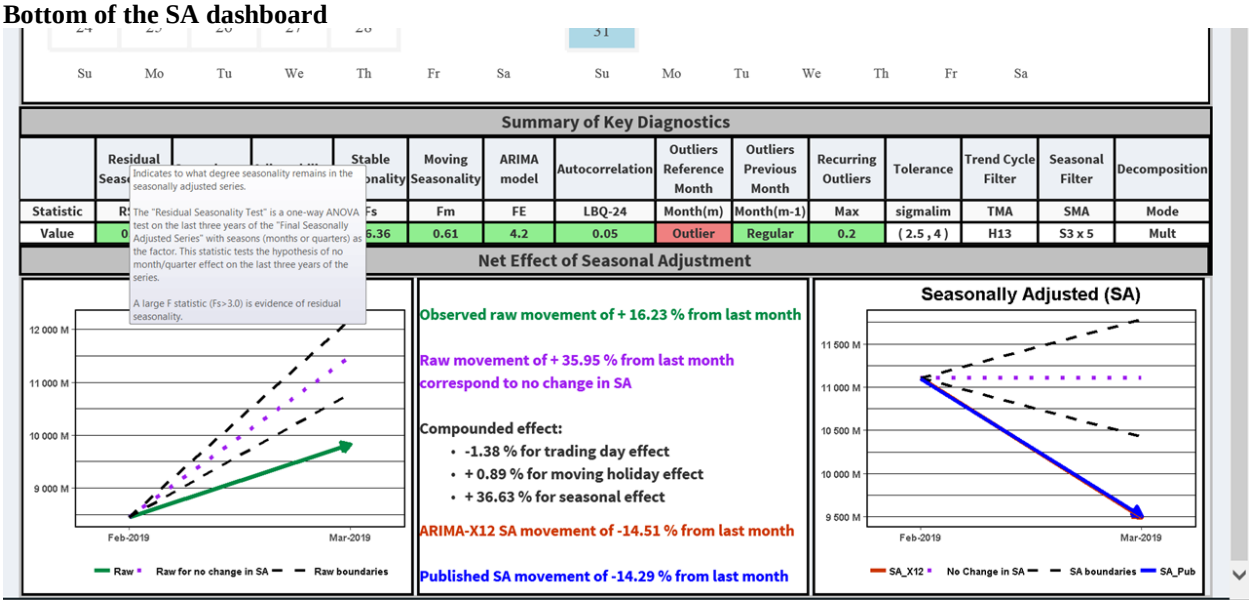
The reg-ARIMA model is also used by X-12-ARIMA to identify outlier values of the irregular component. This corresponds usually to special events occurring, a strike for example. The month of interest's value in fact corresponds to a very negative outlier. This is pretty clear when looking at either the History of the Series graph (left panel of Figure 3-1) or the Year on Year Original graph (first graph of Figure 3-3). On the former, the March 2019 seasonally adjusted data point is far below the trend-cycle datapoint. On the latter, the 2019 February to March movement is noticeably weaker than in previous years. Correcting for that outlier results in the Year on Year Prior Adjusted graph, which is the second graph presented in Figure 3-3. The pattern is made even more regular by that and we can more clearly see the seasonal pattern emerging.

Figure 3-3
Additional year on year graphs



The second half of X-12-ARIMA consists of applying the X11 method which, given the components estimated by the reg-ARIMA model, isolates the seasonal, the trend-cycle and the irregular components. The seasonal component is presented on the last graph of the second row of graphs of the dashboard (top-right graph of Figure 3-2). The seasonal component is allowed to evolve over time. The purple bars show the estimated pattern for the months observed in the current year and the yellow curve shows the pattern of the previous year for comparison. We clearly see in this example that we expect a very strong upward movement from February to March 2019 in the unadjusted series because of the seasonal component. As for the trend-cycle and irregular components, no panel are specifically dedicated to them. However, as was mentioned above, the estimate of the former is presented in black on the History of Series graph and that of the latter can be deducted from the difference between the seasonally adjusted and trend-cycle curves.

Figure 3-4
Bottom of the SA dashboard



On the second to last row of the dashboard (see Figure 3-4), we find a summary of SA diagnostics and of the options used. The color green is used for SA diagnostics that are good, the color yellow for diagnostics that are acceptable and the color red for diagnostics requiring attention. Hovering over a given diagnostic reveals an explanation of that diagnostic. On Figure 3-4, the mouse cursor was hovered over the Residual Seasonality diagnostic. For this particular example, we also see in red that the value of the irregular component for the month of interest is an outlier. The previous month was not an outlier as indicated by the green color. On top of the outlier detection process of the reg-ARIMA model (first half of the X-12-ARIMA method), an extreme value detection process is comprised in the X11 method (second half of the X-12-ARIMA method). Those extreme values are also flagged for the month of interest and the previous month if they occur.

On the last row of the dashboard we find a summary of the net effect on seasonal adjustment. On the left we have in green the observed February to March 2019 movement in the original series with the movement that would result in a flat SA movement based on the estimated calendar and seasonal effects in purple, with (approximate) confidence intervals around it in black. The middle panel provides numeric details. We expect the movement to be an increase of 35.95% and to be driven by the seasonal effect. However, we only observed a 16.23% increase and this results in a negative movement in the SA series (-14.51%). This is represented on the last panel in red (the blue arrow corresponds to the raked movement), with the flat SA movement in purple and corresponding confidence interval in black. Having an SA movement in the opposite direction of the original movement, as is the case here, is a typical situation TSRAC has to look into for SMEs. The dashboard makes it very clear why this is happening and why this is not a problem.

4. Conclusion and next steps

Once shared with SMEs, the SA dashboard clearly and quickly met its capacity building and efficiency goals. It proved to be an invaluable production tool, especially for dealing with pandemic data. It generated very positive reactions from SMEs during production, from time series training attendees and from senior management. It was initially produced for four key monthly economic surveys: the Monthly Retail Trade Survey, the Monthly Survey of Manufacturing, the Monthly Wholesale Trade Survey and the Monthly Survey of Food Services and Drinking Places. All four surveys share a lot of similarities. In particular, data is reported for the whole month (or are calendarized to correspond to whole months) and the reg-ARIMA models used for their flow series typically comprise a trading day regressor and a moving holiday regressor. The dashboard was next extended to the Labor Force Survey (LFS) and the Employment Insurance Statistics (EIS) program. Data for those two statistical programs are reported for a reference week within the month. Consequently, a reference week effect is present in the reg-ARIMA model instead of a trading day effect (Statistics Canada, 2020). The LFS and EIS SA dashboards therefore include a reference week effect panel instead of the trading day panel. Plans are also being made to extend the dashboard to the Survey of Employment, Payroll and Hours and to the Consumer Price Index program. This will be followed by the gradual rollout to other statistical programs (including quarterly programs).

Ongoing content development is also planned. With each new statistical program covered by the SA dashboard comes the incorporation of survey specific features (e.g. additional calendar effects). Moreover, graphic outputs could always be refined and R Shiny does offer many options to do so. A supplementary tabulation will be added to the year on year graphs panel to show the effect of correcting the series for extreme values (see October 2017 value in this paper's example). The addition of a second page for each series might be considered for advanced users. More importantly, adding a page for series that are aggregates of the series presented might prove to be very useful. Summarizing the various SA diagnostics for aggregate of series is not straightforward, but using the outputs of the composite specification of X-12-ARIMA (U.S. Census Bureau, 2017) would be a good starting point.

Acknowledgements

Many past and present members of the TSRAC have contributed to the seasonal adjustment dashboard. The author would thus like to thank Steve Matthews, Michel Ferland, Nada Habli, Joanne Leung, Yu Wang, Lorcan Mischler, Joana Bérubé and Marouane Seffal for their dedication and many contributions to the tool.

References

Ferland, M. and S. Fortier (2009), "Recent Developments in Statistics Canada's Time Series Processing System: Transition to SAS® PROC X12", *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, pp. 1921-1932.

Findley, D., Monsell, B., Bell, W., Otto, M., and B. Chen (1998), "New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program", *Journal of Business & Economic Statistics*, 16(2), pp. 127-152.

Quenneville, B. and S. Fortier (2011), "Restoring Accounting Constraints in Time Series – Methods and Software for a Statistical Agency", in W. R. Bell et al. (eds.) *Economic Time Series: Modeling and Seasonality*, Chapman and Hall/CRC, pp. 231-254.

Statistics Canada (2009), "Seasonal Adjustment and Trend-Cycle Estimation", *Statistics Canada Quality Guideline*, 5th edition, Catalogue no. 12-539-X, available at https://www150.statcan.gc.ca/n1/pub/12-539-x/2009001/seasonal-saisonnal-eng.htm

Statistics Canada (2016), "Chapter 2.1: Organizational Structure and Matrix Management", *Compendium of Management Practices for Statistical Organizations from Statistics Canada's International Statistical*

Fellowship Program, Catalogue no. 11-634-X, available at https://www150.statcan.gc.ca/n1/pub/11-634-x/2016001/section2/chap1-eng.htm

Statistics Canada (2020), "Section 6: Data Processing", *Guide to the Labour Force Survey*, Catalogue no. 71-543-G, available at https://www150.statcan.gc.ca/n1/pub/71-543-g/71-543-g2020001-eng.htm

U.S. Census Bureau (2017), X-13ARIMA-SEATS Reference Manual, Version 1.1, Washington, DC, available at https://www.census.gov/ts/x13as/docX13AS.pdf