Statistics Canada / Statistique Canada

Methodology Branch

Business Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquêtes auprès des entreprises

Canada

WORKING PAPER NO. BSMD-2005-009E
METHODOLOGY BRANCH

CAHIER DE TRAVAIL NO. DMEE-2005-009E
DIRECTION DE LA MÉTHODOLOGIE

A therapy for ill-seasonality in X-11 seasonal adjustment

by

Zhao-Guo Chen and Paul Durk

BSMD-2005-009E

A therapy for ill-seasonality in X-11 seasonal adjustment

Zhao-Guo Chen and Paul Durk

Time Series Research and Analysis Centre
Business Survey Methods Division
Statistics Canada

April 2005

Abstract

When a series for a total, say, the monthly number of travelers from the U.S. to a Canadian province, is broken down into sub-series, say, the number of American travelers by car, plane and other modes of transportation, sometimes the seasonal aspect of one or more sub-series may be the dominating component of these series, and the values in some or all troughs are close to 0. For such sub-series, neither the additive nor the multiplicative adjustment in X-11 related programs can give reasonable results. This paper proposes a treatment of the raw data and the resulting adjusted series so that the final seasonally adjusted series presents the relatively correct trend and irregularity of the original series. The procedure is simple and the method is robust. The insight of such a treatment is explained. An example is provided to illustrate the procedure.

Résumé

Quand une série pour un total, tel que le nombre de voyageurs en provenance des États-Unis vers une province Canadienne, est ventilée par moyen de transport (automobile, avion, autres), la composante saisonnière peut être la composante principale d'une ou de plusieurs sous séries et les valeurs de certains ou de tous les creux peuvent être égales ou proches de 0. Pour de telles sous séries, les modes de décomposition additif et multiplicatif de la méthode X-11 (incluant ses dérivées) ne donnent pas des résultats raisonnables. Cet article propose un traitement des données brutes et de la série ajustée résultante de telle sorte que la série désaisonnalisée finale représente correctement la tendance et les irréguliers de la série originale. La procédure est simple et la méthode et robuste. L'idée derrière le traitement est expliquée. Un exemple illustre la procédure.

1. Introduction

Seasonal adjustment is a widely used methodology within Statistics Canada and around the world to assist in analyzing time series data. By taking out the seasonal pattern in a series we can more easily and accurately do a month-to-month (or quarter-to-quarter) comparison or other analysis because the trend or the seasonally adjusted series (a combination of the trend and the irregular components) may better reflect the real economic situation of the subject under analysis. However, if a given series for the total is broken down into numerous sub-series, sometimes the seasonal aspect of one or more sub-series may become more influential than the trend. We may say these sub-series have "ill-seasonality". For such series the seasonal adjustment may lose its meaning because the corresponding "seasonally adjusted series" will be very sensitive to the specification of the adjustment and the results from the specification may be misleading to analysts. Normally, we would not seasonally adjust these series since they could not be adjusted well enough to be publishable.

On the other hand, if such a situation happens and the seasonally adjusted (SA) series must be reconciled (that is, rake the SA series so that the sum of the SA sub-series equals to the SA series for the total), then all the sub-series have to be seasonally adjusted. However, when the sub-series with "ill-seasonality" are seasonally adjusted, not only do we obtain poor results but also these results may have a noticeably adverse impact on other SA sub-series (and the SA total in more general situation) in the reconciliation procedure. For those sub-series with "ill-seasonality", it is necessary to catch the "real" trend and irregularity in a way that the corresponding SA sub-series do not affect significantly other SA series in the reconciliation procedure.

Specifically, this issue came to the attention of Time Series Research and Analysis Centre (TSRAC) during the migration of the system of seasonal adjustment for the International Travel Survey (ITS) from X-11-ARIMA to X-12-ARIMA. New sub-series were created at fairly low aggregated levels. These sub-series refer to provincial travel totals broken down by mode of transportation (car, plane, etc) and trip duration (same-day, and overnight). It was with some of these sub-series that difficulties in seasonal adjustment arose.

These sub-series, which are said to have "ill-seasonality", all have values equal to or close to 0 in the low season months (troughs) and high values in the peak season months.. The issue with these series is that standard options for seasonal adjustment, either multiplicative or additive, can not provide decent results.

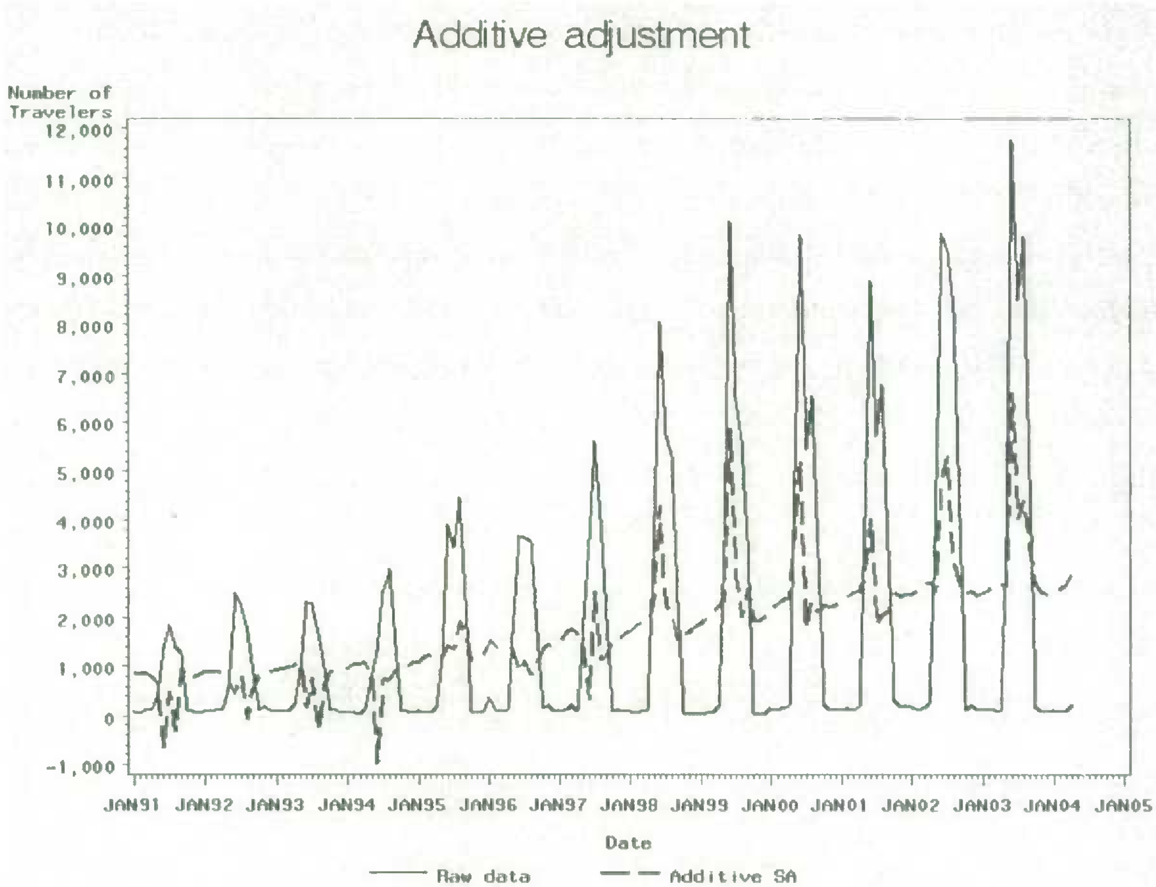

Figure 1

The solid line in Figure 1 is the original series (monthly raw data) for travelers from the United States to Yukon, Canada. The transportation mode is "non-auto", and the trip duration is "same-day" (a provincial sub-series by mode and duration). The dashed line is the seasonally adjusted series using X12-ARIMA with the additive option. During the trough season (winter), most values of the original series are very close to 0. That causes the problem in the adjusted series in higher-value season (not during the winter): many values become negative at the early years where the peaks of the original series are low;

large peaks and dips systematically exist for the later years (1997-2003) where the peaks of the original series are high. Certainly, such a seasonally adjusted series is not acceptable.

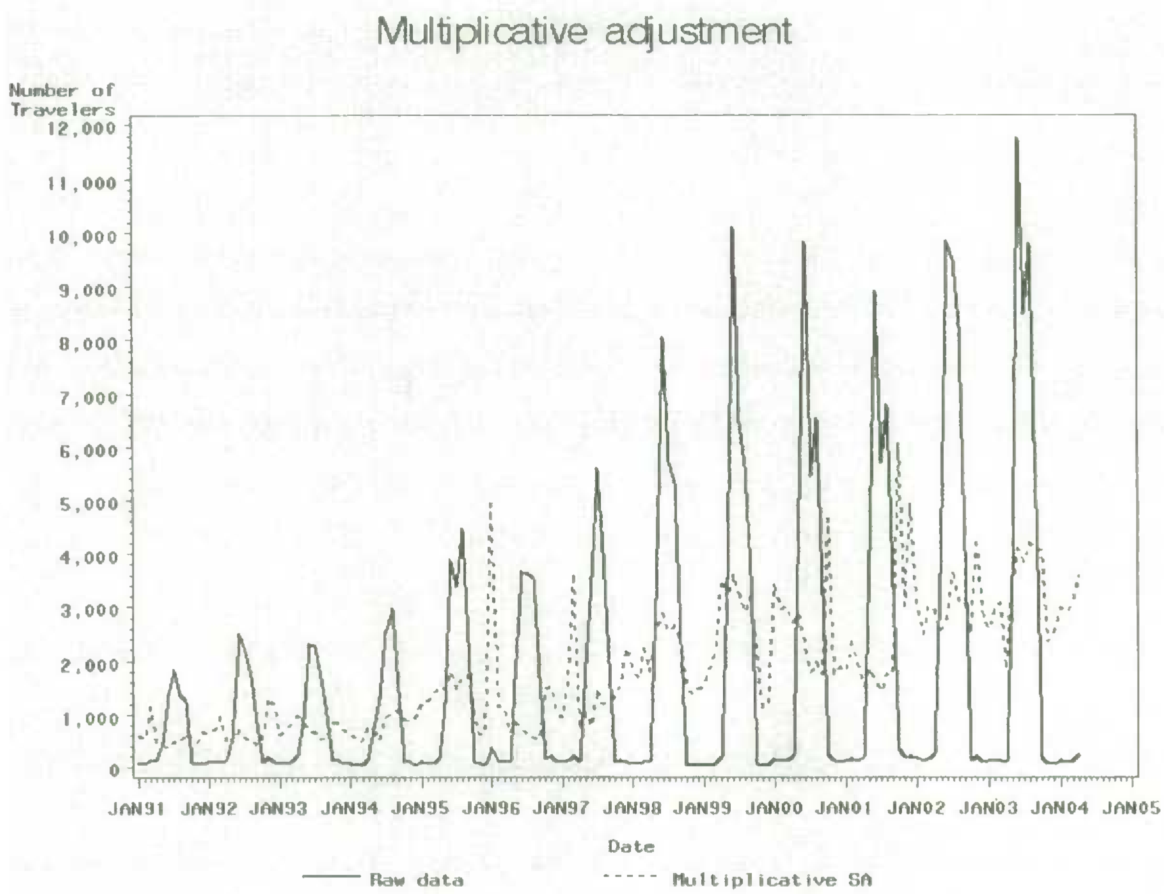

Figure 2

In general, when the raw series appears to have seasonal patterns with amplitude proportional to its level, the multiplicative adjustment should be used. This series seems to fit this situation. However, a multiplicative adjustment is not suitable to a series that contains values close to 0 since it could create spurious excessive spikes or dips during the trough season (see Figure 2, dotted line). Moreover, if there are 0 values, then the multiplicative adjustment simply cannot be done. One such example is the number of travelers from Nova Scotia, Canada, to the United States with "car" as the transportation mode and "overnight" as the duration. The series shows 0 values for most months during the winter season since 1998.

In addition to the additive and the multiplicative adjustments, X-12-ARIMA also provides users with two other options: the log-additive and the pseudo-additive adjustments.

The log-additive adjustment is similar to the multiplicative adjustment. For the example in Figure 2, the SA series obtained from using the log-additive adjustment is much smoother than the SA series shown in Figure 2; however, the spurious high peaks during the trough season still exist (though not very high). For this example, the log-additive adjustment perhaps is acceptable, but like the multiplicative adjustment, it cannot be done when the series has 0 values.

Compared with the additive adjustment in Figure 1, the pseudo-additive adjustment results in a SA series with smaller peaks and dips during the higher-value season of later years (1998-2003), but still with negative values in earlier years (1991, 1994) and in 1997 while the additive adjustment did not result in negative values in 1997.

2. A therapy

How do we resolve the above mentioned issue of seasonally adjusting a series that is clearly seasonal but neither the additive nor the multiplicative options can provide reasonable results? The solution arrived at is as follows: add a positive constant value to each data point in the original series prior to adjustment, run a multiplicative adjustment, and then subtract that constant value from each data point in the adjusted series to get the final results. The issue here is how to choose this constant value so that it has no or very little effect on the outcome.

The observed data can be decomposed as seasonal, trend cycle, and irregular components as

$$O(t) = S(t) + T(t) + I(t)\,.$$

Now we put *O(t)* in the multiplicative expression. In the following, *(t)* may be dropped for simplicity, e.g. $S = S(t)$. Let

$$\widetilde{S} = S/T, \qquad S^* = 1 + \widetilde{S} = (S+T)/T,$$

$$\widetilde{I} = I/(S+T), \qquad I^* = 1 + \widetilde{I} = (S+T+I)/(S+T),$$

then it is easy to see (the multiplicative expression)

$$O = S^* T I^* = S^* N_0^*,$$

where $N_0^* = TI^*$ is the seasonally adjusted component (more precisely, should be called "factor" which is not the same as $N=T+I$). Here the subscript 0 in N_0^* corresponds to $C=0$ in the following.

Now we add a non-negative constant C to all data points, i.e.

$$O_c(t) = O(t) + C = S(t) + T_c(t) + I(t),$$

where $T_c(t) = T(t) + C$. Consider $O_c(t)$ as the original series and put it in the multiplicative expression, again drop (t) from $S(t)$ etc, similarly we have

$$\widetilde{S}_c = S/T_c, \quad S_c^* = 1 + \widetilde{S}_c = (S+T_c)/T_c$$

$$\widetilde{I}_c = I/(S+T+C), \quad I_c^* = 1 + \widetilde{I}_c = (S+T_c+I)/(S+T_c),$$

and

$$O_c = S + T_c + I = S_c^* T_c I_c^* = S_c^* N_c^*,$$

where $N_c^* = T_c I_c^*$. After applying X11-ARIMA or X12-ARIMA to O_c with the multiplicative option, we obtain N_c^*; and then we let the final seasonally adjusted series be

$$N^* = N_c^* - C.$$

Obviously, N^* obtained this way depends on C. We now investigate how C affects the results. Certainly, $N_c^* = N_0^*$ (the result by directly using the multiplicative option on the original raw data), if C=0. When $C \downarrow 0$, the seasonally adjusted series N_c^* will have the feature similar to those of N_0^*: during the trough season, it is very unstable. This can be seen as follows. Put

$$N_c^* = T_c I_c^* = T_c + T_c \widetilde{I}_c$$

The first term $T_c = T + C$ simply raises the level of the series T by C. According to the definitions of T_c and $\widetilde{T}_c$, the second term can be written as $[(T+C)/(S+T+C)]I$. During the trough season, $S+T \approx 0$; when $C \downarrow 0$, the ratio (the coefficient of I) can be very large, and this term is very unstable. So, C should not be too small.

Next, when $C \uparrow \infty$, with either the multiplicative or the additive option for O_c, five basic steps are repeatedly used (c.f. Ladiray and Quenneville, 2001, p16-18).

(A) $O_c \Rightarrow T_c$, using 2×12 moving average to get the trend.

(B) Either $O_c / T_c = S_c^* I_c^*$ (multiplicative) or $O_c - T_c = S + I$ (additive).

(C) $S_c^* I_c^* \Rightarrow S_c^*$ (multiplicative) or $S + I \Rightarrow S$ (additive), using 3×3, or 3×5, or 3×7 moving average to get the seasonality.

(D) Either $O_c / S_c^* = N_c^*$ (multiplicative) or $O_c - S = N_c$ (additive).

(E) Either $N_c^* = T_c I_c^* \Rightarrow T_c$ (multiplicative) or $N_c = T_c + I \Rightarrow T_c$ (additive), using Henderson moving average to get the trend.

Note that, if the additive option is chosen, S and I do not depend on C, so there is no subscript c. Step (D) produces the seasonal adjusted factor or component. As we now use the multiplicative option, the first operation (division) should be applied. According to the definitions given above, we have

$$N_c^* = O_c / S_c^* = O_c (T+C)/(S+T+C) = O_c[\ (S+T+C) - S\ ]/(S+T+C)$$

$$= O_c - O_c S\ /(S+T+C)\ =\ O_c - S\ (S+T+C+I)/(S+T+C).$$

When $C \uparrow \infty$, the coefficient of S converges to 1, then the result of the division becomes the second expression in (D), i.e. the operation for additive adjustment. So if C is too large, the final result will be close to that of directly using the additive adjustment for O.

In summary, C should not be too small and too large. Then what is the guideline for choosing C? Our experience showed that in a large range for C, the result is not affected much by the different values C may take. C may be the overall average of the whole series, or half or a quarter of this average, or sometimes it can be larger than this average. Or, we may take C as the average of the original series over the period where the original series has lower values. Accordingly, we may try different values for C to find such a range where the results are very similar and satisfactory. Then we may choose any value in this range for C. Here a "satisfactory" seasonally adjusted series means that there is no negative value; it is relatively smooth; there is no large "blow-up" of local peaks in the trough season; there are no systematic peaks (or troughs) in several continuous years in the peak season.

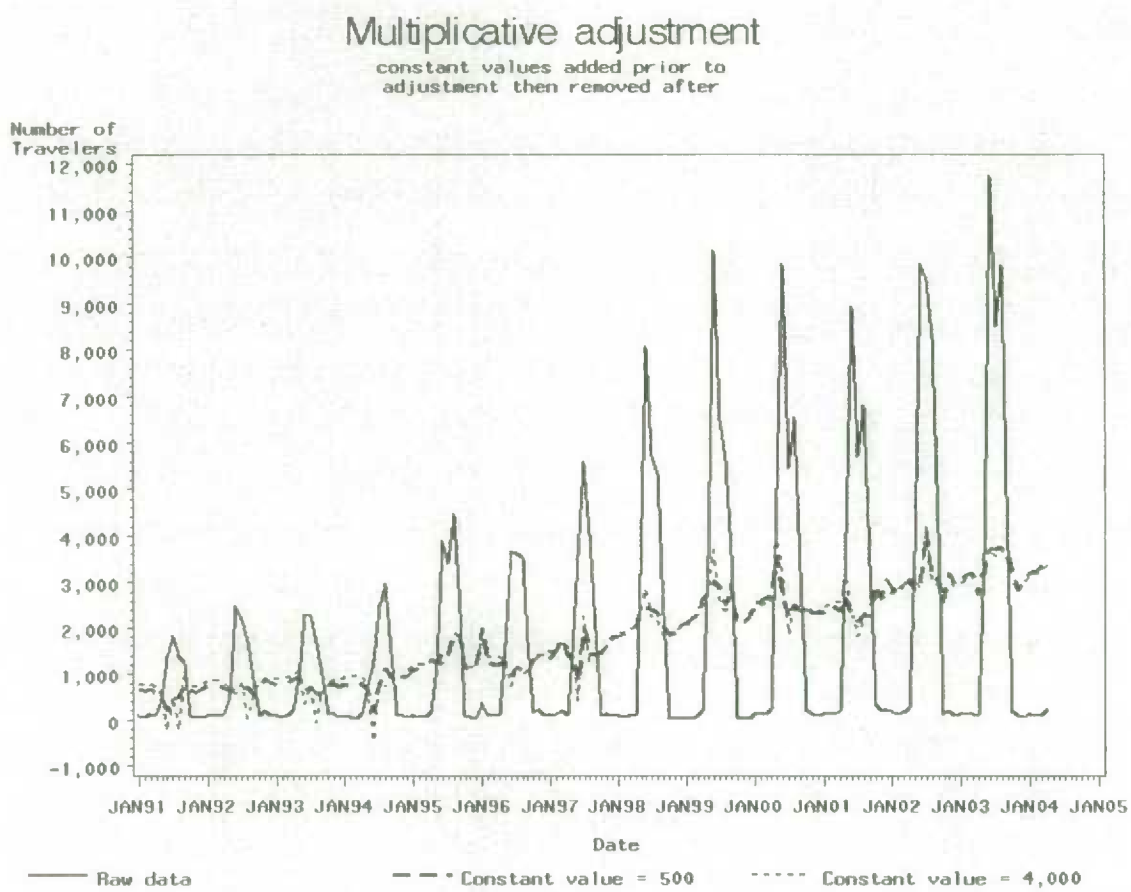

Figure 3

This can be illustrated by working on the abovementioned example, where the overall average of the raw data is about 1600. The original series (solid line) and the final seasonally adjusted series (dashed and dotted lines) are shown in Figure 3. C=500 (dashed) and C=4000 (dotted) give relatively close results in the later part of the series (1997+) where the raw data has a monthly average of approximately 2400. But C=4000 is not as good as C=500 for this period since the corresponding SA series is less smooth. Furthermore, C=4000 results with negative values in the early period of the final SA series (1991-1996) where the raw data has an average of less than 900. Thus, C=4000, which is much larger than the average is not suggested. If we take C=900, the corresponding final SA series is almost the same as the one obtained by using C=500. Thus any C value between 500 and 900 can be used.

3. Conclusion and application

The described method may extract satisfactory trend cycle and irregularity from original series with the abovementioned ill-seasonality. The result is not sensitive to the choice of C in a wide range as described in the previous section. Ultimately, this method was successfully utilized on 28 different series for the International Travel Survey. Due to the development of this method, we were able to seasonally adjust most provincial sub-series by mode/duration which are highly seasonal and with troughs approaching 0, or taking 0 values during the trough season. That made it possible and easier to run the reconciliation program (raking the provincial sub-series to the provincial total). The effects on the outcomes of the other "normal" sub-series and of the total series are minimized.

Acknowledgement

The authors thank Jocelyn Lapierre and Benoit Quenneville for their valuable comments and help with editing.

References

Ladraiy, D. and Quenneville, B. (2001). Seasonal Adjustment with the X-11 Method, Lecture Notes in Statistics, **158**, Springer New York, Berlin.