

# Simple methods to restore the additivity of a system of time series



EUROPEAN  
COMMISSION



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

**New freephone number:**

**00 800 6 7 8 9 10 11**

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.int>).

Luxembourg: Office for Official Publications of the European Communities, 2005

ISSN 1725-4825

ISBN 92-79-01313-0

© European Communities, 2005

# SIMPLE METHODS TO RESTORE THE ADDITIVITY OF A SYSTEM OF TIME SERIES

BY BENOÎT QUENNEVILLE AND ERIC RANCOURT<sup>1</sup>

*Statistics Canada, Time Series Research and Analysis Centre*

We propose and illustrate simple methods based on constrained weighted least squares regression to restore the additivity of a system of time series, with the objective of balancing a table of seasonally adjusted series benchmarked to the corresponding annual totals from the raw series. The methods are designed to be easily programmable with the ©SAS statistical software.

KEYWORDS: Balancing, Benchmarking, Raking, Reconciliation, Regression, SAS, Seasonal adjustment.

JEL CLASSIFICATION:

## 1 Introduction

At Statistics Canada, many surveys publish seasonally adjusted (SA) series that must satisfy some aggregations constraints. For example, the Monthly Retail Trade Survey (MRTS) publishes SA series by industry (19 trade groups) and by region (10 provinces and 3 territories). The SA national total is obtained indirectly as the sum of the 19 industries; however, the sum of the 13 regions does not necessarily equal this national total. In this paper we present simple methods to correct the 13 regional series to eliminate this discrepancy. We call this process *reconciliation* or *balancing*.

Reconciliation is related to the seasonal adjustment of aggregates, often referred to as the direct and indirect approaches to seasonal adjustment<sup>2</sup>. For example, it is possible with X-11-ARIMA, Dagum (1981, 1988), and X-12-ARIMA, Findley, Monsell, Bell, Otto, and Chen (1998), to seasonally adjust the MRTS national total indirectly from the industry breakdown. However, when looking at the regional breakdown, the national total becomes fixed and the sum of the regional estimates is not necessarily equal to this national total; unfortunately, nothing is available in X-11-ARIMA or X-12-ARIMA to impose this national total on the regional estimates.

An alternative to reconciliation is to present the monthly discrepancies openly; however, showing explicit discrepancies usually causes confusion among users and criticism or embarrassment to the publishers.

One simple solution to the reconciliation problem is to seasonally adjust at the industry-region level, and to obtain all the marginal totals indirectly; however, the industry-region

---

<sup>1</sup>Benoît Quenneville, Business Survey Methods Division, 120 Parkdale Ave, Ottawa, Ontario, K1A 0T6, Canada, tel.: +1-613-951-1605, e-mail: Benoit.Quenneville@statcan.ca. Eric Rancourt, tel.: +1-613-951-5046, email: Eric.Rancourt@statcan.ca.

<sup>2</sup>A much detailed discussion on this topic can be found in Bloem, Dippelsman, and Mæhel (2001).

level is too detailed and the impact of the irregular components at that level is too dominant for proper seasonal adjustment.

The MRTS uses the indirect approach to obtain the national total from the industry breakdown, because it is easier to identify breaks, outliers, calendar effects, the seasonal effect narrowly defined, and so on at the industry level. Since the regional series need to be corrected to add up to this national total, reconciliation is performed after seasonal adjustment to restore the accounting relationships that exist between the series.

The simplest way to reconcile the 13 regional MRTS SA series is by pro-rating; however, for the MRTS, as well as many Statistics Canada surveys, the SA series are corrected to match the annual totals from the corresponding raw series; hence, reconciliation must preserve that constraint.

We use a two-step procedure to solve this problem. In the first step, we correct the individual SA series so that their annual totals match those of the corresponding raw series using the benchmarking procedure described in Quenneville, Cholette, Huot, and Di Fonzo (2004) or using Table D11A of X-11-ARIMA or X-12-ARIMA. This first step fixes the level of the individual SA series and is designed to preserve their month to month growth rates<sup>3</sup>. This step also brings the benefits of annual reconciliation and simplifies the rest of the procedure. In the second step, we reconcile all the series so that the aggregation constraints are satisfied, making sure that the annual totals constraints from the first step remain satisfied. This second step is implemented via a constrained weighted regression. The advantage of this two-step approach are:

1. there is no need to preserve the month to month growth rates in the second step because this is done in the first step, and hence,
2. the overall regression in the second step is divided into smaller regressions that only consider either one month at a time, or a year at a time when an annual constraint is involved; also,
3. it is easy to program the regression with the ©SAS statistical software.

For all practical purposes, the reconciliation<sup>4</sup> methodology implemented in this two-step approach gives similar numerical results as those obtained with the general method proposed by Dagum and Cholette (To appear)<sup>5</sup>. Here, however, the two-step approach splits the computations between the seasonal adjustment program (Table D11A of X-12-ARIMA) and a simple regression on only the years and months that need to be reconciled for publication purposes.

---

<sup>3</sup>Unfortunately, consistency with the annual totals from the raw series is achieved at the expense of the quality of the seasonal adjustment and is conceptually wrong: for series with significant calendar-related effects or moving seasonality effects, the annual totals of a seasonally adjusted series should differ from the unadjusted series. This is because moving seasonality implies that the impact of the seasonal effect vary from year to year; and, the number of working days, the impact of moving holidays, and other calendar-related effects vary from year to year.

<sup>4</sup>Like with imposing the annual totals, one can argue that reconciliation is achieved at the expense of damaging the time series quality of the regional (individual) components.

<sup>5</sup>To be consulted for a good overview on how this problem has been dealt with in the past, and for an extensive list of references on the subject.

In this paper, we focus on the proposed reconciliation methodology and illustrate it with various examples. Section 2 illustrates the ideas behind the regression model. We show how pro-rating can be obtained from a weighted regression. We illustrate the concept of *alterability coefficients* as a mean of modifying binding constraints. We show how to balance a  $2 \times 2$  table with appropriate regression models. Section 3 presents case studies. The first case is the reconciliation of historical survey estimates that were converted to a new classification system. The second application is for a prototype we have developed for the seasonal adjustment of the International Travel Survey series with X-12-ARIMA, where reconciliation is needed because the direct SA of an aggregate is imposed on its components. The third application is the reconciliation of the SA regional MRTS series with the SA national total described in this introduction. The last application is with a 2-way classified table of SA series.

## 2 Simple Regression Models for Reconciliation

In some situations, reconciliation is often achieved by simple pro-rating; so, we begin by showing how pro-rating can be obtained from a weighted regression. Suppose there are 3 observations  $y_0, y_1, y_2$  such that  $y_1$  and  $y_2$  must add up to  $y_0$ . One way of reconciling the observed values of  $y_1$  and  $y_2$  with  $y_0$  is by simple pro-rating where the corrected value of  $y_1$  is set equal to  $b_1 = y_1 y_0 / (y_1 + y_2)$ , the corrected value of  $y_2$  is set equal to  $b_2 = y_2 y_0 / (y_1 + y_2)$ , and so  $b_1 + b_2 = y_0$ .

It is easy to set up a regression model to perform the numerical computations for pro-rating. Let  $y_1 = b_1 + e_1$ ,  $y_2 = b_2 + e_2$ ,  $y_0 = b_1 + b_2$ ,  $e_1 \sim (0, y_1)$ ,  $e_2 \sim (0, y_2)$  where  $e_i \sim (0, y_i)$  means that the error  $e_i$  has mean 0 and variance  $y_i$ . Parameter  $b_2$  can be eliminated to obtain the simplified model:  $y_1 = b_1 + e_1$ ,  $y_0 - y_2 = b_1 + e_2$ . Assuming  $e_1$  and  $e_2$  are uncorrelated, the best linear unbiased estimate of  $b_1$  is a weighted average of  $y_1$  and  $y_0 - y_2$  where the weights are inversely proportional to the variances:

$$b_1 = \left( \frac{1}{y_1} + \frac{1}{y_2} \right)^{-1} \left( \frac{y_1}{y_1} + \frac{y_0 - y_2}{y_2} \right) = y_1 \frac{y_0}{y_1 + y_2}.$$

Hence, this regression model is just a mechanical device to perform the numerical computations for pro-rating. This is illustrated with the ©SAS code for the case just presented where  $y_0 = 40$ ,  $y_1 = 5$  and  $y_2 = 25$ .

```
data y;
  input y x1 x2;
  weight = 1/y;
  cards;
    5 1 0
    25 0 1
  ;
run;
Proc reg data=y;
  model y = x1 x2/noint p;
```

```

weight weight;
restrict x1 + x2 = 40;
output out = ypred(keep = y predicted)
       p = predicted;
run;
quit;
Proc Print data=ypred;run;

```

The results are:

y	predicted
5	6.6667
25	33.3333

As expected, the predicted values of  $y_1$  and  $y_2$  add to 40, and are exactly the same as those obtained with pro-rating.

We now extend this simple regression model to deal with more complex cases.

The general model of Dagum and Cholette (To appear) uses *alterability coefficients* to artificially modify the variance of the error associated with an observation, or to modify a binding constraint. In this case, let  $e_1 \sim (0, a_1 y_1)$ ,  $e_2 \sim (0, a_2 y_2)$ ,  $y_0 = b_1 + b_2 + e_0$ ,  $e_0 \sim (0, a_0 y_0)$ , where  $a = (a_0, a_1, a_2)$  is a known vector of alterability coefficients. In this case, it is easy to show that

$$\begin{aligned}
 b_1 &= y_1 + \frac{a_1 y_1}{a_0 y_0 + a_1 y_1 + a_2 y_2} [y_0 - (y_1 + y_2)], \\
 b_2 &= y_2 + \frac{a_2 y_2}{a_0 y_0 + a_1 y_1 + a_2 y_2} [y_0 - (y_1 + y_2)],
 \end{aligned}$$

but now

$$b_1 + b_2 = y_0 - \frac{a_0 y_0}{a_0 y_0 + a_1 y_1 + a_2 y_2} [y_0 - (y_1 + y_2)].$$

The use of an alterability coefficient for  $y_0$  permits not to completely satisfy the constraint. Setting  $a_0$  to a small number is a numerical trick to get close to a binding constraint; moreover, the regression model can be greatly simplified by treating  $y_0$  as an observation with a very small error instead of a constraint on the parameters.

As with the previous example, the ©SAS code for this case is:

```

data y;
input alter y x1 x2;
weight = 1/(alter* y);
cards;
.1    40 1 1
1     5 1 0
1     25 0 1
;
run;

```

```

Proc reg data=y;
  model y = x1 x2/noint p;
  weight weight;
  output out = ypred(keep = alter y predicted)
         p = predicted;
  run;
  quit;
proc print data=ypred;run;

```

And the results are:

alter	y	predicted
0.1	40	38.8235
1.0	5	6.4706
1.0	25	32.3529

When  $a_0 = 0.00001$  the results are:

alter	y	predicted
0.00001	40	39.9999
1.00000	5	6.6666
1.00000	25	33.3332

When dealing with a one-way classification, it may be easier to apply pro-rating without going to the trouble of writing a regression model to perform the arithmetic; however, for higher dimensional tables of time series, it will be much easier to set up a regression model using the above ideas and to use a regression package to perform the computations. This will be illustrated with the case studies, but for now, we show how the regression approach simplifies the computations when dealing with a  $2 \times 2$  table of values.

Suppose the data values are provided in a two-way classification as in the following example:

$y_{1,1}$	$y_{1,2}$	$y_{1,\cdot}$
$y_{2,1}$	$y_{2,2}$	$y_{2,\cdot}$
$y_{\cdot,1}$	$y_{\cdot,2}$	$y_{\cdot,\cdot}$

where  $y_{i,j}$ ,  $i, j = 1, 2$  are the observed values for the cross-classification  $i, j = 1, 2$ ,  $y_{i,\cdot}$ ,  $i = 1, 2$  are the row totals,  $y_{\cdot,j}$ ,  $j = 1, 2$  are the column totals, and  $y_{\cdot,\cdot}$  is the grand total of the table. Balancing this  $2 \times 2$  table of values is obtained with the following

regression model:

$$\begin{aligned}
y_{1,1} &= b_{1,1} + e_{1,1}, e_{1,1} \sim (0, a_{1,1}y_{1,1}) \\
y_{1,2} &= b_{1,2} + e_{1,2}, e_{1,2} \sim (0, a_{1,2}y_{1,2}) \\
y_{2,1} &= b_{2,1} + e_{2,1}, e_{2,1} \sim (0, a_{2,1}y_{2,1}) \\
y_{2,2} &= b_{2,2} + e_{2,2}, e_{2,2} \sim (0, a_{2,2}y_{1,2}) \\
y_{1,.} &= b_{1,1} + b_{1,2} + e_{1,.}, e_{1,.} \sim (0, a_{1,.}y_{1,.}) \\
y_{2,.} &= b_{2,1} + b_{2,2} + e_{2,.}, e_{2,.} \sim (0, a_{2,.}y_{2,.}) \\
y_{.,1} &= b_{1,1} + b_{2,1} + e_{.,1}, e_{.,1} \sim (0, a_{.,1}y_{.,1}) \\
y_{.,2} &= b_{1,2} + b_{2,2} + e_{.,2}, e_{.,2} \sim (0, a_{.,2}y_{.,2}) \\
y_{.,.} &= b_{1,1} + b_{1,2} + b_{2,1} + b_{2,2} + e_{.,.}, e_{.,.} \sim (0, a_{.,.}y_{.,.}).
\end{aligned} \tag{1}$$

In the above set-up, there are 9 observations, 9 known coefficients of alterability ( $a_{i,j}$ ,  $i, j = ., 1, 2$ ) and 4 parameters ( $b_{i,j}$ ,  $i, j = 1, 2$ ) to estimate. As long as  $a_{i,j} > 0$ ,  $i, j = ., 1, 2$ , an unrestricted regression can be run and the predicted values are the balanced or reconciled estimates.

If one of the alterability coefficients is equal to zero, then the corresponding equation should be written as a constraint. For illustration, consider balancing a  $2 \times 2$  table subject to fixed marginal totals by means of a weighted regression.

Equation (1) becomes

$$\begin{aligned}
y_{1,1} &= b_{1,1} + e_{1,1}, e_{1,1} \sim (0, a_{1,1}y_{1,1}) \\
y_{1,2} &= b_{1,2} + e_{1,2}, e_{1,2} \sim (0, a_{1,2}y_{1,2}) \\
y_{2,1} &= b_{2,1} + e_{2,1}, e_{2,1} \sim (0, a_{2,1}y_{2,1}) \\
y_{2,2} &= b_{2,2} + e_{2,2}, e_{2,2} \sim (0, a_{2,2}y_{1,2}) \\
y_{1,.} &= b_{1,1} + b_{1,2} \\
y_{2,.} &= b_{2,1} + b_{2,2} \\
y_{.,1} &= b_{1,1} + b_{2,1} \\
y_{.,2} &= b_{1,2} + b_{2,2}
\end{aligned} \tag{2}$$

where the equation for  $y_{.,.}$  is not needed anymore because it is assumed that  $y_{.,.} = y_{1,.} + y_{2,.} = y_{.,1} + y_{.,2}$ . All the error terms  $e_{1,.}$ ,  $e_{2,.}$ ,  $e_{.,1}$  and  $e_{.,2}$  are set to zero because  $y_{1,.}$ ,  $y_{2,.}$ ,  $y_{.,1}$  and  $y_{.,2}$  represent the control totals.

Using the relations between the parameters and the control totals, we can rewrite Equation (2) as:

$$\begin{aligned}
y_{1,1} &= b_{1,1} + e_{1,1}, e_{1,1} \sim (0, a_{1,1}y_{1,1}) \\
y_{1,.} - y_{1,2} &= b_{1,1} + e_{1,2}, e_{1,2} \sim (0, a_{1,2}y_{1,2}) \\
y_{.,1} - y_{2,1} &= b_{1,1} + e_{2,1}, e_{2,1} \sim (0, a_{2,1}y_{2,1}) \\
y_{2,2} + y_{1,.} - y_{.,2} &= b_{1,1} + e_{2,2}, e_{2,2} \sim (0, a_{2,2}y_{1,2}).
\end{aligned}$$

The best linear unbiased estimator of  $b_{1,1}$ , is thus the weighted average of  $y_{1,1}$ ,  $y_{1.} - y_{1,2}$ ,  $y_{.,1} - y_{2,1}$  and  $y_{2,2} + y_{1.} - y_{.,2}$  where the weights are inversely proportional to their variances:

$$b_{1,1} = \frac{\left( \frac{y_{1,1}}{a_{1,1}y_{1,1}} + \frac{y_{1.} - y_{1,2}}{a_{1,2}y_{1,2}} + \frac{y_{.,1} - y_{2,1}}{a_{2,1}y_{2,1}} + \frac{y_{2,2} + y_{1.} - y_{.,2}}{a_{2,2}y_{2,2}} \right)}{\left( \frac{1}{a_{1,1}y_{1,1}} + \frac{1}{a_{1,2}y_{1,2}} + \frac{1}{a_{2,1}y_{2,1}} + \frac{1}{a_{2,2}y_{2,2}} \right)}.$$

The estimates for the remaining parameters follows from the constraints in Equation (2).

For an  $r \times c$  table, it will be easy to set up the regression model described by Equations (1) and/or (2) and to use a regression package to perform the computations. The difference between Equations (1) and (2) is only in how the constraints on the parameters are treated. Equation (2) assumes binding constraints, whereas Equation (1) assumes non-binding constraints; obviously, there could be a mixture where, for example, the row totals are non-binding and the columns totals are. This will be illustrated in the case studies.

## 3 Case Studies

### 3.1 Reconciliation of historical survey estimates that were converted to a new classification system

This first case study is an illustration of the balancing of a 2-way classified table where all the cells have non-zero alterability coefficients.

Fortier (2003) describes the first part of a process to convert historical survey estimates to a new classification for the Monthly Wholesale and Retail Trade Survey (MWRTS). The MWRTS was developed in the 1980's to produce sales and inventories estimates for the industrial sectors defined by the Standard Industrial Classification (SIC) system. The survey was redesigned to produce estimates for the industrial sectors under the North American Industry Classification System (NAICS) starting in 2004. It was also required to convert the existing series under the SIC classification to the new NAICS classification. For retail trade, backcasting needed to start in January 1991; for wholesale, backcasting started in January 1993; providing at least 10 years of monthly data under the new classification.

It was possible to produce NAICS-based domain estimates for the period January 1998 to December 2001 because the NAICS classification was available for the sampled units. This permitted to calculate conversion factors that were applied to the SIC-based estimates for the period January 1991 (January 1993 for wholesale) to December 1997 to produce NAICS-based estimates starting in 1991. Details are provided in Fortier (2003).

A parallel run of the old and redesigned surveys was available for the period December 2003 to April 2004, allowing the computation of linking factors from the old to the new surveys under NAICS. Without going into details, the linking factors were simply computed as the ratio of the NAICS-based estimates from both surveys. Retail trade series were linked using the March 2004 ratios, and wholesale trade, April 2004.

We now describe in more details how the linking and balancing was done for the retail trade sector using the January 2004 data (this exercise was repeated for every month in the parallel run).

For this discussion, we assume that the survey produces 280 monthly survey estimates decomposed as follows: the grand total; 13 (10 provinces and 3 territories) regional totals; 19 national industry totals; and  $247 = 19 \times 13$  cross-classified industry by region estimates. The parallel run thus provides 280 linking factors.

The linking factor for the grand total is 1.0126. The linking factors at the 13 regional levels vary between 0.94 to 1.22; at the 19 industry levels, between 0.70 to 1.21; at the 247 industry by region levels, between 0 to 13.30 with 0.15 the first non-zero factor and 2.25 the second largest.

To link the historical series it was decided to apply the 280 linking factors to the corresponding series from January 1991 to December 2004. This, however, destroys the additivity of the table for all the months prior to January 2004.

To balance back the tables of estimates, we have applied the method described in the previous section, independently to each of the 156 months between January 1991 to December 2003

The alterability coefficients were selected as follows: 0 for 3 empty cells; 1 for 125 cells including the grand total, the 19 industry totals, the 13 regional totals, most of the series in the 3 territories, and some selected industry-province combinations; 2 for 63 cells, 3 for 59 cells, 4 for 26 cells, and 5 for 4 cells. The same default alterability coefficients were applied to the 156 months.

The regression model corresponding to Equation (1) contains  $247 = 19 \times 13$  free parameters for the 247 industry by region combinations and has 280 observations. It is thus necessary to construct a design matrix, say  $X$ , of dimension  $280 \times 247$ , made of zero and one. Each row of  $X$  corresponds to one of the 280 survey estimates. Each column of  $X$  corresponds to one of the industry-region combination. Given a column  $j$ , row  $i$  takes the value 1 if the corresponding survey estimate includes that industry-region. For example, the column corresponding, say, to industry '010' and region '10' will take the default value of 0, except when the survey estimates is in one of the following categories:

1. industry '010' and region '10',
2. industry '010' and region '00' for the industry total at the national level,
3. industry '000' and region '10' for the regional all industry total, and
4. industry '000' and region '00' for the grand total.

This design matrix  $X$  depends only on the known industry-region structure; so, it is constructed only once.

The presence of zero values is handled by temporary setting them to a small value equal to 0.001 with a corresponding alterability coefficient reset to 0.0001, so that the corresponding weight  $w_{i,j} = 1/(a_{i,j}y_{i,j})$  can be calculated. A flag is also created to reset

the value back to zero. This permits to avoid the removal of the zero values from the computations, which would require a redefinition of the free parameters.

The predicted values from the regression computations correspond to the reconciled estimates. However, these numbers are not rounded. So, rounding is applied to the 247 free estimates, rendering temporary modified zero values back to zero using their flag. Finally, the 19 industry, 13 regional and grand totals are recomputed from the rounded figures. This provides a balanced, rounded and linked table of survey estimates for the month being processed.

## Reconciliation of SA series subject to the annual total constraint

The next case studies deal with the reconciliation of a set of SA series with the additional constraint that the annual totals in the SA series must equal those of the raw series.

We now suppose that the SA series satisfy the annual total constraint. We have discussed in the introduction how to impose that constraint on the individual series. In practice, this simply means getting the table of the SA series with constrained yearly totals from X-12-ARIMA. Next, we split the SA series by years. The complete years, for which an annual total is available, are reconciled one complete year at a time. The current year, where the annual total is not available until the end of the year, is reconciled one month at a time.

For complete years, the months (or quarter) are used as another classification to impose the annual totals on the reconciled estimates. The marginal totals over the months are the annual totals. There are two ways to impose the annual total constraint. One solution is to put the annual total constraints as constraints on the parameters. The other solution is to assign small alterability coefficients to the annual totals. We now illustrate the two approaches with the following two case studies.

### 3.2 Application to the International Travel Survey Series

Figure 1 shows six series from Statistics Canada's International Travel Survey. The series are the SA number of tourists entering the province of Saskatchewan from the United States by various mode of transportation and length of stay: AirOn stands for by plane (Air) staying overnight (On); CarOn, by car staying overnight; CarSd, by car and returning to the United States on the same day (Sd); NonSd, not by car and returning to the United States on the same day; OthOn, not by plane or car (Oth stands for Other) and staying overnight. The Provincial Total (Tot) series is seasonally adjusted directly and the five modes must add up to it. The suffix 'R' at the end of a mode stands for the series after reconciliation. In the figure, the mode NonSd is not seasonally adjusted; so, it is not, and must not be, modified by the reconciliation procedure.

Figure 2 shows the difference between the SA before and after reconciliation. For the total series, the difference is between the direct and indirect SA series, where the indirect SA is obtained by adding the five modes. Figure 3 shows the relative differences. Both

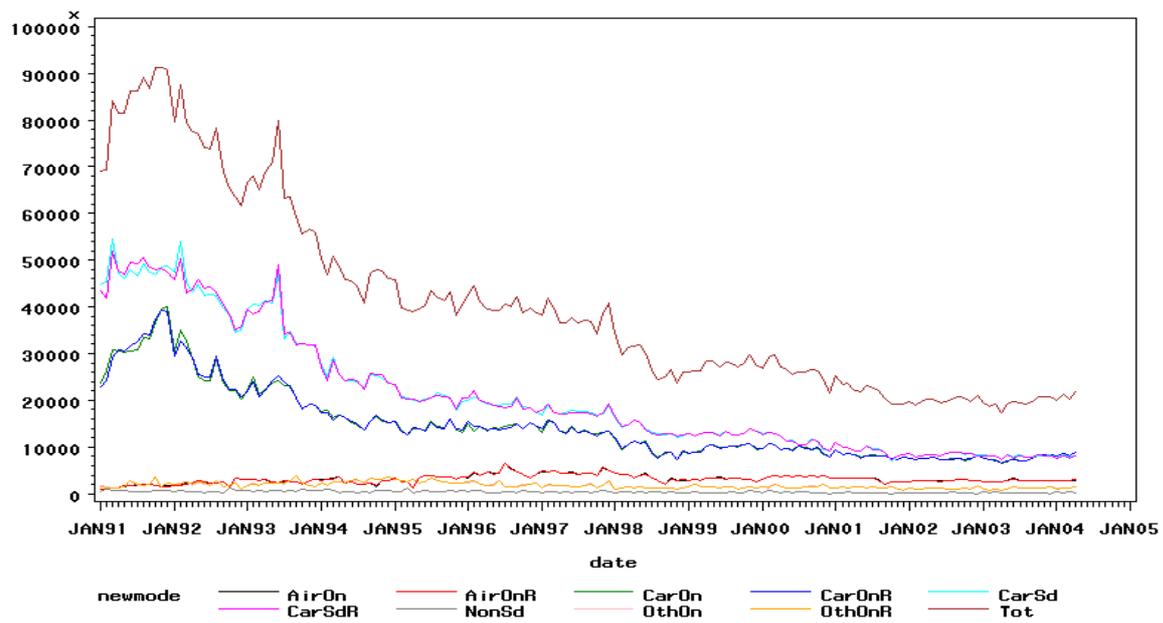


Figure 1: ITS series before and after reconciliation

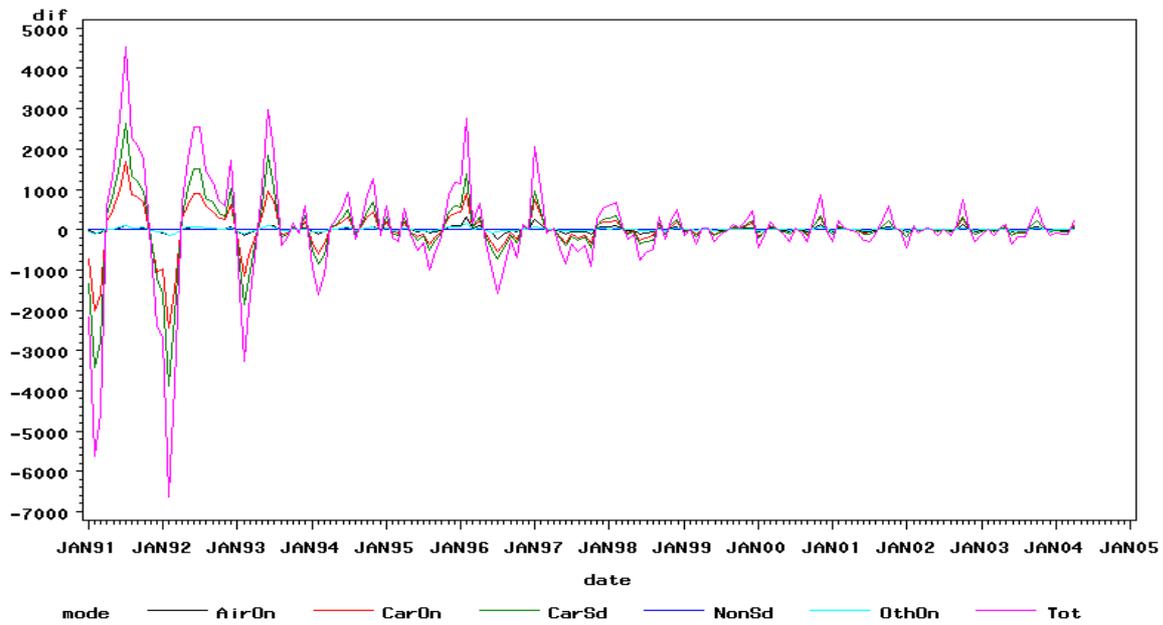


Figure 2: Differences in ITS series before and after reconciliation

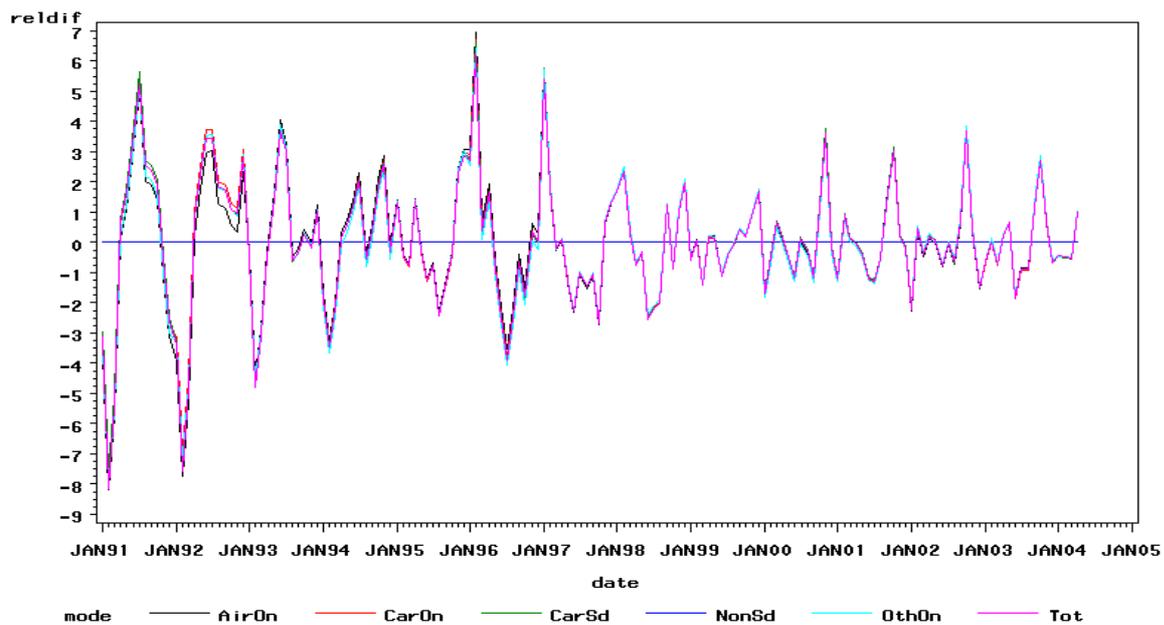


Figure 3: Relative differences in percentage in ITS series before and after reconciliation

figures show how the difference between the five modes and the direct SA total is allocated in proportion to their contribution to the total. Clearly, the figures show that there is residual seasonality in the discrepancy between the direct and indirect SA total. It should be obvious to a seasonal adjustment expert that this residual seasonality would have a multiplicative decomposition model with a lot of moving seasonality.

This example illustrates a case where the direct SA of the aggregate is most likely of better quality and where it is the component series that need to be reconciled, mostly, because some of those components contribute insignificantly to the total, but need to be published for others reasons.

The ©SAS code to reconcile one month of observations, when an annual total constraint is not yet available, is quite simple:

```
Proc reg data = saa;
  model saa = AirOn CarOn OthOn CarSd /noint P;
  weight weight;
  restrict AirOn + CarOn + OthOn + CarSd = &monthtot;
  output out = saa3
    (keep = Year Month Mode Alter positivevalue
      saa predicted)
  p = predicted;
run;
quit;
```

In this code, the dataset saa has 4 observations. Each observation has the variables saa (the SA series corrected to match the annual totals), the year (same for all observations),

the month (same month for the 4 observations), the mode (AirOn, CarOn, OthOn, CarSd) excluding NonSd because it is not seasonally adjusted, the alterability coefficient (Alter), the weight (inverse of alter \* saa), a flag indicating if the SA observation is equal to zero (positivevalue), and 4 explanatory variables taking the value 0 or 1. For example, the variable AirOn takes the value 1 if the mode is AirOn and 0 otherwise. The other explanatory variables in the model statement are defined similarly. The SAS macro variable &monthtot is used to assign the direct SA value of the provincial total (minus the NonSd value); so, the corresponding restriction states that the four reconciled modes must add to that value. The predicted values out of the regression computation are the reconciled values.

The ©SAS code to reconcile one year of observations is also quite simple:

```
Proc reg data = saa;
  model saa =
    AirOnm01 AirOnm02 AirOnm03 AirOnm04 AirOnm05 AirOnm06
    AirOnm07 AirOnm08 AirOnm09 AirOnm10 AirOnm11 AirOnm12
    CarOnm01 CarOnm02 CarOnm03 CarOnm04 CarOnm05 CarOnm06
    CarOnm07 CarOnm08 CarOnm09 CarOnm10 CarOnm11 CarOnm12
    OthOnm01 OthOnm02 OthOnm03 OthOnm04 OthOnm05 OthOnm06
    OthOnm07 OthOnm08 OthOnm09 OthOnm10 OthOnm11 OthOnm12
    CarSdm01 CarSdm02 CarSdm03 CarSdm04 CarSdm05 CarSdm06
    CarSdm07 CarSdm08 CarSdm09 CarSdm10 CarSdm11 CarSdm12
  /noint P;
  weight weight;
  restrict AirOnm01+AirOnm02+AirOnm03+AirOnm04+AirOnm05+AirOnm06+
    AirOnm07+AirOnm08+AirOnm09+AirOnm10+AirOnm11+AirOnm12
    = &airontot;
  restrict CarOnm01+CarOnm02+CarOnm03+CarOnm04+CarOnm05+CarOnm06+
    CarOnm07+CarOnm08+CarOnm09+CarOnm10+CarOnm11+CarOnm12
    = &carontot;
  restrict OthOnm01+OthOnm02+OthOnm03+OthOnm04+OthOnm05+OthOnm06+
    OthOnm07+OthOnm08+OthOnm09+OthOnm10+OthOnm11+OthOnm12
    = &othontot;
  restrict CarSdm01+CarSdm02+CarSdm03+CarSdm04+CarSdm05+CarSdm06+
    CarSdm07+CarSdm08+CarSdm09+CarSdm10+CarSdm11+CarSdm12
    = &carsdtot;

  restrict AirOnm01 + CarOnm01 + OthOnm01 + CarSdm01 = &m1tot;
  restrict AirOnm02 + CarOnm02 + OthOnm02 + CarSdm02 = &m2tot;
  restrict AirOnm03 + CarOnm03 + OthOnm03 + CarSdm03 = &m3tot;
  restrict AirOnm04 + CarOnm04 + OthOnm04 + CarSdm04 = &m4tot;
  restrict AirOnm05 + CarOnm05 + OthOnm05 + CarSdm05 = &m5tot;
  restrict AirOnm06 + CarOnm06 + OthOnm06 + CarSdm06 = &m6tot;
  restrict AirOnm07 + CarOnm07 + OthOnm07 + CarSdm07 = &m7tot;
  restrict AirOnm08 + CarOnm08 + OthOnm08 + CarSdm08 = &m8tot;
  restrict AirOnm09 + CarOnm09 + OthOnm09 + CarSdm09 = &m9tot;
  restrict AirOnm10 + CarOnm10 + OthOnm10 + CarSdm10 = &m10tot;
```

```

restrict AirOnm11 + CarOnm11 + OthOnm11 + CarSdm11 = &m11tot;
restrict AirOnm12 + CarOnm12 + OthOnm12 + CarSdm12 = &m12tot;
output out = saa3
  (keep = Year Month Mode Alter positivevalue
    saa predicted)
p = predicted;

```

The dataset `saa` contains 48 observations for the 12 months of data and the 4 modes to be reconciled. Each observation has the variables `saa` (the SA series corrected to match the annual totals), the year (same for all observations), the month (ranging from 1 to 12), the mode (`AirOn`, `CarOn`, `OthOn`, `CarSd`) excluding `NonSd` because it is not seasonally adjusted, the alterability coefficient (`Alter`), the weight (inverse of `alter * saa`), a flag indicating if the SA observation is equal to zero (`positivevalue`), and 48 explanatory variables taking the value 0 or 1. For example, the variable `AirOnm01` takes the value 1 if the mode is `AirOn` and the month is 1, similarly for the other explanatory variables in the model statement. The macro variable `&airontot` is used to assign the annual total for the series `AirOn`; so, the first restriction basically states that the 12 monthly reconciled values for `AirOn` must add to this annual total; similarly for the other macro variables `&carontot`, `&othontot` and `&carsdtot`. The macro variable `&m1tot` is used to assign the direct SA value of the provincial total (minus the `NonSd` value); so, the corresponding restriction states that the four reconciled modes must add to that value. The predicted values out of the regression computation are the reconciled values. They are further treated to deal with zero values and to be rounded, but details are omitted.

### 3.3 Application to the Monthly Retail Trade Survey

The reconciliation of the 13 regional SA series to the national total for the MRTS discussed throughout this paper can be done the same way as in the reconciliation of the International Travel Survey series, except the 13 regions replace the 5 mode-length of stay classification.

The regression model would then contains 156 explanatory variables for the 13 regions  $\times$  12 months combinations; 13 restrictions to preserve the annual totals for the 13 regional SA series; and 12 restrictions to impose the monthly Canada totals on the corresponding 13 regional estimates.

Another solution that we found useful in practice is to use small alterability coefficients for the annual totals instead of imposing them as binding constraints. This is because, it is often the case that the raw series themselves do not satisfy some basic accounting relationships. This is illustrated in Figure 4 that displays the monthly differences between the Canada totals obtained from summing the 19 industries and from summing the 13 regions. There are small annoying differences that will obviously impact on the annual total constraints; hence, for those series, it was decided to impose the Canada total obtained from the industries on the 13 regional series, and to treat the annual total constraints as observations with small alterability coefficients.

Figure 5 shows the SA series before and after reconciliation for the province of British

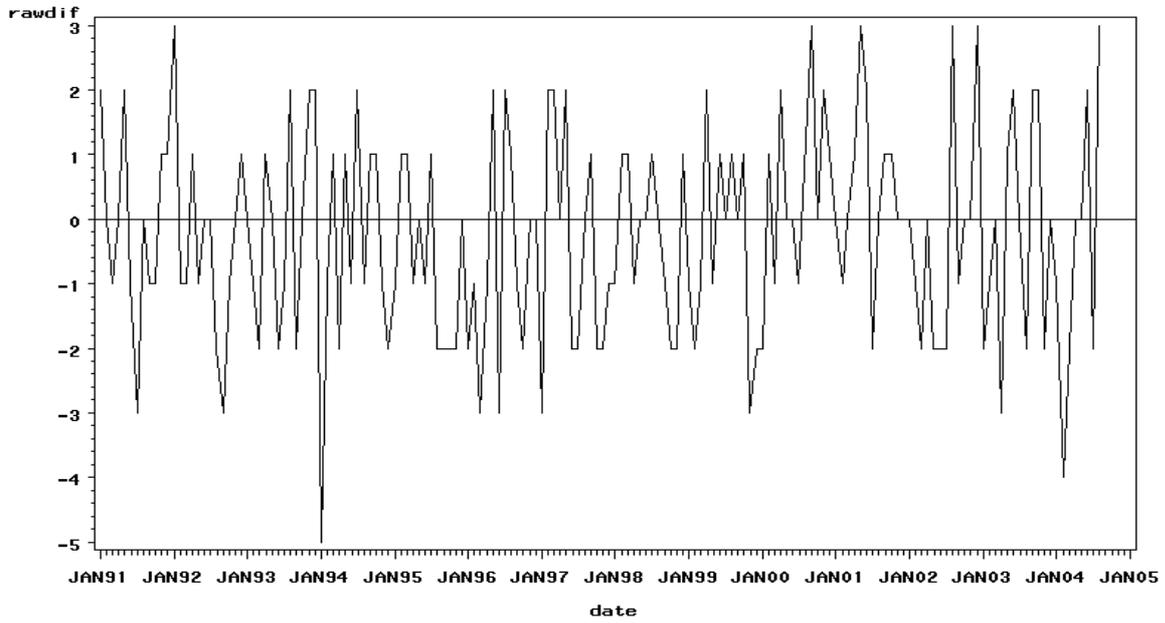


Figure 4: Monthly differences in the MRTS raw series between the Canada totals from the 19 industries and the 13 regions

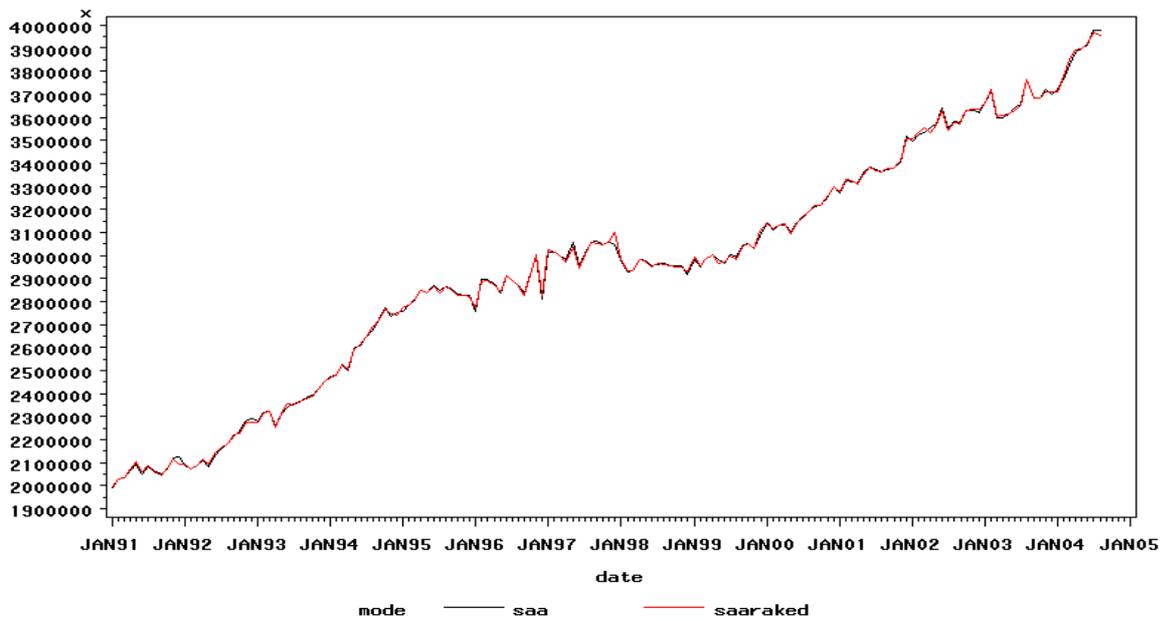


Figure 5: British Columbia MRTS SA series before and after reconciliation

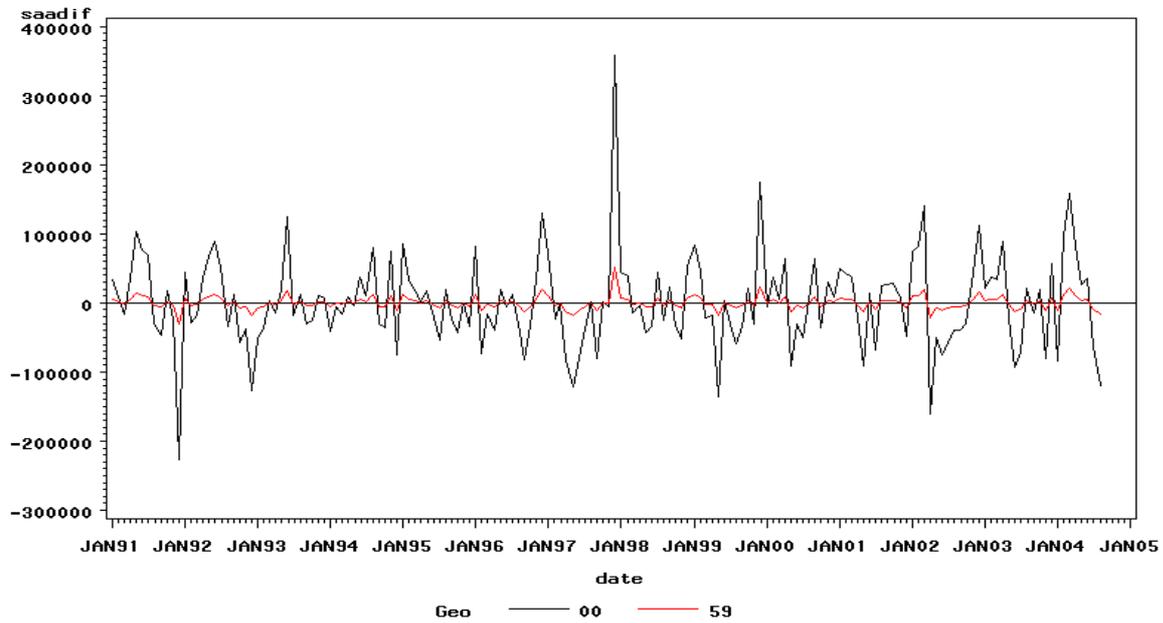


Figure 6: Difference between the National total and the total from the regions (00) for the MRTS series; and the corresponding share allocated British Columbia (59)

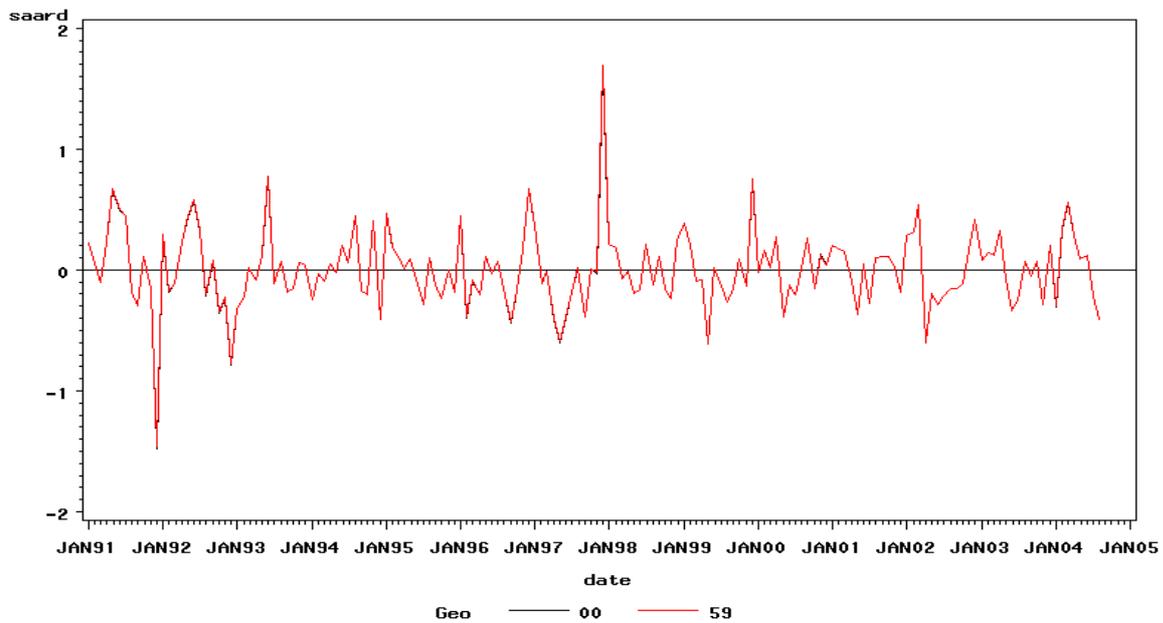


Figure 7: Relative differences in percentage in MRTS series before and after reconciliation for the National total (00) and British Columbia (59)

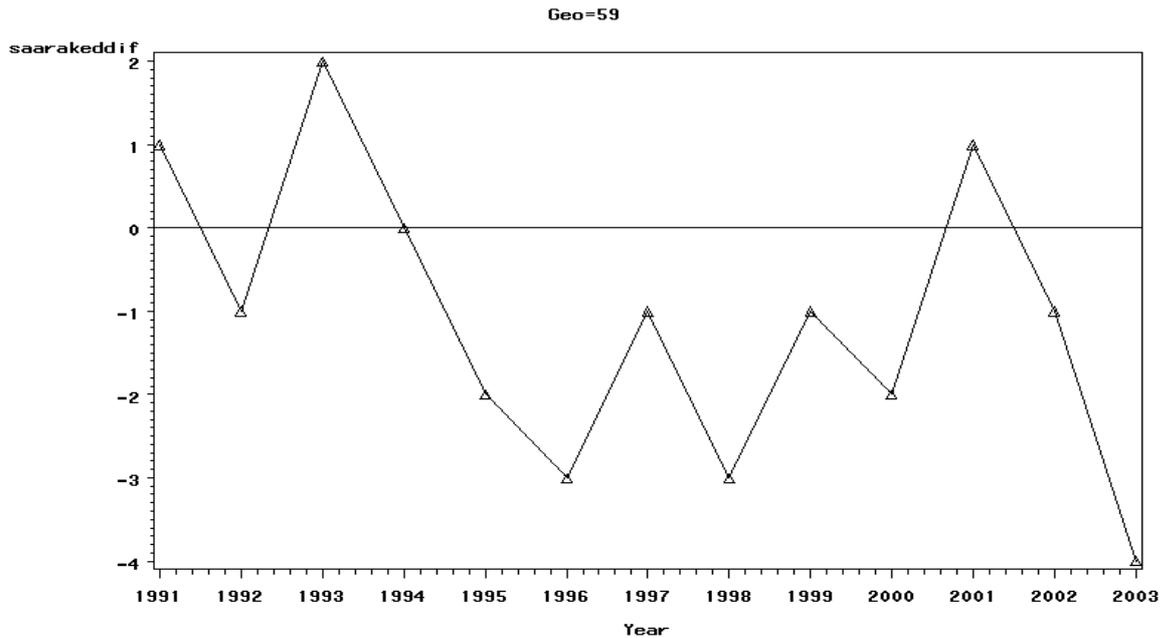


Figure 8: Differences in the annual totals for British Columbia MRTS SA series before and after reconciliation

Columbia. Figure 6 shows the difference between between the National total and the total from the regions, and the corresponding share allocated British Columbia, which corresponds to the differences between the two series in Figure 5. Figure 7 displays the relative differences, which are less than 2% in absolute values. Finally, Figure 8 displays the differences between the annual totals from the SA series before and after reconciliation, which are just small annoying differences due to the use of a small alterability coefficient for the annual totals of the 13 regional SA series.

In a case such as the MRTS, an alternative to using small alterability coefficients for the annual totals is obviously to initially reconcile the annual totals of the 13 regional SA series with the annual totals at the Canada level. This has the operational advantage of identifying cases where unbalanced raw series are being used. Then, once the annual totals are reconciled, they can be imposed as binding constraints. Although not illustrated in this paper, we tested that method and the results came out to be essentially the same.

This example illustrates that reconciliation of the regional SA series is a simple mathematical reallocation of the discrepancy between the national total and the sum of the regional series to avoid embarrassment to the publisher.

### 3.4 Reconciliation of 2-way classified SA series subject to the annual total constraint

Consider again the case of the MRTS where all the SA series at the industry-region classification must be reconciled. This case study pushes to the limit the user-friendly aspect of the regression model for reconciliation, in the sense that advanced programming

techniques are required to generate the ©SAS code for this example.

To be consistent with the previous section, assume that the Canada total is obtained from the 19 industries and that the 13 regions are already reconciled to it. The problem will now be to reconcile the  $19 \times 13 = 247$  SA series at the industry-region classification such that:

1. For every month and given an industry, the sum over the 13 regions matches the industry total. This gives 19 industry restrictions every month.
2. For every month and given a region, the sum over the 19 industries matches the regional total. This gives 13 regional restrictions every month.
3. For every complete year, and given an industry-region, the sum over the 12 months is as close as possible to the annual totals from the corresponding raw series. This gives 247 non-binding constraints.

The regression is done one month at a time in the current year where an annual total may not yet be available; the regression model has  $19 \times 13 = 247$  observations,  $19 \times 13 = 247$  regression variables, 13 regional total restrictions, 19 industry total restriction, and the reconciliation is done with the algorithm described to balance a 2-way table subject to fixed marginal totals. When an annual total is available, the regression is done with one year of observations at a time. The regression model contains:

1. 3211 observations made of the  $247 \times 12 = 2964$  industry-region SA series and their corresponding 247 annual totals.
2.  $19 \times 12 = 228$  industry total restrictions.
3.  $13 \times 12 = 156$  regional total restrictions.

Figure 9 shows the SA series before and after reconciliation for one industry (new car dealers) in the province of British Columbia, and Figure 10 displays their relative differences. Finally, Table 1 displays the annual totals from the raw, the SA series before and after reconciliation. Again, the differences in the annual totals are just small annoying differences.

## 4 Conclusions

This paper presented a simple method based on constrained weighted least squares regression to restore the additivity of a system of time series. The method was designed to be easily programmable with the ©SAS software. The main application of this method is to balance a table of seasonally adjusted series subjected to various constraints.

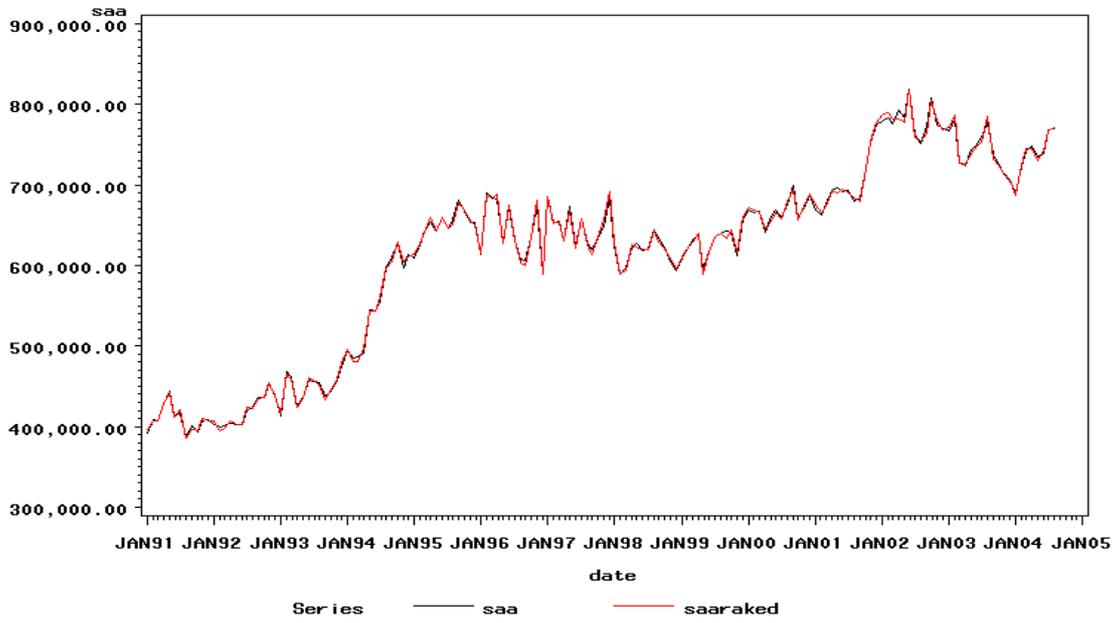


Figure 9: MRTS, New car dealers in British Columbia, SA series before and after reconciliation

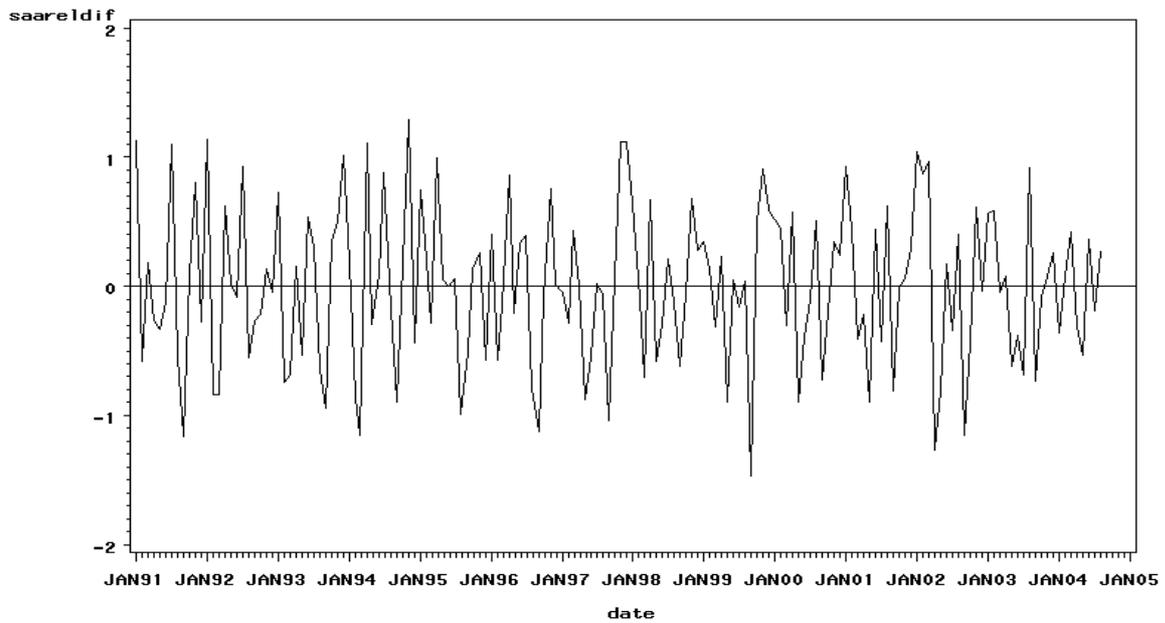


Figure 10: MRTS, New car dealers in British Columbia, relative differences in percentage in the SA series before and after reconciliation

Table 1: MRTS, New car dealers in British Columbia, Annual totals in the raw and SA series before and after reconciliation

Year	Raw	SA	Raked SA
1991	4905149.00	4905150.00	4905148.82
1992	5019263.00	5019264.00	5019263.65
1993	5384431.00	5384430.00	5384430.85
1994	6644360.00	6644360.00	6644360.46
1995	7790625.00	7790626.00	7790625.61
1996	7713968.00	7713969.00	7713968.41
1997	7789048.00	7789049.00	7789047.45
1998	7396166.00	7396166.00	7396164.34
1999	7536090.00	7536090.00	7536091.79
2000	8022211.00	8022210.00	8022209.43
2001	8384530.00	8384531.00	8384533.08
2002	9368059.00	9368060.00	9368059.25
2003	8909165.00	8909166.00	8909165.05

Current work in this project involves solving operational and implementation issues. For example, the regression approach is parameterized to simplify its implementation with ©SAS using the SAS macro language. Rounding algorithms are being added making sure constraints remain satisfied. As illustrated with the MRTS example, edit checks are being added to make sure the input series satisfy the assumptions underlying the method. And finally, a project team approach is used to make sure the resulting software will be documented, easy to maintain and applicable to various statistical programs that publish seasonally adjusted series.

## References

- BLOEM, A., R. DIPPELSMAN, AND N. MÆHEL (2001): *Quarterly National Accounts Manual, Concepts, Data Sources and Compilation*. International Monetary Fund, Washington DC.
- 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.
- DAGUM, E. B., AND P. CHOLETTE (To appear): *Benchmarking and Reconciliation of Time Series Data*. Springer-Verlag, New York.
- FINDLEY, D., B. MONSELL, W. BELL, M. OTTO, AND B. CHEN (1998): “New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program,” *Journal of Business and Economic Statistics*, 16, 127–177.

FORTIER, S. (2003): “La conversion de données historiques selon un nouveau système de classification pour l’Enquête mensuelle sur le commerce de gros et de détail,” *Proceedings of the Survey Method Section, 31st Annual Meeting of the Statistical Society of Canada*, pp. 13–18.

QUENNEVILLE, B., P. CHOLETTE, G. HUOT, AND T. DI FONZO (2004): “Adjustment of Seasonally Adjusted Series to Annual Totals,” *Proceedings of the Business and Economic Section of the American Statistical Association*.