A Mathematical Optimization Approach to Balancing Time Series: Statistics Canada's *GSeriesTSBalancing*

Michel Ferland[1], Susie Fortier[2] and Joana Bérubé[3]
Statistics Canada, 150 Tunney's Pasture Driveway, Ottawa ON, K1A 0T6, CANADA
[1]michel.ferland@canada.ca
[2]susie.fortier@canada.ca
[3]joana.berube@canada.ca

Abstract
Time series data produced by National Statistical Offices and Systems of National Accounts must often respect a vast array of accounting relationships. As the data may come from various sources or undergo non-linear data processing such as seasonal adjustment, the accounting relationships must often be restored before publication. The process used to restore the accounting coherence in the data is referred to as *balancing* or *reconciliation*. This paper introduces the ***GSeriesTSBalancing*** SAS® macro, a numerical optimization solution based on SAS/OR® for balancing a system of time series and the latest addition to Statistics Canada's G-Series software, formerly known as Forillon. The mathematical optimization problem is presented and illustrated with simple examples. A comparison with the TSRAKING procedure, an alternative regression-based solution available in G-Series for simple balancing problems, is also included.

Key Words: Balancing time series, reconciliation, multivariate benchmarking, accounting relationships.

1. Introduction

Time series data produced by National Statistical Offices and Systems of National Accounts must often respect a vast array of accounting relationships. Data from various sources such as surveys, censuses or administrative files may have small conceptual, methodological and/or operational discrepancies that could result in combined data that do not respect the accounting relationships. On the other hand, non-linear data processes such as seasonal adjustment may also break those relationships. Under these circumstances the accounting relationships will usually be restored before publication, a process referred to as *balancing* or *reconciliation*.

This paper introduces the ***GSeriesTSBalancing*** SAS® macro, a numerical optimization solution based on SAS/OR® for balancing a system of time series. It is the latest addition to Statistics Canada's G-Series software, formerly known as Forillon. This tool provides a solution for the second step of the two-step method described in Quenneville and Fortier (2012) where movement preservation is achieved in the first step as temporal benchmarks are used to adjust the level of the time series using a univariate benchmarking procedure such as the G-Series BENCHMARKING procedure (Latendresse, Djona and Fortier 2007). The paper is organized as follows: section 2 presents the mathematical optimization problem; section 3 presents the macro parameters; section 4 illustrates the usage of the macro with simple balancing examples;

section 5 compares the ***GSeriesTSBalancing*** macro with the TSRAKING procedure, an alternative regression-based solution also available in G-Series for simple balancing problems. Finally, a brief conclusion is provided in section 6.

2. Mathematical Optimization Approach

Time series balancing problems can be formulated as minimization problems under constraints. This section presents the two forms of minimization problems solved by the ***GSeriesTSBalancing*** macro followed by an interpretation of the main concepts.

2.1 Minimization Problem – Two Forms

Let's define x_{kt}, $\hat{\theta}_{kt}^{(x)}$ and $c_{kt}^{(x)}$ respectively as the initial value, the balanced value and the alterability coefficient (described later in section 2.2) of time series *k* at time *t*. Without temporal total preservation, the time series balancing problem can be formulated as the following simple quadratic minimization problem, defined for any time period *t*

$$\min_{\theta^{(x)}} \quad \sum_k \frac{\left(x_{kt} - \theta_{kt}^{(x)}\right)^2}{w_{kt}^{(x)}}, \quad \text{where} \quad w_{kt}^{(x)} = \begin{cases} 1 & \text{if } \left|c_{kt}^{(x)} x_{kt}\right| = 0 \\ \left|c_{kt}^{(x)} x_{kt}\right| & \text{otherwise} \end{cases} \tag{1}$$

subject to

$G\theta^{(x)} \text{ op } g$, the linear balancing constraints in matrix form where op is =, ≤ or ≥, $\theta^{(x)}$ is a vector representation of $\theta_{kt}^{(x)}$, *G* is an appropriately sized matrix describing the linear relationships and *g* is vector of constants;

$\theta_{kt}^{(x)} = x_{kt}$, if $\left|c_{kt}^{(x)} x_{kt}\right| = 0$, defining time series value x_{kt} as binding (fixed).

Let's now introduce the concept of temporal group by defining T_j the set of periods *t* belonging to complete temporal group *j* (e.g. a complete year). Let's also denote $a_{kj} = \sum_{t \in T_j} x_{kj}$, $\hat{\theta}_{kj}^{(a)} = \sum_{t \in T_j} \hat{\theta}_{kj}^{(x)}$ and $c_{kj}^{(a)}$ respectively as the initial temporal total, the balanced temporal total and the temporal total alterability coefficient of time series *k* for complete temporal group *j*. Therefore, with temporal total preservation, the quadratic minimization problem for complete temporal group *j* becomes

$$\min_{\theta^{(x,a)}} \quad \sum_{t \in T_j} \sum_k \frac{\left(x_{kt} - \theta_{kt}^{(x)}\right)^2}{w_{kt}^{(x)}} + \sum_k \frac{\left(a_{kj} - \theta_{kj}^{(a)}\right)^2}{w_{kj}^{(a)}} \tag{2}$$

where

$$w_{kt}^{(x)} = \begin{cases} 1 & \text{if } \left|c_{kt}^{(x)} x_{kt}\right| = 0 \\ \left|c_{kt}^{(x)} x_{kt}\right| & \text{otherwise} \end{cases} \quad \text{and} \quad w_{kj}^{(a)} = \begin{cases} 1 & \text{if } \left|c_{kj}^{(a)} a_{kj}\right| = 0 \\ \left|c_{kj}^{(a)} a_{kj}\right| & \text{otherwise} \end{cases}$$

subject to

$G\theta^{(x)} \text{ op } g$, the linear balancing constraints in matrix form described in (1);

$\theta_{kt}^{(x)} = x_{kt}$, if $\left|c_{kt}^{(x)} x_{kt}\right| = 0$, defining time series value x_{kt} as binding (fixed);

$\sum_{t \in T_j} \theta_{kt}^{(x)} = \theta_{kj}^{(a)}$, the implicit temporal constraints;

$\theta_{kj}^{(a)} = a_{kj}$, if $\left|c_{kj}^{(a)} a_{kj}\right| = 0$, defining time series temporal total a_{kj} as binding (fixed).

While minimization problem (2) involves several periods (the set of periods belonging to a complete temporal group), minimization problem (1) involves a single period. Balancing a system of time series usually requires solving several independent quadratic minimization problems of either form (1) or (2). The total number of problems to be solved depends on the number of periods available as input and whether temporal totals are preserved or not; we refer to the set of periods involved in a given problem as a *processing group*.

2.2 Interpretation

The general idea behind the time series balancing solution presented here is to minimize the relative distance between the time series initial values, denoted x_{kt} in minimization problems (1) and (2), and their balanced (final) values $\hat{\theta}_{kt}^{(x)}$ subject to two types of constraints:

- **Balancing constraints**: multivariate linear constraints provided by the user that define the relationships that need to be restored (must hold after balancing) between the different time series, for each individual time period t of the system of time series. They are the $G\theta^{(x)}\ \texttt{op}\ g$ constraints, where $\texttt{op}$ stands for operator (=, ≤ or ≥). These constraints are sometimes called *cross-sectional* or *contemporaneous* constraints.
- **Implicit temporal constraints**: univariate temporal aggregation constraints automatically added to the minimization problem when temporal totals are preserved. Temporal constraints are defined for every time series of the minimization problem and impose that the sum of the balanced values $\hat{\theta}_{kt}^{(x)}$ of a given time series k over the periods t of a complete temporal group j $\left(\text{for all } t \in T_j\right)$ must be equal to the balanced (final) temporal total $\hat{\theta}_{kj}^{(a)}$. They are constraints $\sum_{t\in T_j} \theta_{kt}^{(x)} = \theta_{kj}^{(a)}$, where the distance between the initial temporal total a_{kj} and the balanced temporal total $\hat{\theta}_{kj}^{(a)}$ is minimized in the objective function.

The implicit temporal constraints are not defined for incomplete temporal groups (e.g. incomplete years) resulting in minimization problem (1) being used for all periods belonging to an incomplete temporal group. If the balancing problem does not involve temporal total preservation, then the minimization problem (1) would be used for all periods.

Alterability coefficients are non-negative numbers that change the relative cost of modifying an initial value. By changing the actual objective function to minimize, they allow the generation of a wide range of solutions. Similarly to the constraints, there are two types of alterability coefficients: *regular alterability coefficients* $c_{kt}^{(x)}$ and *temporal total alterability coefficients* $c_{kj}^{(a)}$. Since alterability coefficients appear in the denominator of the objective function distance terms, the larger the coefficient the less costly it is to modify a time series value (or temporal total) and, conversely, the smaller the coefficient the more costly it becomes. This results in time series values (and temporal totals) with larger alterability coefficients proportionally changing more than the ones with smaller alterability coefficients. When an alterability coefficient is 0, an additional constraint on the corresponding time series value (or temporal total) is added to the problem $\left(\text{constraints } \theta_{kt}^{(x)} = x_{kt} \text{ and } \theta_{kj}^{(a)} = a_{kj}\right)$ and we say that the corresponding time series value (or temporal total) is *binding*, referring to the fact that it cannot be modified (it is non-alterable or fixed). Conversely, we say that a time series value (or temporal total) is *non-binding* when its alterability coefficient is greater than 0. Time series temporal totals are usually binding as well as values of time series corresponding to marginal totals in aggregation table raking problems (i.e. time series with a coefficient of -1 in the balancing aggregation constraints) while the values of component series in those raking problems (i.e. time series with a coefficient of 1 in the balancing aggregation constraints) are usually non-binding. In practice, *almost binding* time series values (or temporal totals) can be achieved by specifying very small alterability coefficients (almost 0). This approach can be used to avoid infeasible problems that may result from small inconsistencies in the data (e.g. fully specified multi-dimensional raking problems with temporal total preservation). The term *temporal total preservation* refers to the fact that temporal totals are always kept "as close as possible" to their initial value. *Pure preservation* is achieved for binding temporal totals $\left(c_{kj}^{(a)}=0\right)$ while the change is minimized for non-binding temporal totals $\left(c_{kj}^{(a)}>0\right)$.

3. The *GSeriesTSBalancing* Macro

The ***GSeriesTSBalancing*** macro solves the minimization problems under constraints presented in section 2 using numerical solvers available in SAS/OR through the use of the OPTMODEL procedure. SAS/OR version 9.3 or more recent is required in order to use ***GSeriesTSBalancing*** V1.01.

3.1 Input Time Series Data Set

The time series data may be provided to the ***GSeriesTSBalancing*** macro either in a *wide* format, where time series data appear in separate columns (variables) in the data set (one column per time series), or in a *tall* format where the time series data are stacked in a single column instead. The output (balanced) time series data set is generated in the same format as the input data set. Balancing examples provided in section 4 use the default *wide* data format. A time variable that contains SAS date, time or date-time values identifying time periods *t* should be included in the input data set when the balancing problem involves temporal total preservation.

3.2 Problem Specification Data Set

The details of the balancing problem are provided to the ***GSeriesTSBalancing*** macro through the *problem specification data set*. Using a sparse format, this data set allows the user to specify only the relevant information such as the non-zero coefficients of the balancing constraints, the non-default alterability coefficients and the time series lower and upper bounds. The problem specification data set is inspired from the LP procedure's *sparse data input format* in SAS (SAS Institute 2011). The problem specification data set is composed of 4 mandatory variables, `_TYPE_`, `_COL_`, `_ROW_`, `_COEF_`, and one optional variable, `_TIMEVAL_`. An observation in the problem specification data set either defines a label for one of the 7 types of the balancing problem elements with variables `_TYPE_` and `_ROW_` (see **Table 1**) or specifies coefficients or numerical values for those balancing problem elements with variables `_COL_`, `_ROW_`, `_COEF_` and `_TIMEVAL_` (see **Table 2**). Variable `_ROW_` identifies the elements of the balancing problem and is the key variable that makes the link between both types of records in the problem specification data set. The same label (`_ROW_`) cannot be associated with more than one problem element (_TYPE_) and multiple labels cannot be defined for a given type of problem element (_TYPE_), except for constraints (_TYPE_=EQ, =LE or =GE). The order of the observations in the problem specification data set is unimportant and character values are case insensitive (variables _TYPE_, _COL_ and _ROW_). Examples of the problem specification data set are provided in section 4.

Table 1: Label Definition Records

Variable	Type	Description
TYPE	Character	Reserved keyword that identifies one of the 7 types of balancing problem elements and tells the macro how to interpret *Information Specification Records*: • EQ: equality (=) constraint • LE: lower or equal (≤) inequality constraint • GE: greater or equal (≥) inequality constraint • lowerBd: time series lower bound • upperBd: time series upper bound • alter: time series regular alterability coefficient • alterTmp: time series temporal total alterability coefficient.
ROW	Character	Label to be associated with the "type keyword" (balancing problem element).

Table 2: Information Specification Records

Variable	Type	Description
COL	Character	Reserved word _rhs_ to specify a right-hand-side constraint bound or the name of the time series to which the information applies.
ROW	Character	Label associated with a "type keyword" (balancing problem element) identifying the type information being specified.
COEF	Numeric	Value of a constraint bound (non-zero element of vector g) when _COL_=_rhs_ or, otherwise, time series coefficient or numerical value in accordance with the label (_ROW_): • non-zero constraint coefficient (element of matrix G); • or lower or upper bound on $\theta_{kt}^{(x)}$; • or non-default regular alterability coefficient $c_{kt}^{(x)}$ (default coefficient is 1); • or non-default temporal total alterability coefficient $c_{kj}^{(a)}$ (default coefficient is 0).
TIMEVAL	Numeric	Optional time value to restrict the application of alterability coefficients or bounds to a specific time period. By default, alterability coefficients and bounds apply to all time periods (when _TIMEVAL_ is not provided in the problem specification data set or when its value is *missing*). For problems without a time variable, _TIMEVAL_ values correspond to observation numbers in the input data set.

Table 3: ***GSeriesTSBalancing*** Macro Parameters

Parameter Name	**Short Description**
`inTS`	*Mandatory*. Input (unbalanced) time series data set.
`inProblemSpecs`	*Mandatory*. Problem specification data set.
`OutTS`	*Mandatory*. Output (balanced) time series data set.
`timeVarName` (i)	Time variable in the input data set (contains the SAS date, time or date-time values).
`periodInterval` (i)	SAS date, time or date-time interval associated with the periods in the input time series data set.
`temporalGrpInterval` (i)	SAS date, time or date-time interval associated with the temporal totals to be preserved. *Not to be specified for balancing problems without temporal total preservation*.
`optModelSolverOptions` (ii)	Options for the SOLVE statement of the OPTMODEL procedure (default is `WITH QP`).
`displayLevel`	Level of information displayed in the Log and active ODS destinations (e.g. Output window).
`outOptModelSummary`	*OPTMODEL summary* output data set.
`outDetailedResults`	*Detailed balancing results* output data set.
`outEvaluatedConstraints`	*Evaluated constraints* output data set.
`outTemporalTotals`	*Temporal totals* output data set.
`outProcessingGrps`	*Processing groups* output data set.
`alterPos`, `alterNeg` and `alterMix`	Regular alterability coefficients (default is 1).
`alterTemporalTotals`	Temporal total alterability coefficients (default is 0).
`lowerBound` and `upperBound`	Bounds for the time series values.
`toleranceValue`	Tolerance in absolute value for the balancing constraints.
`toleranceValueTemporal` and `tolerancePercentTemporal`	Tolerance for the implicit temporal constraints specified either in absolute value or in percentage.
`TSFormat`	Definition of the format of the input and output data sets (`WIDE` or `TALL`; default is `WIDE`).
`tallTSIDVarName`	Variable in the input data set that contains the time series names when parameter `TSFormat=TALL` (default is `_NAME_`).
`tallTSValueVarName`	Variable in the input data set that contains the time series values when parameter `TSFormat=TALL` (default is `_VALUE_`).
`tallTSAlterVarName`	Variable in the input data set that contains the time series alterability coefficients when parameter `TSFormat=TALL`.
`Language`	Display language for messages in the Log (`EN` for English or `FR` for French; default is `EN`).

(i) Consult the SAS documentation for more details on SAS date, time or date-time values and intervals.
(ii) Consult the SAS OPTMODEL procedure documentation for more details on the SOLVE statement including available solvers and options.

3.3 *GSeriesTSBalancing* Macro Parameters

Parameters allow the user to specify the input and output data sets, the problem specification data set as well as time related information such as the frequency of the time series and temporal totals to be preserved (when applicable). Other features available through the ***GSeriesTSBalancing*** macro parameters include changing the default alterability coefficients, specifying time series lower or upper bounds, setting the time series data set format, creating optional balancing diagnostics data sets, specifying tolerances for the balancing constraints, etc.

The following example of SAS code illustrates a basic call to the ***GSeriesTSBalancing*** macro where the most commonly used parameters are shown.

```
%GSeriesTSBalancing
(
   /* Mandatory Parameters */
   inTS                 = myDataIn,
   inProblemSpecs       = myProblem,
   outTS                = myDataOut,

   /* Optional Parameters */
   timeVarName          = Date,
   periodInterval       = MONTH,
   temporalGrpInterval = YEAR
)
```

Table 3 on the previous page lists all available macro parameters along with a brief description of each parameter including the default value of optional parameters (when relevant). Mandatory parameters are highlighted in light red in the table while the most commonly used optional parameters are highlighted in light blue. A complete description of all parameters can be found in the ***GSeriesTSBalancing*** macro user guide (Statistics Canada 2016).

4. Examples

4.1 Example 1

In this first example, the objective is to balance the following simple accounting table (`Profits` = `Revenues` – `Expenses`) without modifying the `Profits` and where `Revenues` ≥ 0 and `Expenses` ≥ 0.

Table 4: Unbalanced Accounting Table (million $)

Row	Revenues	Expenses	Profits
1	15	10	10
2	4	8	−1
3	250	250	5
4	8	12	0
5	0	45	-55

In addition to the accounting rule, expressed as `Revenues` – `Expenses` – `Profits` = 0, the problem specification data set should contain the non-default alterability coefficient of 0 for `Profits` and lower bounds of 0 for `Revenues` and `Expenses`.

Table 5: Problem Specification Data Set

TYPE	_COL_	_ROW_	_COEF_
EQ		Accounting Rule	.
	Revenues	Accounting Rule	1
	Expenses	Accounting Rule	-1
	Profits	Accounting Rule	-1
alter		Alterability Coefficient	.
	Profits	Alterability Coefficient	0
lowerBd		Lower Bound	.
	Revenues	Lower Bound	0
	Expenses	Lower Bound	0

This simple example involves 5 independent minimization problems (processing groups), one for each row of the accounting table.

Table 6: Balanced Accounting Table (million $)

Row	Revenues	Expenses	Profits
1	18	8	10
2	5	6	−1
3	252.5	247.5	5
4	9.6	9.6	0
5	0	55	-55

The solution returned by the ***GSeriesTSBalancing*** macro corresponds to equal proportional changes (pro-rating) and is related to the default alterability coefficients of 1. Equal absolute changes could be obtained instead by specifying alterability coefficients equal to one over the initial values $\left(c_{kt}^{(x)} = 1/x_{kt}\right)$ in the problem specification data set. Let's do this for the 2nd row of the accounting table.

Table 7: Problem Specification Data Set – Equal Change for Row 2

TYPE	_COL_	_ROW_	_COEF_	_TIMEVAL_
		...		
alter		Alterability Coefficient	.	.
	Profits	Alterability Coefficient	0	.
	Revenues	*Alterability Coefficient*	*0.25*	*2*
	Expenses	*Alterability Coefficient*	*0.125*	*2*
		...		

Table 8: Balanced Accounting Table (million $) – Equal Change for Row 2

Row	Revenues	Expenses	Profits
1	18	8	10
2	*5.5*	*6.5*	*−1*
3	252.5	247.5	5
4	9.6	9.6	0
5	0	55	-55

The SAS code used to obtain the balanced data in **Table 8** is provided on the next page.

```
data myData1;
   input Row Revenues Expenses Profits;
   datalines;
1   15   10   10
2    4    8   -1
3  250  250    5
4    8   12    0
5    0   45  -55
;

data myProblem1;
   length _TYPE_ $7 _COL_ $8 _ROW_ $24;
   input _TYPE_= _COL_= _ROW_= _COEF_= _TIMEVAL_=;
   datalines;
_TYPE_=EQ      _COL_=.        _ROW_=Accounting Table         _COEF_=.     _TIMEVAL_=.
_TYPE_=.       _COL_=Revenues _ROW_=Accounting Table         _COEF_=1     _TIMEVAL_=.
_TYPE_=.       _COL_=Expenses _ROW_=Accounting Table         _COEF_=-1    _TIMEVAL_=.
_TYPE_=.       _COL_=Profits  _ROW_=Accounting Table         _COEF_=-1    _TIMEVAL_=.
_TYPE_=alter   _COL_=.        _ROW_=Alterability Coefficient _COEF_=.     _TIMEVAL_=.
_TYPE_=.       _COL_=Profits  _ROW_=Alterability Coefficient _COEF_=0     _TIMEVAL_=.
_TYPE_=.       _COL_=Revenues _ROW_=Alterability Coefficient _COEF_=0.25  _TIMEVAL_=2
_TYPE_=.       _COL_=Expenses _ROW_=Alterability Coefficient _COEF_=0.125 _TIMEVAL_=2
_TYPE_=lowerBd _COL_=.        _ROW_=Lower Bound              _COEF_=.     _TIMEVAL_=.
_TYPE_=.       _COL_=Revenues _ROW_=Lower Bound              _COEF_=0     _TIMEVAL_=.
_TYPE_=.       _COL_=Expenses _ROW_=Lower Bound              _COEF_=0     _TIMEVAL_=.
;

%GSeriesTSBalancing
(
   inTS           = myData1,
   inProblemSpecs = myProblem1,
   outTS          = outBalanced1
)
```

4.2 Example 2

In this second example, consider the simulated data on vehicle sales shown in the table on the following page. The data consist of quarterly sales of vehicles by region (West, Centre and East), along with a national total representing the three regions, and by type of vehicles (cars, trucks and a total that may include other types of vehicles). Column "Sum" is not part of the input data set but has been added to illustrate the initial discrepancies between the sum of the regions and the national total. The data in this table correspond to directly seasonally adjusted data that have been benchmarked to the annual totals of the corresponding unadjusted time series data as part of the seasonal adjustment process (e.g. with the FORCE spec in the X-12-ARIMA software). This explains why the sum of the regional sales agree with the national sales for rows "Total for 2015" in the table. Column "Sum" would be the indirectly seasonally adjusted national sales. The objective is to reconcile the regional sales to the national sales without modifying the latter while ensuring that the sum of the sales of cars and trucks do not exceed 95% of the sales for all types of vehicles in any quarter. For illustrative purposes, suppose that the sales of trucks in the Centre region for the second quarter of 2015 cannot be modified. This balancing problem includes 2 sets balancing constraints:

$$\begin{aligned} &\texttt{West\_AllTypes} + \texttt{Centre\_AllTypes} + \texttt{East\_AllTypes} = \texttt{National\_AllTypes} \\ &\texttt{West\_Cars} + \texttt{Centre\_Cars} + \texttt{East\_Cars} = \texttt{National\_Cars} \\ &\texttt{West\_Trucks} + \texttt{Centre\_Trucks} + \texttt{East\_Trucks} = \texttt{National\_Trucks} \end{aligned}$$

$$\begin{aligned} &\texttt{West\_Cars} + \texttt{West\_Trucks} \le 0.95 * \texttt{West\_AllTypes} \\ &\texttt{Centre\_Cars} + \texttt{Centre\_Trucks} \le 0.95 * \texttt{Centre\_AllTypes} \\ &\texttt{East\_Cars} + \texttt{East\_Trucks} \le 0.95 * \texttt{East\_AllTypes} \end{aligned}$$

Table 9: Number of Vehicles Sold (in thousands) – Seasonally Adjusted

Period	Type of Vehicle	West	Centre	East	Sum	National
2015Q1	All Types	43	49	47	139	136
	Cars	20	18	12	50	53
	Trucks	20	22	26	68	61
2015Q2	All Types	40	45	42	127	114
	Cars	16	16	19	51	44
	Trucks	21	26	21	68	59
2015Q3	All Types	35	47	40	122	133
	Cars	14	15	16	45	50
	Trucks	19	25	19	63	71
2015Q4	All Types	44	44	45	133	138
	Cars	19	20	14	53	52
	Trucks	21	18	27	66	74
Total for 2015	**All Types**	**162**	**185**	**174**	**521**	**521**
	Cars	**69**	**69**	**61**	**199**	**199**
	Trucks	**81**	**91**	**93**	**265**	**265**
2016Q1	All Types	46	48	55	149	135
	Cars	16	15	19	50	51
	Trucks	27	25	28	80	54

Table 10: Problem Specification Data Set

TYPE	_COL_	_ROW_	_COEF_	_TIMEVAL_
EQ		AllTypes Agg	.	.
	West_AllTypes	AllTypes Agg	1	.
	Centre_AllTypes	AllTypes Agg	1	.
	East_AllTypes	AllTypes Agg	1	.
	National_AllTypes	AllTypes Agg	-1	.
EQ		Cars Agg	.	.
	West_Cars	Cars Agg	1	.
	Centre_Cars	Cars Agg	1	.
	East_Cars	Cars Agg	1	.
	National_Cars	Cars Agg	-1	.
EQ		Trucks Agg	.	.
	West_Trucks	Trucks Agg	1	.
	Centre_Trucks	Trucks Agg	1	.
	East_Trucks	Trucks Agg	1	.
	National_Trucks	Trucks Agg	-1	.
LE		West Reg Sum	.	.
	West_Cars	West Reg Sum	1	.
	West_Trucks	West Reg Sum	1	.
	West_AllTypes	West Reg Sum	-0.95	.
LE		Centre Reg Sum	.	.
	Centre_Cars	Centre Reg Sum	1	.
	Centre_Trucks	Centre Reg Sum	1	.
	Centre_AllTypes	Centre Reg Sum	-0.95	.
LE		East Reg Sum	.	.
	East_Cars	East Reg Sum	1	.
	East_Trucks	East Reg Sum	1	.
	East_AllTypes	East Reg Sum	-0.95	.
Alter		Alter Coef	.	.
	National_AllTypes	Alter Coef	0	.
	National_Cars	Alter Coef	0	.
	National_Trucks	Alter Coef	0	.
	Centre_Trucks	Alter Coef	0	2015Q2

The national sales time series (`National_AllTypes`, `National_Cars` and `National_Trucks`) are binding and alterability coefficients of 0 are therefore specified in the problem specification data set found on the previous page. All other time series alterability coefficients are 1 (default value), except for `Centre_Trucks` on 20015Q2 where it is 0.

Annual totals for all time series are automatically preserved since the default value of temporal total alterability coefficients is 0 in the ***GSeriesTSBalancing*** macro. All time series of this balancing problem are positive, which can be simply specified in the macro call with the `lowerBound` parameter. This balancing problem involves only 2 processing groups (2 independent minimization problems): i) the 4 quarters of 2015 and ii) period 2016Q1. The reconciled data resulting from this time series balancing problem solved with ***GSeriesTSBalancing*** are presented in the following table, where numbers in red *italic* characters correspond to values that were not modified (alterability coefficients of 0).

Table 11: Reconciled Number of Vehicles Sold (in thousands) - Seasonally Adjusted

Period	Type of Vehicle	West	Centre	East	Sum	National
2015Q1	All Types	42.109	47.637	46.254	136	*136*
	Cars	21.156	19.134	12.710	53	*53*
	Trucks	18.561	18.594	23.845	61	*61*
2015Q2	All Types	35.311	41.409	37.280	114	*114*
	Cars	14.005	13.338	16.657	44	*44*
	Trucks	16.615	*26*	16.385	59	*59*
2015Q3	All Types	38.895	50.581	43.524	133	*133*
	Cars	15.241	16.848	17.911	50	*50*
	Trucks	21.710	27.229	22.061	71	*71*
2015Q4	All Types	45.685	45.373	46.942	138	*138*
	Cars	18.598	19.680	13.722	52	*52*
	Trucks	24.114	19.177	30.709	74	*74*
Total for 2015	**All Types**	***162***	***185***	***174***	**521**	***521***
	Cars	***69***	***69***	***61***	**199**	***199***
	Trucks	***81***	***91***	***93***	**265**	***265***
2016Q1	All Types	41.678	43.490	49.832	135	*135*
	Cars	16.320	15.300	19.380	51	*51*
	Trucks	18.225	16.875	18.900	54	*54*

The SAS code used to solve this example is provided below.

```
data myData2;
   input Date
         West_AllTypes Centre_AllTypes East_AllTypes National_AllTypes
         West_Cars     Centre_Cars     East_Cars     National_Cars
         West_Trucks   Centre_Trucks   East_Trucks   National_Trucks;
   informat date yyq6.;
   format date yyq6.;
   datalines;
2015q1  43  49  47  136  20  18  12  53  20  22  26  61
2015q2  40  45  42  114  16  16  19  44  21  26  21  59
2015q3  35  47  40  133  14  15  16  50  19  25  19  71
2015q4  44  44  45  138  19  20  14  52  21  18  27  74
2016q1  46  48  55  135  16  15  19  51  27  25  28  54
;
```

```
data myProblem2;
   length _TYPE_ $5 _COL_ $17 _ROW_ $14;
   input _TYPE_= _COL_= _ROW_= _COEF_= _TIMEVAL_=;
   informat _TIMEVAL_ yyq6.;
   format _TIMEVAL_ yyq6.;
   datalines;
_TYPE_=EQ    _COL_=.                 _ROW_=AllTypes Agg   _COEF_=.    _TIMEVAL_=.
_TYPE_=.     _COL_=West_AllTypes     _ROW_=AllTypes Agg   _COEF_=1    _TIMEVAL_=.
_TYPE_=.     _COL_=Centre_AllTypes   _ROW_=AllTypes Agg   _COEF_=1    _TIMEVAL_=.
_TYPE_=.     _COL_=East_AllTypes     _ROW_=AllTypes Agg   _COEF_=1    _TIMEVAL_=.
_TYPE_=.     _COL_=National_AllTypes _ROW_=AllTypes Agg   _COEF_=-1   _TIMEVAL_=.
_TYPE_=EQ    _COL_=.                 _ROW_=Cars Agg       _COEF_=.    _TIMEVAL_=.
_TYPE_=.     _COL_=West_Cars         _ROW_=Cars Agg       _COEF_=1    _TIMEVAL_=.
_TYPE_=.     _COL_=Centre_Cars       _ROW_=Cars Agg       _COEF_=1    _TIMEVAL_=.
_TYPE_=.     _COL_=East_Cars         _ROW_=Cars Agg       _COEF_=1    _TIMEVAL_=.
_TYPE_=.     _COL_=National_Cars     _ROW_=Cars Agg       _COEF_=-1   _TIMEVAL_=.
_TYPE_=EQ    _COL_=.                 _ROW_=Trucks Agg     _COEF_=.    _TIMEVAL_=.
_TYPE_=.     _COL_=West_Trucks       _ROW_=Trucks Agg     _COEF_=1    _TIMEVAL_=.
_TYPE_=.     _COL_=Centre_Trucks     _ROW_=Trucks Agg     _COEF_=1    _TIMEVAL_=.
_TYPE_=.     _COL_=East_Trucks       _ROW_=Trucks Agg     _COEF_=1    _TIMEVAL_=.
_TYPE_=.     _COL_=National_Trucks   _ROW_=Trucks Agg     _COEF_=-1   _TIMEVAL_=.
_TYPE_=LE    _COL_=.                 _ROW_=West Reg Sum   _COEF_=.    _TIMEVAL_=.
_TYPE_=.     _COL_=West_Cars         _ROW_=West Reg Sum   _COEF_=1    _TIMEVAL_=.
_TYPE_=.     _COL_=West_Trucks       _ROW_=West Reg Sum   _COEF_=1    _TIMEVAL_=.
_TYPE_=.     _COL_=West_AllTypes     _ROW_=West Reg Sum   _COEF_=-.95 _TIMEVAL_=.
_TYPE_=LE    _COL_=.                 _ROW_=Centre Reg Sum _COEF_=.    _TIMEVAL_=.
_TYPE_=.     _COL_=Centre_Cars       _ROW_=Centre Reg Sum _COEF_=1    _TIMEVAL_=.
_TYPE_=.     _COL_=Centre_Trucks     _ROW_=Centre Reg Sum _COEF_=1    _TIMEVAL_=.
_TYPE_=.     _COL_=Centre_AllTypes   _ROW_=Centre Reg Sum _COEF_=-.95 _TIMEVAL_=.
_TYPE_=LE    _COL_=.                 _ROW_=East Reg Sum   _COEF_=.    _TIMEVAL_=.
_TYPE_=.     _COL_=East_Cars         _ROW_=East Reg Sum   _COEF_=1    _TIMEVAL_=.
_TYPE_=.     _COL_=East_Trucks       _ROW_=East Reg Sum   _COEF_=1    _TIMEVAL_=.
_TYPE_=.     _COL_=East_AllTypes     _ROW_=East Reg Sum   _COEF_=-.95 _TIMEVAL_=.
_TYPE_=alter _COL_=.                 _ROW_=Alter Coef     _COEF_=.    _TIMEVAL_=.
_TYPE_=.     _COL_=National_AllTypes _ROW_=Alter Coef     _COEF_=0    _TIMEVAL_=.
_TYPE_=.     _COL_=National_Cars     _ROW_=Alter Coef     _COEF_=0    _TIMEVAL_=.
_TYPE_=.     _COL_=National_Trucks   _ROW_=Alter Coef     _COEF_=0    _TIMEVAL_=.
_TYPE_=.     _COL_=Centre_Trucks     _ROW_=Alter Coef     _COEF_=0    _TIMEVAL_=2015Q2
;

%GSeriesTSBalancing
(
   /* Mandatory Parameters */
   inTS                 = myData2,
   inProblemSpecs       = myProblem2,
   outTS                = outBalanced2,

   /* Optional Parameters */
   timeVarName          = Date,
   periodInterval       = QTR,
   temporalGrpInterval = YEAR,   /* annual total preservation */
   lowerBound           = 0      /* positive data */
)
```

5. Comparison with the TSRAKING Procedure

Statistics Canada’s G-Series software includes another tool to reconcile time series, the TSRAKING procedure. It is a SAS-based procedure that reconciles aggregation tables using the Dagum and Cholette (2006) regression based approach. The solution is obtained with the Generalized Least Squares method and matrix manipulation. See Bérubé and Fortier (2009) for more details on the TSRAKING procedure. A comparison of the both tools follows.

- TSRAKING is limited to one- and two-dimensional aggregation table raking problems (with temporal total preservation if required) while ***GSeriesTSBalancing*** handles more general balancing problems (e.g. higher dimensional raking problems, non-negative solutions, general linear equality and inequality constraints as opposed to aggregation rules only, etc).

- While ***GSeriesTSBalancing*** and TSRAKING allow the preservation of temporal totals, time management is not incorporated in TSRAKING. For example, the construction of the processing groups is left to the user with TSRAKING and separate calls for each processing group must be submitted.

- ***GSeriesTSBalancing*** accommodates the specification of sparse problems in their reduced form. This is not the case of TSRAKING where aggregation rules must always be fully specified.

- Both tools handle negative values in the input data differently. While the solutions of raking problems obtained from ***GSeriesTSBalancing*** and TSRAKING are identical when all input data points are positive, they may differ if some data points are negative. For example, negative data points in a one-dimensional raking problem would move in opposite directions.

- ***GSeriesTSBalancing*** is usually faster than TSRAKING (see **Table 12**) but the usage of the *Moore-Penrose* general inverse in TSRAKING makes it generally less sensitive than ***GSeriesTSBalancing*** to small inconsistencies that sometimes occur in fully specified (over-specified) multi-dimensional raking problems with temporal total preservation.

Table 12: Comparison of Processing Times (real time on a Windows® PC)

<table>
<tr><td colspan="2">Test Case #1
Small two-dimensional raking problem (3 by 2 aggregation table) for daily values with annual constraints (annual total preservation):
• 365 periods (days) consisting of 1 complete year;
• 1 processing group.</td></tr>
<tr><td>TSRAKING
8 min 32 sec</td><td>GSeriesTSBalancing
Less than 10 sec</td></tr>
<tr><td colspan="2">Test Case #2
17 raking problems solved successively (15 one-dim. and 2 large two-dim. tables) for monthly values with annual constraints (annual total preservation):
• 171 periods (months) consisting of 14 complete years and one incomplete year composed of 3 months;
• 289 processing groups in total (17 problems times 17 processing groups).</td></tr>
<tr><td>TSRAKING
32 min 46 sec</td><td>GSeriesTSBalancing
1 min 33 sec</td></tr>
</table>

Test Case #3	
Complex mixture of 62 one- and two-dimensional raking problems solved successively for monthly values without temporal constraints: • 456 periods (months) for most problems, fewer for some problems; • 27,500 processing groups in total.	
TSRAKING	***GSeriesTSBalancing***
19 min 15 sec	8 min 53 sec

Two main factors can explain the gains in processing time of the ***GSeriesTSBalancing*** macro over the TSRAKING procedure:

a) the efficiency of the optimization approach, namely SAS/OR's QP solver, over the matrix manipulation (inversion) involved in the regression approach;
b) the internal management of time and construction of the processing groups resulting in a single call to ***GSeriesTSBalancing*** for a given raking problem versus several calls to TSRAKING which involves duplication of pre- and post-processing operations.

6. Conclusion

With ***GSeriesTSBalancing***, a numerical optimisation solution based on SAS/OR for balancing a system of time series is now available in Statistics Canada's G-Series software. The macro can be used to reconcile a system of time series measured at the same frequency according to a flexible set of linear constraints. It offers an efficient way of applying the second step of the two-step method described in Quenneville and Fortier (2012) where movement preservation is achieved in the first step by applying a univariate benchmarking procedure. In terms of performance, the new macro compares favorably to the TSRAKING procedure, an alternative tool based on matrix manipulation also available in G-Series.

For more information on ***GSeriesTSBalancing*** or on Statistics Canada's G-Series software, please contact the G-Series support team at G-Series@statcan.ca.

References

Bérubé, J. and Fortier, S. (2009). PROC TSRAKING: An in-house SAS® procedure for balancing time series. *JSM Proceedings, Business and Economic Section*. Alexandria, VA: American Statistical Association.

Dagum, E. B. and Cholette, P. (2006). *Benchmarking, Temporal Distribution and Reconciliation Methods of Time Series*. Springer-Verlag, New York, Lecture notes in Statistics, #186.

Latendresse, E., Djona, M., and Fortier, S. (2007). Benchmarking Sub-Annual Series to Annual Totals – From Concepts to SAS® Procedure and SAS® Enterprise Guide® Custom Task. *Proceedings of the SAS Global Forum*, April 2007.

Quenneville, B., and Fortier, S. (2012). Restoring accounting constraints in time series – Methods and software for a statistical agency. *Economic Time Series: Modeling and Seasonality*. Chapman & Hall/CRC.E.

SAS Institute Inc. (2011). *SAS/OR 9.3 Online Help and Documentation*. https://support.sas.com/documentation/onlinedoc/or/index.html#or93

Statistics Canada (2016). The ***GSeriesTSBalancing*** Macro User Guide. *G-Series 2.00.001 User Guides*. Statistics Canada, Ottawa, Canada.