# The G-Series software version 2.00

This set of documents is intended to guide you in using the G-Series SAS system. The options, statements and macro parameters for each of the G-Series procedures and macros are explained. The guides also offer one or more examples of SAS code that you can copy and paste into your SAS session editor and run as is.

---

## Overview

G-Series is the project name grouping together different SAS procedures and macros related to time series. These tools are developed at Statistics Canada by the System Engineering Division in collaboration with Business Survey Methods Division's Time Series and Research Analysis Centre.

Two procedures and one macro are available:

- PROC BENCHMARKING: This procedure will restore coherence between time series data of the same target variable measured at different frequencies (e.g. sub-annually and annually). This procedure is available in G-Series production versions v1.04 and v2.0.

- PROC TSRAKING: This procedure will restore additivity in a system of time series measured at the same frequency. This procedure is available in G-Series production versions v1.04 and v2.0.

- Macro **GSeriesTSBalancing**: This macro balances (reconciles) a system of time series measured at the same frequency according to a set of linear constraints. This procedure is available in G-Series production version v2.0.

For more information, please contact the G-Series support team using the G-Series e-mail address (G-Series@statcan.gc.ca) or consult the web site at G-Series (Statistics Canada intranet only).

# The BENCHMARKING Procedure

## Overview

This procedure will ensure coherence between time series data of the same target variable measured at different frequencies, e.g. sub-annually and annually. Benchmarking consists of imposing the level of the benchmark series while minimizing the revisions of the observed movement in the sub-annual series as much as possible. The procedure also allows nonbinding benchmarking where the benchmark series can also be revised.

The procedure may also be used for benchmarking-related topics such as temporal distribution (disaggregation of the benchmark series into more frequent observations), calendarization (special case of temporal distribution) and linkage (connection of different time series segments into a consistent single time series).

---

## Procedure Syntax

**PROC BENCHMARKING** <*option(s)*>;

VAR var1 </alt1> <…varN </altN> >;

WITH var1 </alt1> <…varN </altN> >;

**BY** *variable(s)*;

| To do this | Use this statement |
|---|---|
| Perform benchmarking separately for each BY group | BY |
| Identify, as needed, the sub-annual series to benchmark along with their associated alterability coefficients | VAR |
| Identify, as needed, the benchmark series along with their associated alterability coefficients | WITH |

---

## PROC BENCHMARKING Statement

**PROC BENCHMARKING** <option(s)>;

| To do this | Use this option |
|---|---|
| Specify the input benchmarks data set | BENCHMARKS= |
| Specify a value for the bias | BIAS= |
| Specify the bias estimation option | BIASOPTION= |
| Specify the value for the $\lambda$ (lambda) parameter | LAMBDA= |
| Specify the output benchmarks data set | OUTBENCHMARKS= |
| Specify the output data set that will contain supplementary data useful to produce analytical tables and graphs | OUTGRAPHTABLE= |
| Specify the output benchmarked sub-annual series data set | OUTSERIES= |
| Specify the value for the $\rho$ (rho) parameter | RHO= |

| | |
|---|---|
| Specify the input sub-annual series data set | SERIES= |
| Specify the tolerance level for the ultimate test in percentage or absolute value | TOLERANCEPERCENT= \| TOLERANCEVALUE= |
| Specify the threshold for identification of negative values. | TOLNEGRESULT= |
| Print information about CPU time and print intermediate results | VERBOSE |
| Display a warning message in the log when a negative value is found (see option TOLNEGRESULT=) | WARNNEGRESULT \| NOWARNNEGRESULT |

# Options

**BENCHMARKS**=*SAS-data-set*

specifies the input SAS data set that contains the benchmarks. It is mandatory. The following numeric variables define the coverage period of each benchmark and must be in this data set: STARTYEAR, STARTPERIOD, ENDYEAR and ENDPERIOD. In addition, a numeric variable named VALUE containing the values of the benchmark must be present by default. To specify another variable name or more than one benchmark series, use the VAR and/or WITH statements.

**BIAS**=*real number*

specifies the value of the user-defined bias to be used for the correction of the sub-annual series prior to benchmarking. The bias is added to the sub-annual series with an additive model (when LAMBDA=0.0) while it is multiplied otherwise (when LAMBDA≠0.0). Option BIAS= is not mandatory. The default value is BIAS=0.0 when LAMBDA=0.0 and BIAS=1.0 otherwise. Note that option BIAS= is not used when option BIASOPTION=3 (see below).

**BIASOPTION**=*1, 2 or 3*

specifies the bias estimation option. It is mandatory.

| If you specify this bias estimation option value | Then the procedure |
|---|---|
| 1 | Will not estimate the bias parameter. The bias used to correct the sub-annual series will be the value specified with the BIAS= option or the default value (see option BIAS=). |
| 2 | Will estimate the bias parameter, print the result to the log, but will not use it. The bias used to correct the sub-annual series will be the value specified with the BIAS= option or the default value (see option BIAS=). |
| 3 | Will estimate the bias parameter, print the result to the log and apply the estimated bias correction. Any value specified with the BIAS= option will be printed to the log but not used. |

**LAMBDA**=*real number (suggested values are between -3 and 3)*

specifies the value of the adjustment model parameter $\lambda$. This option is mandatory. Typical values are LAMBDA=0.0 for an additive model and LAMBDA=1.0 for a proportional model. Prorating can be obtained by setting LAMBDA=0.5 and RHO=0.0.

**OUTBENCHMARKS**=*SAS-data-set*

names the output SAS data set that will contain the benchmarks used by the procedure. If it is not specified, PROC BENCHMARKING will create it by using the DATA*n* naming convention. The variables of this data set will have the same names as the ones in the BENCHMARKS= input data set. If BY variables are specified, the BY variables will also appear on this data set.

**OUTGRAPHTABLE**=*SAS-data-set*

names the output SAS data set that will contain supplementary data useful to produce analytical tables and graphs in the case of non-overlapping benchmarks. It is optional and is not created if not specified. The following table describes the variables that will be written to the data set. If BY variables are specified, they will also appear on this data set.

| Variable Name | Description |
|---|---|
| ALTBENCHMARKS | Name of the variable used to specify the benchmark series alterability coefficients |
| ALTBENCHMARKSVALUE | The benchmark series alterability coefficient values |
| ALTSERIES | Name of the variable used to specify the sub-annual series alterability coefficients |
| ALTSERIESVALUE | The sub-annual series alterability coefficient values |
| AVGBENCHMARK | Value of the benchmark divided by the number of covered periods; missing if there is no benchmark |
| AVGBENCHMARKSUBANNUALRATIO | Ratios** of the values of variables AVGBENCHMARK and AVGSUBANNUAL; missing if there is no benchmark |
| AVGSUBANNUAL | Values of the initial sub-annual series averaged over the benchmark coverage period; missing if there is no benchmark |
| BENCHMARKED | Benchmarked values of the sub-annual series |
| BENCHMARKEDSUBANNUALRATIO | Ratios** of the values of variables BENCHMARKED and SUBANNUAL |
| BIAS | The value of BIAS used (calculated by the procedure or supplied by user) |
| DATE | Character variable combining the values of variables YEAR and PERIOD |
| GROWTHRATEBENCHMARKED | Growth rates** in benchmarked sub-annual series. |
| GROWTHRATESUBANNUAL | Growth rates** in initial sub-annual series |
| LAMBDA | The value of the input parameter $\lambda$ (lambda) |
| M | The benchmark coverage period identifiers, m = 1, .., M; missing if there is no benchmark |
| PERIOD | Period values |
| PERIODICITY | The maximum number of periods in a year, for example 4 for a quarterly series |
| RHO | The value of the input parameter $\rho$ (rho) |
| SUBANNUAL | Sub-annual series initial values |
| SUBANNUALCORRECTED | Sub-annual series rescaled values ('corrected for bias') |
| T | The sub-annual period identifiers, t = 1, .., T |

| | |
|---|---|
| VARBENCHMARKS | Name of the variable used to specify the benchmark series |
| VARSERIES | Name of the variable used to specify the sub-annual series |
| YEAR | Year values |

** Terms "ratios" and "growth rates" refer here to a terminology commonly used in time series and actually mean "differences" and "first differences" respectively when the additive model is used (LAMBDA=0.0).

**OUTSERIES**=*SAS-data-set*

names the output SAS data set that will contain the benchmarked sub-annual series. If it is not specified, PROC BENCHMARKING will create it by using the DATA*n* naming convention. The variables of this data set will have the same names as the ones in the SERIES= input data set. If BY variables are specified, they will also appear on this data set.

**RHO**=*real number between 0 and 1 (inclusive)*

specifies the value of the autoregressive parameter $\rho$ in the [0,1] interval. It is mandatory.

**SERIES**=*SAS-data-set*

specifies the input SAS data set that contains the sub-annual series to benchmark. It is mandatory. The following numeric variables must be in this data set: YEAR and PERIOD. In addition, a numeric variable named VALUE containing the values of the sub-annual series to be processed must be present by default. To specify another variable name or more than one sub-annual series, use the VAR statement.

**TOLERANCEPERCENT=**
**TOLP**=*positive real number (including 0)*

**TOLERANCEVALUE=**
**TOLV**=*positive real number (including 0)*

specifies the tolerance, in percentage or absolute value, to be used when performing the ultimate test in the case of binding benchmarks (alterability coefficient = 0.0 for the benchmarks). The test compares the input annual benchmarks with benchmarks calculated from the benchmarked series. This number is optional.

**Default:** TOLERANCEVALUE=1E-3
**Example:** To set a tolerance of 1%, specify TOLERANCEPERCENT=0.01
To set a tolerance of 10, specify TOLERANCEVALUE=10

**TOLNEGRESULT=**
**TOLN=***strictly negative* r*eal number (excluding 0)*

specifies the threshold for identification of negative values. A value is considered negative when it is smaller than this threshold. This number is optional. The default value for this option is -1E -3.

**VERBOSE**

use this option to tell the procedure to print intermediate results and CPU time. With this option, intermediate matrices will be printed to the log as well as the CPU time taken to process them. (Inverse of matrices, multiplication of matrices, etc.).

**WARNNEGRESULT | NOWARNNEGRESULT**

specifies whether PROC BENCHMARKING displays a warning message in the log when a negative value created by the procedure is smaller than the threshold specified by the TOLNEGRESULT= option. The default option is WARNNEGRESULT.

---

# VAR Statement

**VAR** var1 </alt1> <…varN </altN> >;

## Required arguments

***var1…varN***

specifies the variable(s) containing the values of the sub-annual series to be benchmarked. The variables must be numeric. The VAR statement is optional. If not specified, the procedure will look for a variable called VALUE in the SERIES= input data set.

***alt1…altN***

specifies, as needed, the variable(s) containing the alterability coefficients that apply to the sub-annual series. The alterability coefficients variables should be numeric. If an alterability coefficients variable is not specified for a given sub-annual series, the procedure will use the default alterability coefficient value for sub-annual series, which is 1.0. Alterability coefficients only come into play after the initial sub-annual series has been corrected for the bias (when applicable). For example, this means that specifying an alterability coefficient of 0.0 for a given sub-annual series data point *will not* result in an unchanged value after benchmarking if bias correction is used (see options BIAS= and BIASOPTION=). Specification of alterability coefficients is not allowed when RHO=1.

---

# WITH Statement

**WITH** var1 </alt1> <…varN </altN> >;

## Required arguments

***var1…varN***

specifies the variable(s) that contain the values of the benchmarks. The variables must be numeric. The WITH statement is optional. If not specified, the procedure will look in the BENCHMARKS= input dataset for either the variable VALUE or variables corresponding to the names specified in the VAR statement.

***alt1…altN***

specifies, as needed, the variable(s) containing the alterability coefficients that apply to the benchmarks. The alterability coefficients variables should be numeric. If an alterability coefficients variable is not specified for a given benchmark, the procedure will use the default alterability coefficient value for a benchmark, which is 0.0. Specification of alterability coefficients is not allowed when RHO=1.

---

# BY Statement

**BY** *variable-1 <... variable-n>*;

### Required arguments

***variable(s)***

specifies the variables used by the procedure to form BY groups. The BY variables must be present in the two input data sets (BENCHMARKS= and SERIES=) and will appear in all three output data sets (OUTBENCHMARKS=, OUTGRAPHTABLE and OUTSERIES=). You can specify more than one variable. This statement is optional. The variables can be numeric or character.

---

# Details

- If a missing value appears in one of the variables of the BENCHMARKS= input data set (other than the BY variables), the observations with the missing values are dropped, a warning message is displayed in the log and the procedure runs.
- If a missing value appears in the YEAR and/or PERIOD variables of the SERIES= input data set and BY variables are specified, the corresponding BY group is skipped, a warning message is displayed in the log and the procedure moves on to the next BY group. If no BY variables are specified, a warning message is displayed in the log and no processing is done at all.
- If a missing value appears in a variable identifying a sub-annual series variables in the SERIES= input data set (see the VAR statement) and BY variables are specified, the corresponding BY group is skipped, a warning message is displayed in the log and the procedure moves on to the next BY group. If no BY variables are specified, the affected sub-annual series is not processed, a warning message is displayed in the log and the procedure moves on to the next sub-annual series (when applicable).
- The procedure does not allow proportional benchmarking (LAMBDA≠0) of a sub-annual series that is null over the entire coverage period of a non-null benchmark.
- With a value of RHO=1, only the default alterability coefficients (0 for a benchmark and 1 for a sub-annual series) are valid. The specification of alterability coefficients variables is therefore not allowed. If such variables are specified, the procedure ignores them and displays a warning message in the log.
- Alterability coefficients only come into play after the initial sub-annual series has been corrected for the bias (when applicable). For example, this means that specifying an alterability coefficient of 0 for a given sub-annual series data point *will not* result in an unchanged value after benchmarking if bias correction is used.
- Two-level names such as *libref.SAS-data-set* can be used to specify input data sets or to make output data sets permanent.
- If BY variables are specified, they will appear in all the output data sets.

# Examples

```
/* Example 1: Simple case with a single quarterly series to benchmark to annual values */

/* Sub-annual (quarterly) series */
data mySeries;
input year
      period
      value;
datalines;
1998 1 1.9
1998 2 2.4
1998 3 3.1
1998 4 2.2
1999 1 2.0
1999 2 2.6
1999 3 3.4
1999 4 2.4
2000 1 2.3
;

/* Annual benchmarks */
data myBenchmarks;
input startyear
      startperiod
      endyear
      endperiod
      value;
datalines;
1998 1 1998 4 10.3
1999 1 1999 4 10.2
;

/* Benchmarking using...
      - recommended RHO value for quarterly series (RHO=0.729)
      - proportional model (LAMBDA=1)
      - bias-corrected sub-annual series with the bias estimated by the procedure
        (BIASOPTION=3)
*/
proc benchmarking
   benchmarks=myBenchmarks
   series=mySeries
   outbenchmarks=outBenchmarks
   outseries=outSeries
   outgraphtable=outGraphTable
   rho=0.729
   lambda=1
   biasoption=3;
run;

proc print data=outseries;
run;



/* Example 2: Two quarterly series to benchmark to annual values,
              with by groups and user-defined alterability coefficients */

/* Sub-annual (quarterly) series where 1999 quarters 1 and 2 for BY group A
   of series van_sales_quarterly are non-alterable (see variable alt_van) */
```

```
data mySeries;
length group $1;
input group   /* BY group id */
      year
      period
      car_sales_quarterly
      van_sales_quarterly
      alt_van @@;   /* Alterability coefficients for van_sales_quarterly */
datalines;
A 1998 1 1851 1900 1    A 1998 2 2436 2200 1
A 1998 3 3115 3000 1    A 1998 4 2205 2000 1
A 1999 1 1987 1900 0    A 1999 2 2635 2500 0
A 1999 3 3435 3800 1    A 1999 4 2361 2500 1
A 2000 1 2183 2100 1    A 2000 2 2822 3100 1
A 2000 3 3664 3650 1    A 2000 4 2550 2950 1
A 2001 1 2342 3300 1    A 2001 2 3001 4000 1
A 2001 3 3779 3290 1    A 2001 4 2538 2600 1
A 2002 1 2363 2010 1    A 2002 2 3090 3600 1
A 2002 3 3807 3500 1    A 2002 4 2631 2100 1
A 2003 1 2601 2050 1    A 2003 2 3063 3500 1
A 2003 3 3961 4290 1    A 2003 4 2774 2800 1
A 2004 1 2476 2770 1    A 2004 2 3083 3080 1
A 2004 3 3864 3100 1    A 2004 4 2773 2800 1
A 2005 1 2489 3100 1    A 2005 2 3082 2860 1
B 1998 1 1851 1900 1    B 1998 2 2436 2200 1
B 1998 3 3115 3000 1    B 1998 4 2205 2000 1
B 1999 1 1987 1900 1    B 1999 2 2635 2500 1
B 1999 3 3435 3800 1    B 1999 4 2361 2500 1
B 2000 1 2183 2100 1    B 2000 2 2822 3100 1
B 2000 3 3664 3650 1    B 2000 4 2550 2950 1
B 2001 1 2342 3300 1    B 2001 2 3001 4000 1
B 2001 3 3779 3290 1    B 2001 4 2538 2600 1
B 2002 1 2363 2010 1    B 2002 2 3090 3600 1
B 2002 3 3807 3500 1    B 2002 4 2631 2100 1
B 2003 1 2601 2050 1    B 2003 2 3063 3500 1
B 2003 3 3961 4290 1    B 2003 4 2774 2800 1
B 2004 1 2476 2770 1    B 2004 2 3083 3080 1
B 2004 3 3864 3100 1    B 2004 4 2773 2800 1
B 2005 1 2489 3100 1    B 2005 2 3082 2860 1
;

/* Annual benchmarks */
data myBenchmarks;
length group $1;
input group    /* BY group id */
      startYear
      startPeriod
      endYear
      endPeriod
      car_sales_annual
      van_sales_annual @@;
datalines;
A 1998 1 1998 4 10324 12000    A 1999 1 1999 4 10200 10400
A 2000 1 2000 4 10582 11550    A 2001 1 2001 4 11097 11400
A 2002 1 2002 4 11582 14500    A 2003 1 2003 4 11092 16000
B 1998 1 1998 4 10324 12000    B 1999 1 1999 4 10200 10400
B 2000 1 2000 4 10582 11550    B 2001 1 2001 4 11097 11400
B 2002 1 2002 4 11582 14500    B 2003 1 2003 4 11092 16000
;
```

```
/* Benchmarking...
      - using recommended RHO value for quarterly series (RHO=0.729)
      - with proportional model (LAMBDA=1)
      - without bias correction (BIASOPTION=1 and option BIAS= not specified)
*/
proc benchmarking
    benchmarks=myBenchmarks
    series=mySeries
    outbenchmarks=outBenchmarks
    outseries=outSeries
    outgraphtable=outGraphTable
    rho=0.729
    lambda=1
    biasoption=1;

    var car_sales_quarterly van_sales_quarterly/alt_van;
    with car_sales_annual van_sales_annual;
    by group;
run;
```

---

# Notes

This document is a guide for the use of the BENCHMARKING procedure. PROC BENCHMARKING is part of Statistics Canada's G-Series software formerly known as Forillon.

For more information, please contact the G-Series support team using the G-Series e-mail address (G-Series@statcan.gc.ca) or consult the web site at G-Series (Statistics Canada intranet only).

---

# References


Bloem, A. M., R. J. Dippelsman, and N. Ø. Mæhle (2001). **Quarterly National Accounts Manual, Concepts, Data Sources and Compilation**, International Monetary Fund, Washington DC.

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

Fortier, S. and B. Quenneville (2007). "Theory and Application of Benchmarking in Business Surveys", **ICES III** proceedings, Introductory Overview Lecture at the International Conference on Establishment Surveys III, June 2007.

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

Quenneville, B., S. Fortier, Z.-G. Chen and E. Latendresse (2006). "Recent Developments in Benchmarking to Annual Totals in X-12-ARIMA and at Statistics Canada", proceedings of the **2006 Eurostat Conference on Seasonality, Seasonal Adjustment and Their Implications for Short-Term Analysis and Forecasting**, Luxembourg, May 2006.


---

# The TSRAKING Procedure

## Overview

This procedure will restore cross-sectional aggregation constraints in a system of time series. The aggregation constraints may come from a 1 or 2-dimensional table. Optionally, temporal constraints can also be preserved.

Procedure Syntax

**PROC TSRAKING** <option(s)>;

**ID** *variable(s);*

| To do this | Use this statement |
|---|---|
| Identify the variables of the input data set to be transferred to the output data set | ID |

---

## PROC TSRAKING Statement

**PROC TSRAKING** <option(s)>;

| To do this | Use this option |
|---|---|
| Specify the alterability coefficients data set | ALTERABILITY= |
| Specify the default alterability coefficient for the component series annual totals (temporal constraints) | ALTERANNUAL= |
| Specify the default alterability coefficient for the SERIES variables (component series) | ALTERSERIES= |
| Specify the default alterability coefficient for the TOTAL1 variables (1st dimension control totals) | ALTERTOTAL1= |
| Specify the default alterability coefficient for the TOTAL2 variables (2nd dimension control totals) | ALTERTOTAL2= |
| Specify the input data set | DATA= |
| Specify the metadata data set | METADATA= |
| Specify the output data set | OUT= |
| Specify the tolerance level for the ultimate test in percentage or absolute value | TOLERANCEPERCENT= \| TOLERANCEVALUE= |
| Specify the threshold for identification of negative values. | TOLNEGRESULT= |
| Display intermediate results and information about CPU time in the log | VERBOSE |
| Display a warning message in the log when a negative value is found (see option TOLNEGRESULT=) | WARNNEGRESULT \| NOWARNNEGRESULT |

# Options

**ALTERABILITY**=*SAS-data-set*

specifies the SAS data set that contains the alterability coefficients variables. Any alterability coefficients variable must correspond to a component series or a control total, that is, a variable with the same name must be present in the DATA= input data set. If the optional ALTERABILITY= data set is used, the values of these alterability coefficients will override the values specified using the ALTERxxx options (see below). When the DATA= data set contains several observations and the ALTERABILITY= data set contains only one, the alterability coefficients are used (repeated) for all observations of the DATA= data set. Alternatively, the ALTERABILITY= data set may contain as many observations as the DATA= data set.

**ALTERANNUAL=*positive*** *real number (including 0)*

specifies the default alterability coefficient for the component series temporal constraints (annual totals). It is optional and has a default value of 0.0 (binding totals). It will apply to temporal constraints for which an alterability coefficient has not already been specified using variable ALTERANNUAL of the METADATA= data set.

**ALTERSERIES=*positive*** *real number (including 0)*

specifies the default alterability coefficient for the component series. It is optional and has a default value of 1.0. It will apply to component series for which an alterability coefficient has not already been specified using the ALTERABILITY= data set.

**ALTERTOTAL1=*positive*** *real number (including 0)*

specifies the default alterability coefficient for the 1st dimension control totals. It is optional and has a default value of 0.0 (binding totals). It will apply to control totals for which an alterability coefficient has not already been specified using the ALTERABILITY= data set.

**ALTERTOTAL2=*positive*** *real number (including 0)*

specifies the default alterability coefficient for the 2nd dimension control totals. It is optional and has a default value of 0.0 (binding totals). It will apply to control totals for which an alterability coefficient has not already been specified using the ALTERABILITY= data set.

**DATA**=*SAS-data-set*

specifies the SAS data set that contains the variables representing the time series system, that is the component series and cross-sectional control totals to be raked. If more than one observation is provided, the sum of the provided values will also be preserved as part of implicit temporal constraints. The DATA= data set is mandatory.

**METADATA**=*SAS-data-set*

specifies the metadata SAS data set that describes the aggregation constraints (additivity rules) used by the procedure. The METADATA= data set is mandatory. Two character variables must be in this data set: SERIES and TOTAL1. Two variables are optional: TOTAL2 (character) and ALTERANNUAL (numeric). The values of variable SERIES represent the variable names of the component series in the

DATA= input data set. Similarly, the values of variables TOTAL1 and TOTAL2 represent the variable names of the 1st and 2nd dimensions control totals in the DATA= input data set. Variable ALTERANNUAL contains the alterability coefficient for the temporal constraint associated to the component series.

**OUT**=*SAS-data-set*

specifies the output SAS data set that contains all the raked component series, raked cross-sectional control totals and any variables specified in the ID statement. It is optional. If it is not specified, PROC TSRAKING will create it by using the DATA*n* naming convention.

**TOLERANCEPERCENT=**
**TOLP**=*positive real number (including 0)*

**TOLERANCEVALUE=**
**TOLV**=*positive real number (including 0)*

specifies the tolerance, in percentage or absolute value, to be used when performing the ultimate test in the case of binding totals (alterability coefficient = 0.0 for the control totals). The test compares the binding control totals with the ones calculated from raked component series. This number is optional.

**Default:** TOLERANCEVALUE=1E-3.
**Example:** To set a tolerance of 1%, specify TOLERANCEPERCENT=0.01
To set a tolerance of 10, specify TOLERANCEVALUE=10

**TOLNEGRESULT=**
**TOLN=***strictly negative real number (excluding 0)*

specifies the threshold for identification of negative values. A value is considered negative when it is smaller than this threshold. This number is optional. The default value for this option is -1E -3.

**VERBOSE**

displays intermediate results and information about CPU time in the log.

**WARNNEGRESULT | NOWARNNEGRESULT**

specifies whether PROC TSRAKING displays a warning message in the log when a negative value created by the procedure is smaller than the threshold specified by the TOLNEGRESULT= option. The default option is WARNNEGRESULT.

---

# ID Statement

**ID** *variable*-1 ... *variable-n*;

Use the ID statement to name additional *variables* to be transferred from the DATA= input data set to the OUT= output data set. By default, the output data set only contains the variables described in the metadata (see option METADATA=). *Variables* listed in the ID statement must be present in the DATA= input data set.

---

# Details

- With the exception of the variables listed in the ID statement, *missing values in the DATA= input data set* will stop the procedure.
- Missing and/or *negative values in the ALTERABILITY= input data set* will stop the procedure.
- Two-level names such as *libref.SAS-data-set* can be used to specify input data sets or to make the output data set permanent.

---

# Examples

```
/* EXAMPLE 1  */

/* In this 1-dimensional example, the values of "cars" and "vans" must sum up to the
value of "total"  */

data myMetadata;
input series $5.
      Total1 $5.;
datalines;
cars total
vans total
;

data myData;
input cars
      vans
      total;
datalines;
25 5 40
;

proc tsraking
   metadata=myMetadata
   data=myData
   out=outData;
run;


/* EXAMPLE 2  */

/* In this 2-dimensional example, we have estimates of "cars" and "vans" sales for the
"alb", "sask" and "man" regions.  The sum of the 3 regions for "cars" must be equal to
its control total (cars_alb + cars_sask + cars_man = cars_total).  The sum of the 3
regions for "vans" must be equal to its control total (vans_alb + vans_sask + vans_man =
vans_total).  Also, in each of the regions, the sum of "cars" and "vans" must be equal to
the regional control total (cars_alb + vans_alb = alb_total;  cars_sask + vans_sask =
sask_total and cars_man + vans_man = man_total).  Finally, "vans" sales in region "sask"
(vans_sask) are non-alterable (alterability coefficient = 0) */

data myMetadata;
input series $9.
      Total1 $11.
      Total2 $11.;
datalines;
cars_alb  cars_total alb_total
cars_sask cars_total sask_total
cars_man  cars_total man_total
```

```
vans_alb  vans_total alb_total
vans_sask vans_total sask_total
vans_man  vans_total man_total
;

data myData;
input cars_alb cars_sask cars_man vans_alb vans_sask vans_man
      alb_total sask_total man_total cars_total vans_total;
datalines;
12 14 13 20 20 24 30 31 32 40 53
;

data myalter;
input cars_alb cars_sask cars_man vans_alb vans_sask vans_man
      alb_total sask_total man_total cars_total vans_total;
datalines;
1 1 1 1 0 1 0 0 0 0 0
;

proc tsraking
   metadata=myMetadata
   data=myData
   alterability=myAlter
   out=outData;
run;


/* EXAMPLE 3  */

/* In this 2-dimensional example with temporal constraints, we have the constraints A_1 +
A_2 + A_3 = A; B_1 + B_2 + B_3 = B; A_1 + B_1 = _1; A_2 + B_2 = _2 and A_3 + B_3 = _3.
We have 4 (quarterly) observations of each value and we want to preserve annual totals.
For each variable, the sum of the 4 quarterly values must stay the same after the
reconciliation step. */

data mymetadata;
input series $3.
      total1 $3.
      total2 $3.;
datalines;
A_1 _1 A
B_1 _1 B
A_2 _2 A
B_2 _2 B
A_3 _3 A
B_3 _3 B
;

data mydata;
input A_1 A_2 A_3 B_1 B_2 B_3 A B _1 _2 _3;
datalines;
12 14 13 20 20 24 40 53 30 31 32
10  9 15 21 29 20 25 80 35 35 35
12  8 17 15 20 30 40 59 23 32 44
 9  9 14 17 24 23 37 71 28 35 45
;

proc tsraking
   metadata=myMetadata
   data=myData
   out=outData;
run;
```

# Notes

This document is a guide for the use of the TSRAKING procedure. PROC TSRAKING is part of Statistics Canada's G-Series software formerly known as Forillon.

For more information, please contact the G-Series support team using the G-Series e-mail address (G-Series@statcan.gc.ca) or consult the web site at G-Series (Statistics Canada intranet only).

# References


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

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

Fortier, S. and B. Quenneville (2009). "Reconciliation and Balancing of Accounts and Time Series", **JSM 2009** proceedings, Business and Economic Section. Alexandria, VA: American Statistical Association.

# The GSeriesTSBalancing Macro

## Overview

This macro balances (reconciles) a system of time series according to a set of linear constraints. The balancing solution is obtained by solving a quadratic optimization problem (see the appendix) with the OPTMODEL procedure. Given the feasibility of the balancing problem, the result is a time series data set in which the specified constraints are respected for every time period. Linear equality and inequality constraints are allowed. Optionally, the preservation of temporal totals may also be specified.

This macro is meant to be used with SAS version 9.3 or later and requires SAS/OR.

---

## Macro Syntax

The complete set of available parameters is provided below, along with their default value when applicable.

```
%GSeriesTSBalancing
(
   /* Mandatory Parameters */

   inTS                      = ,
   inProblemSpecs            = ,
   outTS                     = ,

   /* Optional Parameters */

   timeVarName               = ,
   periodInterval            = ,
   temporalGrpInterval       = ,

   optModelSolverOptions     = WITH QP,
   displayLevel              = 2,

   outOptModelSummary        = ,
   outDetailedResults        = ,
   outEvaluatedConstraints   = ,
   outTemporalTotals         = ,
   outProcessingGrps         = ,

   alterPos                  = 1,
   alterNeg                  = 1,
   alterMix                  = 1,
   alterTemporalTotal        = 0,

   lowerBound                = ,
   upperBound                = ,

   toleranceValue            = ,
   toleranceValueTemporal    = ,
   tolerancePercentTemporal  = ,

   TSFormat                  = WIDE,
   tallTSIDVarName           = _NAME_,
   tallTSValueVarName        = _VALUE_,
   tallTSAlterVarName        = ,

   language                  = EN
)
```

# Parameters

The following table gives a brief description of all macro parameters.

| Parameter Name | Short Description |
|---|---|
| inTS | Mandatory. Input data set. |
| inProblemSpecs | Mandatory. Problem specification data set. |
| outTS | Mandatory. Output (balanced) data set. |
| timeVarName | Time variable (SAS date, time or datetime numeric variable) in the input data set. |
| periodInterval | SAS date, time or datetime interval associated with the periods in the input data set. |
| temporalGrpInterval | SAS date, time or datetime interval associated with the temporal totals to be preserved. |
| optModelSolverOptions | Solver options for the OPTMODEL procedure. |
| displayLevel | Level of information to be 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. |
| alterTemporalTotals | Temporal total alterability coefficients. |
| lowerBound | Lower bound for the time series values. |
| upperBound | Upper bound for the time series values. |
| toleranceValue | Tolerance 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. |
| tallTSIDVarName | Input data set variable that contains the time series names when parameter TSFormat=TALL. |
| tallTSValueVarName | Input data set variable that contains the time series values when parameter TSFormat=TALL. |
| tallTSAlterVarName | Input data set variable that contains the time series alterability coefficients when parameter TSFormat=TALL. |
| language | Display language for messages in the Log. |

The following section provides more details on all parameters, listed in alphabetical order.

alterMix = *positive real number (including 0)*

Alterability coefficient associated to the values of time series with a mix of positive and negative coefficients in the balancing constraints in which they are involved. Alterability coefficients provided in the problem specification data set override this value.

This parameter is optional and the default value is 1.0.

`alterNeg` = *positive real number (including 0*

Alterability coefficient associated to the values of time series with negative coefficients in all balancing constraints in which they are involved (e.g. marginal totals in raking problems). Alterability coefficients provided in the problem specification data set override this value.

This parameter is optional and the default value is 1.0.

`alterPos` = *positive real number (including 0)*

Alterability coefficient associated to the values of time series with positive coefficients in all balancing constraints in which they are involved (e.g. component series in raking problems). Alterability coefficients provided in the problem specification data set override this value.

This parameter is optional and has a default value of 1.0.

`alterTemporalTotals` = *positive real number (including 0)*

Alterability coefficient associated to the time series temporal totals. Alterability coefficients provided in the problem specification data set override this value.

This parameter is optional and the default value is 0.0.

`displayLevel` = `0`, `1`, `2`, `3`, `4` *or* `5`

Level of information to be displayed in the Log and active ODS destinations (e.g. Output window).

| Displayed information | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Macro header and completion message in the Log | √ | √ | √ | √ | √ | √ |
| Error and Warning messages in the Log | √ | √ | √ | √ | √ | √ |
| Content of macro variable `_OROPTMODEL_` in the Log | | √ | | | | √ |
| Notes from OPTMODEL in the Log | | | √ | √ | √ | √ |
| The following OPTMODEL ODS tables in the active ODS destinations:<br>• "Problem Summary"<br>• "Solution Summary" | | | √ | √ | √ | √ |
| The following OPTMODEL ODS tables in the active ODS destinations:<br>• "Methods of Derivative Computation" (NLP solver)<br>• "Solver Options"<br>• "Optimization Statistics"<br>• other solver-specific ODS tables (when applicable) | | | | √ | √ | √ |
| Balancing results (balanced time series) in the active ODS destinations | | | | | √ | √ |
| OPTMODEL expanded problem (OPTMODEL EXPAND statement) in the active ODS destinations | | | | | | √ |

This parameter is optional and the default value is `2`.

inProblemSpecs = *SAS-data-set*

Name of problem specification data set. The purpose of this file is to specify the linear constraints (balancing rules) that define the relationships that need to be restored and, optionally, series-specific alterability coefficients and lower/upper bounds that would take precedence over the generic values defined with parameters alterPos, alterNeg, alterMix, alterTemporalTotals, lowerBound and upperBound.

The information is provided using 4 mandatory variables (_type_, _col_, _row_ and _coef_) and one optional variable (_timeVal_). There are 2 types of records in the problem specification data set:

- Label definition records (when _type_ is not blank)
- Information specification records (when _type_ is blank)

| Variable | Type | Details – *Label Definition Records* | Details – *Information Specification Records* |
|---|---|---|---|
| _type_ | Char | Reserved keyword that tells the macro how to interpret *Information Specification Records*:<br>EQ: equality (=) constraint<br>LE: lower or equal (≤) inequality constraint<br>GE: greater or equal (≥) inequality constraint<br>lowerBd: time series lower bound<br>upperBd: time series upper bound<br>alter: time series regular alter. coef.<br>alterTmp: time series temp. tot. alter. coef. | Not applicable (*blank* value) |
| _col_ | Char | Not applicable (*blank* value) | Name of a time series or reserved word _rhs_ to specify a right-hand-side constraint bound (constant) different than 0. |
| _row_ | Char | Label to be associated to the "type keyword" | Label associated with a "type keyword" identifying the type of information being specified. |
| _coef_ | Num | Not applicable (*missing* value) | Constraint bound (_rhs_) or time series information, e.g.<br>• constraint coefficient (default is 0)<br>• lower bound (default is given by parm. lowerBound)<br>• upper bound (default is given by parm. upperBound)<br>• regular alterability coefficient (default is given by parameters alterPos, alterNeg and alterMix)<br>• temporal total alterability coefficient (default is given by parameter alterTemporalTotals) |
| _timeVal_ | Num | Not applicable (*missing* value) | 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 (within the time series) in the input data set. |

The information provided in the problem specification data set is not case-sensitive.

This parameter is mandatory.

inTS = *SAS-data-set*

Name of the input data set. This data set contains the data for the time series involved in the balancing problem. It can also contain a date, time or datetime variable for preserving temporal totals (see parameter timeVarName). Data can be provided in a *wide* format (default) or in a *tall* format (see parameter TSFormat for more details). In *tall* mode, parameters tallTSIDVarName, tallTSValueVarName and tallTSAlterVarName identify the role of the variables in the input data set.

This parameter is mandatory.

`language` = `EN` *or* `FR`

Display language for messages in the Log. The default language is English (`EN`).

`lowerBound` = *real number*

Lower bound for the time series values. If not specified, the lower bound is -∞. Lower bounds provided in the problem specification data set override this value.

This parameter is optional and does not have a default value.

`optModelSolverOptions` = *character string*

Solver options for the OPTMODEL procedure. Two solvers are appropriate for time series balancing problems: QP (quadratic programming) and NLP (nonlinear programming), each having their own specific options (see the SAS documentation). The QP solver is much faster than the NLP solver and is therefore the recommended (and default) solver for this optimization problem. If keyword `NONE` is specified, the solver choice is left to the OPTMODEL procedure, which will result in using the NLP solver (with default options) for this problem in SAS version 9.3.

Valid values for this parameter in the context of this macro are:

| | |
|---|---|
| QP solver: | `optModelSolverOptions = WITH QP </ options>` |
| NLP solver: | `optModelSolverOptions = WITH NLP </ options>` |
| Let OPTMODEL decide: | `optModelSolverOptions = NONE` |

This macro is meant to be used with SAS version 9.3 or later and requires SAS/OR. It is not compatible with earlier versions of SAS. This parameter is optional and the default value is `WITH QP`.

`outDetailedResults` = *SAS-data-set*

Name of the *detailed balancing results* output data set. This data set contains detailed information on the time series involved in the balancing problem. It is created only if this parameter is specified.

| Variable Name | Description |
|---|---|
| `_NAME_`[1] | Time series name. |
| `_T_` | Period counter used internally by the macro. |
| `timeVarName` parameter value | Time variable (when specified). |
| `_VALUE_IN_` | Value before balancing. |
| `_ALTER_` | Alterability coefficient. |
| `_VALUE_OUT_` | Value after balancing. |
| `_DIF_` | Difference (`_VALUE_OUT_ - _VALUE_IN_`) |
| `_RDIF_` | Relative difference (`_DIF_ / _VALUE_IN_`); missing when `_VALUE_IN_ = 0`. |

This parameter is optional and does not have a default value.

[1] In *tall* mode (parameter TSFormat=TALL), the name of this variable is based on the value of parameter tallTSIDVarName.

`outEvaluatedConstraints` = *SAS-data-set*

Name of the *evaluated constraints* output data set. This data set contains detailed information on the specified balancing constraints. E.g. it shows how close to the bounds of inequality constraints the optimal solution is. It may also be useful in identifying constraints that are not respected in non-optimal solutions (e.g. infeasible complex balancing problems). The data set is created only if this parameter is specified.

| Variable Name | Description |
|---|---|
| `_CON_` | Constraint label. |
| `_T_` | Period counter used internally by the macro. |
| `timeVarName` parameter value | Time variable (when specified). |
| `_TYPE_` | Constraint type (`EQ`, `LE` or `GE`). |
| `_RHS_` | Constraint right-hand-side constant (constraint bound). |
| `_LHS_IN_` | Constraint left-hand-side value evaluated before balancing. |
| `_LHS_OUT_` | Constraint left-hand-side value evaluated after balancing. |
| `_DIF_` | Difference (`_LHS_OUT_ - _LHS_IN_`) |
| `_RDIF_` | Relative difference (`_DIF_ / _LHS_IN_`); missing when `_LHS_IN_ = 0`. |

This parameter is optional and does not have a default value.

`outOptModelSummary` = *SAS-data-set*

Name of the *OPTMODEL summary* output data set. This data set contains the information included in the `_OROPMODEL_` macro variable for each processing group (solution status, objective function value, number of iterations, execution time, etc.), thus providing a summary of the OPTMODEL execution by processing group. The file is created only if this parameter is specified.

| Variable Name | Description |
|---|---|
| `_PROC_GRP_` | Processing group counter used internally by the macro.[2] |
| `_PROC_GRP_LABEL_` | Processing group label providing a more meaningful description.[2] |
| $term_1$ | 1st term of macro variable `_OROPTMODEL_`. |
| $term_2$ | 2nd term of macro variable `_OROPTMODEL_`. |
| ... | … |
| $term_k$ | $k^{th}$ term of macro variable `_OROPTMODEL_`. |

Consult the OPTMODEL procedure documentation for the actual terms included in the `_OROPTMODEL_` macro variable for each solver.

This parameter is optional and does not have a default value.

[2] When preserving temporal totals, periods in a complete temporal group (e.g. calendar years that contain all 12 months) share the same processing group value with a label corresponding to the formatted temporal group value (e.g. 2011) while periods in an incomplete temporal group (e.g. calendar years that contain fewer than 12 months) have different values (e.g. one value per month) with labels corresponding to formatted period values (e.g. 2011-01, 2011-02, etc.). Otherwise, when temporal totals are not preserved, all periods have a distinct processing group value with a label corresponding to the formatted period value. The label variable (`_PROC_GRP_LABEL_`) is not included in the data set when parameter `temporalGrpInterval=_ALL_` or when a time variable (parameter `timeVarName`) is not specified.

outProcessingGrps = *SAS-data-set*

Name of the *processing groups* output data set. This data set contains detailed information on the periods and processing groups that were involved in the balancing problem. It is created only if this parameter is specified.

| Variable Name | Description |
|---|---|
| _T_ | Period counter used internally by the macro. |
| timeVarName parameter value | Time variable (when specified). |
| _PROC_GRP_ | Processing group counter used internally by the macro.[2] |
| _PROC_GRP_LABEL_ | Processing group label providing a more meaningful description.[2] |

This parameter is optional and does not have a default value.

outTemporalTotals = *SAS-data-set*

Name of the *temporal totals* output data set. This data set contains detailed information on the temporal totals that were involved in the balancing problem, i.e. the time series values cumulated over each complete temporal group (e.g. calendar years that contain all 12 months). It is empty when temporal totals are not preserved. See parameter temporalGrpInterval for more details on temporal total preservation. The data set is created only if this parameter is specified.

| Variable Name | Description |
|---|---|
| _NAME_[1] | Time series name. |
| _PROC_GRP_ | Processing group counter used internally by the macro.[2] |
| _PROC_GRP_LABEL_ | Processing group label providing a more meaningful description.[2] |
| _TOTAL_IN_ | Temporal total before balancing. |
| _ALTERTMP_ | Temporal total alterability coefficient. |
| _TOTAL_OUT_ | Temporal total after balancing. |
| _DIF_ | Difference (_TOTAL_OUT_ - _TOTAL_IN_) |
| _RDIF_ | Relative difference (_DIF_ / _TOTAL_IN_); missing when _TOTAL_IN_ = 0. |

This parameter is optional and does not have a default value.

outTS = *SAS-data-set*

Name of the output balanced data set. This data set is identical to the input data set, except that the initial (unbalanced) time series values have been replaced with the resulting balanced values. It is the main output and is always created by the macro.

This parameter is mandatory.

periodInterval = *SAS date, time or datetime interval*

SAS date, time or datetime interval associated with the periods in the input data set; requires parameter timeVarName to be specified. If this parameter is not specified but a time variable (parameter timeVarName) is specified, the macro automatically determines the period interval based on the first 3 (non missing) periods of the input data set using the INTGET function in SAS.

See the "Details" section for more information on the concept of time in the **GSeriesTSBalancing** macro.

This parameter is optional and does not have a default value.

`tallTSAlterVarName` = *SAS-variable-name*

Name of the input data set variable that contains the time series alterability coefficients in *tall* mode (parameter `TSFormat=TALL`). This parameter is ignored in *wide* mode (parameter `TSFormat=WIDE`). Non-missing alterability coefficients provided in the input data set override the alterability coefficients specified with parameters `alterPos`, `alterNeg` and `alterMix` as well as any alterability coefficient provided in the problem specification data set

This parameter is optional and does not have a default value (no alter. coef. variable by default).

`tallTSIDVarName` = *SAS-variable-name*

Name of the input data set variable that contains the names of the time series in *tall* mode (parameter `TSFormat=TALL`). This parameter is ignored in *wide* mode (parameter `TSFormat=WIDE`).

This parameter is optional and the default value is `_NAME_`.

`tallTSValueVarName` = *SAS-variable-name*

Name of the input data set variable that contains the values of the time series in *tall* mode (parameter `TSFormat=TALL`). This parameter is ignored in *wide* mode (parameter `TSFormat=WIDE`).

This parameter is optional and the default value is `_VALUE_`.

`temporalGrpInterval` = *SAS date, time or datetime interval or reserved word* `_ALL_`

SAS date, time or datetime interval associated with the temporal totals to be preserved (set of periods to be processed "at once"). Note:

- If this parameter is not specified, temporal totals are not preserved and each period of the input data set is balanced independently (period by period processing).
- Reserved word `_ALL_` can be specified to preserve temporal totals corresponding to the sum of the values over all periods of the input data set (process all periods "at once"); value `_ALL_` can also be used for problems without a time variable.
- Temporal totals are preserved only if the input data set contains more than one period and the frequency of the temporal group interval is lower than the frequency of the period interval (e.g. `periodInterval=MONTH` and `temporalGrpInterval=YEAR`). Otherwise, if the input data set contains a single period or the frequency of the temporal group interval is higher or the same as the frequency of the period interval (e.g. `periodInterval=QTR` and `temporalGrpInterval=MONTH` or `=QTR`), the concept of temporal total does not apply and period by period processing is automatically implemented.

See the "Details" section for more information on the concept of time in the **GSeriesTSBalancing** macro.

This parameter is optional and does not have a default value (no temporal total preservation by default).

`timeVarName` = *SAS-variable-name*

Name of the time variable (SAS date, time or datetime numeric variable) in the input data set. If a time variable is not specified, each period of the input data set is balanced independently, unless parameter `temporalGrpInterval=_ALL_` (see the description of parameter `temporalGrpInterval`).

See the "Details" section for more information on the concept of time in the **GSeriesTSBalancing** macro.

This parameter is optional and does not have a default value (no time variable by default).

`toleranceValue` = *positive real number*

Tolerance for the balancing constraints. When a tolerance is specified,

| | | | |
|---|---|---|---|
| EQ constraints | $G\theta = b$ | become | $b - \texttt{toleranceValue} \le G\theta \le b + \texttt{toleranceValue}$ |
| GE constraints | $G\theta \ge b$ | become | $G\theta \ge b - \texttt{toleranceValue}$ |
| LE constraints | $G\theta \le b$ | become | $G\theta \le b + \texttt{toleranceValue}$ |

This parameter is optional and does not have a default value (the default is no tolerance).

`toleranceValueTemporal` = *positive real number*
`tolerancePercentTemporal` = *positive real number*

Tolerance for the implicit temporal constraints corresponding to binding temporal totals specified either in absolute value or in percentage. Both `toleranceValueTemporal` and `tolerancePercentTemporal` cannot be specified at the same time and do not apply to (are ineffective for) non-binding temporal totals.

These parameters modify the implicit temporal constraints $\sum_{t \in T_j} \theta_{kt}^{(x)} = a_{kj}$ as follows, where initial temporal total $a_{kj}$ is binding $\left(\left|c_{kj}^{(a)} a_{kj}\right| = 0\right)$ and *j* designates a complete temporal group:

$$a_{kj} - \texttt{toleranceValueTemporal} \le \sum_{t \in T} \theta_{kt}^{(x)} \le a_{kj} + \texttt{toleranceValueTemporal}$$

or

$$a_{kj}\left(1 - \texttt{tolerancePercentTemporal}\right) \le \sum_{t \in T_j} \theta_{kt}^{(x)} \le a_{kj}\left(1 + \texttt{tolerancePercentTemporal}\right)$$

These parameters are optional and do not have a default value (the default is no tolerance).

`TSFormat` = `WIDE` *or* `TALL`

Defines the format of the input and output data sets (parameters `inTS` and `outTS`). The `WIDE` format means that time series data appear in separate columns (variables) in the data set (one column per time series). With the `TALL` format, the time series data are stacked in a single column (variable). Parameters `tallTSIDVarName`, `tallTSValueVarName` and `tallTSAlterVarName` identify the role of the variables in the input data set when the `TALL` format is specified.

This parameter is optional and the default value is `WIDE`.

`upperBound` = *real number*

Upper bound for the time series values. If not specified, the upper bound is ∞. Upper bounds provided in the problem specification data set override this value.

This parameter is optional and does not have a default value.

---

# Details

- The information provided as input to the macro is not case sensitive. This applies to the macro parameter values and to the information provided in the problem specification data set and the input data set (e.g. time series name in *tall* mode).

- Precedence rules for the different ways to specify alterability coefficients and lower/upper bounds are listed in the following table, higher precedence appearing first:

| | |
|---|---|
| 1 | Non-missing alterability coefficients provided in the data file in *tall* mode (see parameter `tallTSAlterVarName`). |
| 2 | Dated (non-missing `_timeVal_`) alterability coefficients and lower/upper bounds specified in the problem specification data set. |
| 3 | Undated (missing `_timeVal_`) alterability coefficients and lower/upper bounds specified in the problem specification data set. |
| 4 | Alterability coefficients and lower/upper bounds defined with parameters `alterPos`, `alterNeg`, `alterMix`, `alterTemporalTotals`, `lowerBound` and `upperBound`. |

- Time is a key concept in the **GSeriesTSBalancing** macro, namely for preserving temporal totals. Fortunately the concept of time is well incorporated in SAS and the macro takes advantage of the versatile and powerful set of date and time intervals and functions available in SAS. Three types of elements related to time are used in the **GSeriesTSBalancing** macro:

  - ***Periods***: high frequency time values that correspond to the periodicity of the time series to be balanced (e.g. months, quarters). They correspond to observations (rows) of the input data set in *wide* mode.

  - ***Temporal groups***: low frequency time values associated to temporal totals to be preserved (e.g. calendar years). A given temporal group is said to be *complete* when the provided data contain all periods that belong to the temporal group (e.g. calendar years for which all 12 months are present in the input data set) or *incomplete* otherwise (e.g. current calendar year with monthly data up to October, i.e. with only 10 months instead of 12). The *cardinality* of a complete temporal group is defined as the number of periods it contains (e.g. 12 and 4 for calendar years for monthly and quarterly data respectively). ***Temporal totals*** are defined for all time series and for all *complete temporal groups* involved in the balancing problem. A given temporal total corresponds to the sum of the values of a given time series over the periods of a given complete temporal group.

  - ***Processing groups***: sets of periods that need to be processed (optimized) at once by the OPTMODEL procedure. A processing group either corresponds to a single period (period by period processing) or to the set of periods that belong to a complete temporal group when preserving temporal totals.

Handling of time in the **GSeriesTSBalancing** macro is specified with 3 optional parameters: `timeVarName`, `periodInterval`, and `temporalGrpInterval`. The following table summarizes the possible scenarios when the input data set contains more than 1 period.

| The problem includes a time variable | Temporal group interval (`temporalGrpInterval` value) | Time processing scenario | Number of processing groups |
|---|---|---|---|
| No | *Not specified* | Period by period | Number of periods |
| | `_ALL_` | All periods at once (temporal total preservation) | 1 |
| Yes | *Not specified* | Period by period | Number of periods |
| | `_ALL_` | All periods at once (temporal total preservation) | 1 |
| | Lower frequency than `periodInterval` (e.g. `periodInterval=MONTH temporalGrpInterval=YEAR`) | • Within complete temporal groups: all periods at once (temporal total preservation)<br>• Within incomplete temporal groups: period by period | Number of complete temporal groups + Number of periods in incomplete temporal groups |
| | Higher frequency than `periodInterval` (e.g. `periodInterval=QTR temporalGrpInterval=MONTH`) | The specified intervals are not compatible with the concept of temporal total preservation, resulting in period by period processing. | Number of periods |
| | Same as `periodInterval` | | |
| | Incompatible with `periodInterval` (e.g. `periodInterval=WEEK temporalGrpInterval=MONTH`) | | |

Consult the SAS documentation for the complete list of available date, time and datetime intervals.

- After calls to the **GSeriesTSBalancing** macro, the _GSERIESRC_ macro variable can be used to verify if the execution was successful (0) or not (VA###), where VA### refers to a G-Series error message.

---

# Comparison with PROC TSRaking

- **PROC TSRaking** is limited to one- and two-dimensional raking problems (with temporal total preservation if required) while the **GSeriesTSBalancing** macro 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 both **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 **PROC TSRaking** and separate calls for each processing group must be submitted.

- The **GSeriesTSBalancing** macro accommodates the specification of sparse problems in their reduced form. This is not the case of **PROC 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 will differ if some data points are negative.

- The **GSeriesTSBalancing** macro is usually faster than **PROC TSRaking** but is generally more sensitive to small inconsistencies that sometime occur in fully specified (over-specified) multi-dimensional raking problems with temporal total preservation.

---

# Example 1

# One-dimensional raking problem

In this example, we have estimates of "cars" sales for the "alb", "sask" and "man" regions as well as total sales, for the first three quarters of 2011 (no annual total preservation). Total sales are binding (specified with macro parameter `alterNeg=0`). The balancing constraint is

- cars_alb + cars_sask + cars_man = cars_tot.

The problem specification data set, the input data set and the macro call would be

```
data myProblem;
   length _type_ $2 _col_ $9 _row_ $22;
   input _type_= _col_= _row_= _coef_=;
   datalines;
_type_=eq   _col_=.           _row_=Total cars aggregation   _coef_=.
_type_=.    _col_=cars_alb    _row_=Total cars aggregation   _coef_=1
_type_=.    _col_=cars_sask   _row_=Total cars aggregation   _coef_=1
_type_=.    _col_=cars_man    _row_=Total cars aggregation   _coef_=1
_type_=.    _col_=cars_tot    _row_=Total cars aggregation   _coef_=-1
;

data myData;
   input date cars_alb cars_sask cars_man cars_tot;
   informat date yyq6.;
   format date yyq6.;
   datalines;
2011q1 20 18 12 53
2011q2 16 16 19 44
2011q3 14 15 16 50
;

%GSeriesTSBalancing
(
   inTS           = myData,
   inProblemSpecs = myProblem,
   outTS          = outBalanced,
   timeVarName    = date,
   alterNeg       = 0   /* binding total sales (CARS_TOT) */
)
```

# Example 2

# Two-dimensional raking problem with non-negativity constraints

In this example, we have estimates of "cars" and "vans" sales for the "alb", "sask" and "man" regions from the 2nd quarter of 2010 to the 1st quarter of 2012 and we want to preserve annual totals for 2011. The raking marginal totals are binding as well as the annual totals. The balancing constraints are

- cars_alb + cars_sask + cars_man = cars_tot;
- vans_alb + vans_sask + vans_man = vans_tot;
- cars_alb + vans_alb = alb_tot;
- cars_sask + vans_sask = sask_tot;
- cars_man + vans_man = man_tot.

- cars_alb, cars_sask, cars_man, vans_alb, vans_sask, vans_man, cars_tot, vans_tot, alb_tot, sask_tot and man_tot should be ≥ 0.

An alterability coefficient of 0 is used for time series "vans_sask" for the 2nd quarter of 2010 while the default value of 1 is used for all other time periods. Time series non-negativity is specified with macro parameter `lowerBound=0`. While the annual totals are binding by default, binding marginal totals are specified with parameter `alterNeg=0`. Display level 4 is specified in order to display the balancing results in the Output window and three optional output data sets are requested: the *OPTMODEL summary* data set, the *detailed balancing results* data set and the *temporal totals* data set.

The problem specification data set, the input data set and the macro call would be

```
data myProblem;
   length _type_ $5 _col_ $9 _row_ $30;
   input _type_= _col_= _row_= _coef_= _timeVal_=;
   informat _timeVal_ yyq6.;
   format _timeVal_ yyq6.;
   datalines;
_type_=eq      _col_=.          _row_=Total cars aggregation           _coef_=.    _timeVal_=.
_type_=.       _col_=cars_alb   _row_=Total cars aggregation           _coef_=1    _timeVal_=.
_type_=.       _col_=cars_sask  _row_=Total cars aggregation           _coef_=1    _timeVal_=.
_type_=.       _col_=cars_man   _row_=Total cars aggregation           _coef_=1    _timeVal_=.
_type_=.       _col_=cars_tot   _row_=Total cars aggregation           _coef_=-1   _timeVal_=.
_type_=eq      _col_=.          _row_=Total vans aggregation           _coef_=.    _timeVal_=.
_type_=.       _col_=vans_alb   _row_=Total vans aggregation           _coef_=1    _timeVal_=.
_type_=.       _col_=vans_sask  _row_=Total vans aggregation           _coef_=1    _timeVal_=.
_type_=.       _col_=vans_man   _row_=Total vans aggregation           _coef_=1    _timeVal_=.
_type_=.       _col_=vans_tot   _row_=Total vans aggregation           _coef_=-1   _timeVal_=.
_type_=eq      _col_=.          _row_=Alberta total aggregation        _coef_=.    _timeVal_=.
_type_=.       _col_=cars_alb   _row_=Alberta total aggregation        _coef_=1    _timeVal_=.
_type_=.       _col_=vans_alb   _row_=Alberta total aggregation        _coef_=1    _timeVal_=.
_type_=.       _col_=alb_tot    _row_=Alberta total aggregation        _coef_=-1   _timeVal_=.
_type_=eq      _col_=.          _row_=Saskatchewan total aggregation   _coef_=.    _timeVal_=.
_type_=.       _col_=cars_sask  _row_=Saskatchewan total aggregation   _coef_=1    _timeVal_=.
_type_=.       _col_=vans_sask  _row_=Saskatchewan total aggregation   _coef_=1    _timeVal_=.
_type_=.       _col_=sask_tot   _row_=Saskatchewan total aggregation   _coef_=-1   _timeVal_=.
_type_=eq      _col_=.          _row_=Manitoba total aggregation       _coef_=.    _timeVal_=.
_type_=.       _col_=cars_man   _row_=Manitoba total aggregation       _coef_=1    _timeVal_=.
_type_=.       _col_=vans_man   _row_=Manitoba total aggregation       _coef_=1    _timeVal_=.
_type_=.       _col_=man_tot    _row_=Manitoba total aggregation       _coef_=-1   _timeVal_=.
_type_=alter   _col_=.          _row_=Alterability coefficient         _coef_=.    _timeVal_=.
_type_=.       _col_=vans_sask  _row_=Alterability coefficient         _coef_=0    _timeVal_=2010q2
;

data myData;
   input date cars_alb cars_sask cars_man vans_alb vans_sask vans_man alb_tot sask_tot
         man_tot cars_tot vans_tot;
   informat date yyq6.;
   format date yyq6.;
   datalines;
2010q2 14 18 14 20 21 27 32 38 43 58 55
2010q3 17 14 16 29 25 26 41 35 39 44 71
2010q4 14 19 18 20 28 27 41 40 45 58 68
2011q1 20 18 12 20 22 26 32 39 43 53 61
2011q2 16 16 19 21 26 21 40 30 33 44 59
2011q3 14 15 16 19 25 19 34 47 40 50 71
2011q4 19 20 14 21 18 27 44 44 38 52 74
2012q1 16 15 19 27 25 28 35 38 32 51 54
;
```

```
%GSeriesTSBalancing
```

```
(
   inTS                = myData,
   inProblemSpecs      = myProblem,
   outTS               = outBalanced,
   timeVarName         = date,
   periodInterval      = QTR,
   temporalGrpInterval = YEAR,  /* preserve annual totals */
   displayLevel        = 4,     /* display the balancing results */
   outOptModelSummary  = outSummary,
   outDetailedResults  = outResults,
   outTemporalTotals   = outAnnTotals,
   alterNeg            = 0,     /* binding marginal totals */
   lowerBound          = 0      /* non-negativity constraints */
)
```

---

# Appendix – Mathematical formulation of the balancing problem

# Notation

| Symbol | Description |
|---|---|
| $P$ | Total number of independent optimization problems (processing groups) to be solved with the OPTMODEL procedure in order to balance the system of time series. The set of periods included in processing group $j$ is given by $T_j$ (see below). |
| $K$ | Total number of time series (problem variables) in the system. |
| $T$ | Total number of periods in the time series system. |
| $T_j$ | Represents the partition of the time series system in the time dimension. When temporal totals are preserved, $T_j$ either corresponds to the set of periods belonging to a complete temporal group (e.g. all periods of a complete year) or to a specific period for incomplete temporal groups (e.g. individual periods of incomplete years). Otherwise, when temporal totals are not preserved, $T_j$ always represents a specific period. |
| $x_{kt}$ | Initial (unbalanced) time series values, $k=1,\ldots,K;\quad t=1,...,T.$ |
| $\theta_{kt}^{(x)}$ | Balanced time series values, $k=1,\ldots,K;\quad t=1,...,T.$ |
| $c_{kt}^{(x)}$ | Alterability coefficient associated with $x_{kt}$ (regular alterability coefficients). |
| $a_{kj}$ | Initial temporal totals for time series $x_k$, $k=1,...,K$; $j$ belonging to the set of processing groups that correspond to complete temporal groups (e.g. complete years). |
| $\theta_{kj}^{(a)}$ | Final temporal totals for time series $x_k$, $k=1,...,K$; $j$ belonging to the set of processing groups that correspond to complete temporal groups (e.g. complete years). |
| $c_{kj}^{(a)}$ | Alterability coefficient associated with $a_{kj}$ (temporal total alterability coefficients). |
| $I^{(eq)}$ | Total number of linear *equality* constraints. |
| $b_i^{(eq)}$ | Right-hand-side value (constant) of linear *equality* constraint $i$, $i=1,\ldots,I^{(eq)}$. |
| $G_{ik}^{(eq)}$ | Coefficient of time series $x_{kt}$ in *equality* constraint $i$. Same value for each period $t$, $t=1,\ldots,T$. |
| $I^{(le)}$ | Total number of linear *lower-or-equal-to* inequality constraints. |

| Symbol | Description |
|---|---|
| $b_i^{(le)}$ | Right-hand-side value (constant) of linear *lower-or-equal-to* inequality constraint $i$, $i = 1,\ldots,I^{(le)}$. |
| $G_{ik}^{(le)}$ | Coefficient of time series $x_{kt}$ in *lower-or-equal-to* inequality constraint $i$. Same value for each period $t$, $t = 1,\ldots,T$. |
| $I^{(ge)}$ | Total number of linear *greater-or-equal-to* inequality constraints. |
| $b_i^{(ge)}$ | Right-hand-side value (constant) of linear *greater-or-equal-to* inequality constraint $i$, $i = 1,\ldots,I^{(ge)}$. |
| $G_{ik}^{(ge)}$ | Coefficient of time series $x_{kt}$ in *greater-or-equal-to* inequality constraint $i$. Same value for each period $t$, $t = 1,\ldots,T$. |

# Minimization problem

Balancing a system of time series with data for $T$ periods, $\{1,2,\ldots,T\} = \bigcup_j T_j$, can be formulated as a sequence of $P$ quadratic minimization problems to be solved independently. Using the notation defined previously, the minimization problem for each processing group $j$, $j = 1,...,P$, is defined as follows.

If $j$ corresponds to a complete temporal group $\left(\text{card}\left(T_j\right) > 1\right)$:

$$\min_{\theta} \quad \sum_{k=1}^{K}\sum_{t\in T_j}\frac{\left(x_{kt}-\theta_{kt}^{(x)}\right)^2}{s_{kt}^{(x)}}+\sum_{k=1}^{K}\frac{\left(a_{kj}-\theta_{kj}^{(a)}\right)^2}{s_{kj}^{(a)}},$$

$$\text{where} \quad s_{kt}^{(x)} = \begin{cases} 1 & \text{if } \left|c_{kt}^{(x)}x_{kt}\right| = 0 \\ \left|c_{kt}^{(x)}x_{kt}\right| & \text{otherwise} \end{cases}, \quad a_{kj} = \sum_{t\in T_j} x_{kt}, \quad s_{kj}^{(a)} = \begin{cases} 1 & \text{if } \left|c_{kj}^{(a)}a_{kj}\right| = 0 \\ \left|c_{kj}^{(a)}a_{kj}\right| & \text{otherwise} \end{cases}$$

subject to:

$$\left.\begin{aligned} &\sum_{k=1}^{K} G_{ik}^{(eq)}\theta_{kt}^{(x)} = b_i^{(eq)} \quad i = 1,...,I^{(eq)},\ t\in T_j \\ &\sum_{k=1}^{K} G_{ik}^{(le)}\theta_{kt}^{(x)} \le b_i^{(le)} \quad i = 1,...,I^{(le)},\ t\in T_j \\ &\sum_{k=1}^{K} G_{ik}^{(ge)}\theta_{kt}^{(x)} \ge b_i^{(ge)} \quad i = 1,...,I^{(ge)},\ t\in T_j \end{aligned}\right\} \text{ equivalent to } G\theta \text{ OP } b \text{ where OP is } =,\ \le \text{ or } \ge.$$

$$\theta_{kt}^{(x)} = x_{kt} \quad \text{if } \left|c_{kt}^{(x)}x_{kt}\right| = 0,\ t\in T_j$$

$$\sum_{t\in T_j}\theta_{kt}^{(x)} = \theta_{kj}^{(a)} \quad k = 1,...,K$$

$$\theta_{kj}^{(a)} = a_{kj} \quad \text{if } \left|c_{kj}^{(a)}a_{kj}\right| = 0,\ k = 1,...,K$$

Otherwise, if $j$ corresponds to a single period $\left(\text{card}\left(T_j\right)=1\right)$:

$$\min_{\theta} \quad \sum_{k=1}^{K}\sum_{t\in T_j}\frac{\left(x_{kt}-\theta_{kt}^{(x)}\right)^2}{s_{kt}^{(x)}}, \quad \text{where} \quad s_{kt}^{(x)}=\begin{cases}1 & \text{if } \left|c_{kt}^{(x)}x_{kt}\right|=0\\ \left|c_{kt}^{(x)}x_{kt}\right| & \text{otherwise}\end{cases}$$

subject to:

$$\left.\begin{aligned}&\sum_{k=1}^{K}G_{ik}^{(eq)}\theta_{kt}^{(x)}=b_i^{(eq)} \quad i=1,...,I^{(eq)},\ t\in T_j\\ &\sum_{k=1}^{K}G_{ik}^{(le)}\theta_{kt}^{(x)}\le b_i^{(le)} \quad i=1,...,I^{(le)},\ t\in T_j\\ &\sum_{k=1}^{K}G_{ik}^{(ge)}\theta_{kt}^{(x)}\ge b_i^{(ge)} \quad i=1,...,I^{(ge)},\ t\in T_j\end{aligned}\right\} \text{ equivalent to } G\theta \text{ OP } b \text{ where OP is } =,\ \le \text{ or } \ge.$$

$$\theta_{kt}^{(x)}=x_{kt} \qquad \text{if } \left|c_{kt}^{(x)}x_{kt}\right|=0,\ t\in T_j$$

# Interpretation

The general idea is to minimize the distance between the time series initial values, denoted $x_{kt}$ in the minimization problem, and their balanced (final) values $\theta_{kt}^{(x)}$ 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 of the minimization problem $\left(\text{for each } t,\ t=1,...,T\right)$. They are the $G\theta$ OP $b$ constraints, where OP stands for operator (=, ≤ or ≥). They 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 $\theta_{kt}^{(x)}$ 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 temporal total $\theta_{kj}^{(a)}$. They are constraints $\sum_{t\in T_j}\theta_{kt}^{(x)}=\theta_{kj}^{(a)}$, where the distance between the initial temporal total $\left(a_{kj}\right)$ and the final temporal total $\left(\theta_{kj}^{(a)}\right)$ of time series $k$ for complete temporal group $j$ is minimized in the objective function. These implicit temporal constraints *do not exist for incomplete temporal groups or when temporal totals are not preserved*.

There are as many independent quadratic minimization problems to solve as there are processing groups $j$ to balance. The total number of processing groups $P$ depends on the number of periods provided in the input data and whether temporal totals are preserved or not.

**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* $\left(c_{kt}^{(x)}\right)$

and *temporal total alterability coefficients* $\left(c_{kj}^{(a)}\right)$. Since alterability coefficients appear at 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. Whenever 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. Conversely, we say that a time series value (or temporal total) is *non-binding* when its alterability coefficient is not 0. Time series temporal totals are usually binding as well as values of time series corresponding to marginal totals in raking problems (i.e. time series with a coefficient of -1 in the balancing aggregation constraints) while the values of component series in 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 sometimes be useful in order 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).

**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 while the change is minimized for non-binding temporal totals.

---

# Notes

This document is a guide for the use of the **GSeriesTSBalancing** macro. For more information, please contact the G-Series support team using the G-Series e-mail address (G-Series@statcan.gc.ca) or consult the web site at G-Series (Statistics Canada intranet only).

---

# References


Bikker R., Daalmans J. and Mushkudiani N. (2013). Benchmarking large accounting frameworks: A generalized multivariate model. *Economic Systems Research*, DOI:10.1080/09535314.2013.801010.

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.


---