Paper 146-2007

Benchmarking Sub-Annual Series to Annual Totals – From Concepts to SAS® Procedure and SAS® Enterprise Guide® Custom Task

Edith Latendresse, Statistics Canada, Ottawa, Canada
Maurice Djona, Statistics Canada, Ottawa, Canada
Susie Fortier, Statistics Canada, Ottawa, Canada

ABSTRACT

Situations that require benchmarking are very common in statistical agencies. Benchmarking is defined as an adjustment of the level of a sub-annual series using auxiliary annual benchmarks. The sub-annual series is modified so that the annual sums of the sub-annual series are equal to the corresponding benchmarks. This is done while preserving the movement in the sub-annual series as much as possible as well as considering that the benchmarks at the end of the series might not be available yet. This paper illustrates the benchmarking methodology developed at Statistics Canada. The presented method is a special case of the general regression-based benchmarking model proposed by Dagum and Cholette (2006).

The paper also presents the innovative implementation of the methodology with a complete SAS procedure called PROC BENCHMARKING, developed at Statistics Canada for UNIX and Microsoft Windows operating systems, using SAS/TOOLKIT®. The procedure is presented through a custom add-in task for SAS Enterprise Guide and Microsoft Office, developed to provide a user-friendly interface and produce analytical tables and graphs based on the benchmarking results.

The methodology presented is primarily used by economists and analysts, while the accompanying SAS procedure and custom task require basic knowledge of SAS and SAS Enterprise Guide.

INTRODUCTION

Benchmarking is defined as an adjustment of the level of a sub-annual series to the level of an annual series. Both time series measure the same variable, but at different time intervals and different levels of accuracy. The annual series is usually more accurate than the sub-annual series and, for this reason, it is considered as benchmark. Because the two series are coming from different sources, discrepancies are usually observed between the annual benchmarks and the annual sums of the sub-annual series. With benchmarking, the sub-annual series is modified so that the annual sums of the sub-annual series are equal to the corresponding benchmarks. This is done considering two main challenges. The first one is to preserve the movement in the sub-annual series as much as possible. The second one is to account for the timeliness of annual benchmarks, in the sense that the benchmarks for the observations at the end of the series may not be available yet. The presented method is a special case of the general regression-based benchmarking model proposed by Dagum and Cholette (2006).

First, the notation and the parameters involved will be introduced before setting out the benchmarking methodology. Then, the SAS procedure and the custom task developed to implement the methodology will be presented. An example will be used to demonstrate navigating through the different custom task window panes.

NOTATION

The sub-annual series $s_t, t = 1,...T$, is associated with a series of dates. In general, the date values take the form $yyyypp$, where $yyyy$ represents the year and pp represents the period (for example, from 1 to 4 for quarterly series and from 1 to 12 for monthly series). It is assumed that the dates are mapped into the set of integers $1 \text{ to } T$. Sub-annual series are also referred to as indicator series. Let $s = (s_1,..., s_T)'$.

The annual benchmarks $a_m, m = 1,..., M$ are also associated with dates. The starting date is $t_{1,m}$ and the ending date is $t_{2,m}$ such that $1 \le t_{1,m} \le t_{2,m} \le T$. With this notation, a benchmark is not restricted to cover a calendar year. It could be, for example, a fiscal year, a six-month period or even individual values at arbitrary points along the series (in that situation $t_{1,m} = t_{2,m}$). In the following, we will refer to an *annual* benchmark without loss of generality. Let $a = (a_1,..., a_M)'$.

The benchmarked series $\hat{\theta}_t, t = 1,..., T$, is such that $\sum_{t=t_{1,m}}^{t_{2,m}} \hat{\theta}_t = a_m; m = 1,..., M$. Let $\hat{\theta} = (\hat{\theta}_1,..., \hat{\theta}_T)'$.

We finally define the matrix operator J with dimensions $M \times T$ such that Js is the vector of the annual totals according to the sub-annual data. To do so, the matrix J takes the form

$$J = \begin{bmatrix} j_{1,1} & j_{1,2} & \cdots & j_{1,T} \\ \vdots & \vdots & \ddots & \vdots \\ j_{M,1} & j_{M,2} & \cdots & j_{M,T} \end{bmatrix}$$

where, for $m = 1,\ldots,M$ and $t = 1,\ldots,T$, $j_{m,t} = \begin{cases} 1 & t_{1,m} \le t \le t_{2,m} \\ 0 & \text{otherwise} \end{cases}$.

PARAMETERS

The benchmarking method considered is driven by a few parameters.

The first one is the autoregressive parameter $\rho : 0 \le \rho \le 1$, a smoothing parameter with suggested default values $\rho = 0.9$ for monthly sub-annual series and $\rho = 0.9^3 = 0.729$ for quarterly sub-annual series.

The second parameter is λ, the adjustment model parameter. It is a real number with default values $\lambda = 1$ for a proportional benchmarking model, $\lambda = 0$ for an additive benchmarking model and $\lambda = 0.5$, with $\rho = 0$, for pro-rating.

The last parameter concerns the bias. The bias is defined as the expected discrepancy between an annual benchmark and its related sub-annual values. Bias estimation is useful to account for the timeliness of annual benchmarks and helps minimize the revision of the benchmarked series when a new benchmark becomes available. A consistent estimate of the bias is the average discrepancy (when $\lambda = 0$):

$$b = \frac{\sum_{m=1}^{M} a_m - \sum_{m=1}^{M} \sum_{t=t_{1,m}}^{t_{2,m}} s_t}{\sum_{m=1}^{M} \sum_{t=t_{1,m}}^{t_{2,m}} 1}. \qquad (1)$$

When it is more convenient to express the bias in terms of a ratio instead of a difference in level (when $\lambda \ne 0$), the bias parameter can be estimated as

$$b = \frac{\sum_{m=1}^{M} a_m}{\sum_{m=1}^{M} \sum_{t=t_{1,m}}^{t_{2,m}} s_t}. \qquad (2)$$

Once a bias parameter is estimated, you can apply it or not. Let $c = b$ if the bias is applied. Default values are used for c when the bias is not applied[1]. If c is expressed in terms of a difference in level such as in Equation (1), then let $s_t^* = c + s_t$. If *c* is expressed as a ratio such as in Equation (2), then let $s_t^* = c \cdot s_t$. The series s_t^* is called the *re-scaled sub-annual series* or the *indicator series corrected for bias*. Let $s^* = (s_1^*,\ldots,s_T^*)'$.

All the details about this method of estimating the bias parameter are explained in Dagum and Cholette (2006).

BENCHMARKING

Methodological details for the benchmarking formulae in this section are provided in Quenneville *et al.* (2006).

Define C as the $T \times T$ matrix with $\left|s_t^*\right|^{\lambda}$ as the *t*-th element of the main diagonal[2] and with 0 elsewhere.

For *ρ*<1, the benchmarked series $\hat{\theta} = (\hat{\theta}_1,\ldots,\hat{\theta}_T)$ is

$$\hat{\theta} = s^* + V_e J' V_d^{-1} (a - Js^*),$$

where $V_d = JV_eJ'$; $V_e = C\Omega_e C$ and Ω_e is the $T \times T$ matrix[3] defined by $\Omega_e = ((\rho^{|i-j|}))$, $i, j = 1,\ldots,T$.

---

[1] The default value $c = 0$ is used when $\lambda = 0$; $c = 1$ is used otherwise.

[2] The diagonal elements of the C matrix can be re-scaled to avoid numerical problems. For example, they can be divided by their overall mean.

[3] By defining $\rho^0 = 1$, note that Ω_e is the Identity matrix when $\rho = 0$.

For $\rho = 1$, the benchmarked series $\hat{\theta}$ is

$$\hat{\theta} = s^* + W(a - Js^*),$$

where W is the $T \times M$ upper-right corner matrix from the following matrix product:

$$\begin{bmatrix} C^{-1}\Delta'\Delta C^{-1} & J' \\ J & 0 \end{bmatrix}^{-1} \begin{bmatrix} C^{-1}\Delta'\Delta C^{-1} & 0 \\ J & I_M \end{bmatrix} = \begin{bmatrix} I_T & W \\ 0 & W_\upsilon \end{bmatrix},$$

Δ is the $(T-1)\times T$ matrix with -1 at index (i,i), 1 at index $(i,i+1)$, $i = 1,...,T-1$ and 0 elsewhere and I_M is the $M \times M$ Identity matrix. (The $M \times M$ matrix W_υ is associated with the Lagrange multipliers.)

PROC BENCHMARKING AND SAS ENTERPRISE GUIDE CUSTOM TASK

The methodology has been implemented with a complete SAS procedure called PROC BENCHMARKING developed at Statistics Canada under the project name Forillon. The in-house SAS procedure has been created using the C language and SAS/TOOLKIT. The development of a SAS procedure includes four basic steps:

(1) writing a grammar that defines the syntax of the procedure,
(2) writing the procedure itself using C, FORTRAN or PL/I language,
(3) using the SAS-supplied USERPROC procedure to create some required intermediate files, and
(4) compiling and linking the procedure with the SAS system.

Detailed information on the process and examples can be found in the book by the SAS Institute Inc. (1991).

Furthermore, a custom add-in task for SAS Enterprise Guide and Microsoft Office has been developed to provide a user-friendly interface for the use of PROC BENCHMARKING. Building a custom task basically involves the required implementation of three interfaces supplied by SAS in the SAS.Shared.AddIns namespace. This is done using any Microsoft Visual Studio .NET language (VB or C#). The implementation of these interfaces makes the component a valid add-in for SAS Enterprise Guide and Microsoft Office. A link to the information on creating a custom task using Visual Studio .NET is given in the references (SAS Institute Inc. (2006)). The developed custom task allows you to easily provide input files and parameters to the procedure. It can also produce analytical tables and graphs based on the benchmarking results. An example illustrating how to use the procedure through the custom task is presented below.

The look and feel of the benchmarking custom task is very similar to that of the SAS-supplied Enterprise Guide Tasks. When accessing the custom task through the SAS Enterprise Guide, you first have to provide the location and the name of the sub-annual series data. The sub-annual SAS data set must contain three mandatory numerical variables: YEAR, PERIOD and VALUE. Once the sub-annual file is provided, the Input Data pane appears (as in Figure 1). The procedure requires the sub-annual data set to be sorted by year and period in ascending order. The *Sort* box has to be checked if the data set is not already sorted. You then have to provide the name of the benchmarks file. The benchmarks SAS data set must contain five mandatory numerical variables: STARTYEAR, STARTPERIOD, ENDYEAR, ENDPERIOD, and VALUE.

The following statements read a seasonal quarterly series into a SAS data set called mySeries, and the corresponding annual benchmarks into a SAS data set called myBenchmarks.

```
DATA SGF2007.mySeries;
INPUT @01 year 4.
      @06 period 1.
      @08 value;
CARDS;
1998 1 1851
1998 2 2436
1998 3 3115
1998 4 2205
1999 1 1987
1999 2 2635
1999 3 3435
1999 4 2361
2000 1 2183
...
;
RUN;
```

```
DATA SGF2007.myBenchmarks;
   INPUT @01 startYear 4.
         @06 startPeriod 1.
         @08 endYear 4.
         @13 endPeriod
         @15 value;
   CARDS;
   1998 1 1998 4 10324
   1999 1 1999 4 10200
   ...
   ;
   RUN;
```

These two data sets are given as input files in Figure 1. Note that the procedure allows for more generalized benchmarks. They could be non-consecutive – say, with benchmarks every other year – and could even overlay, as long as they are coherent.

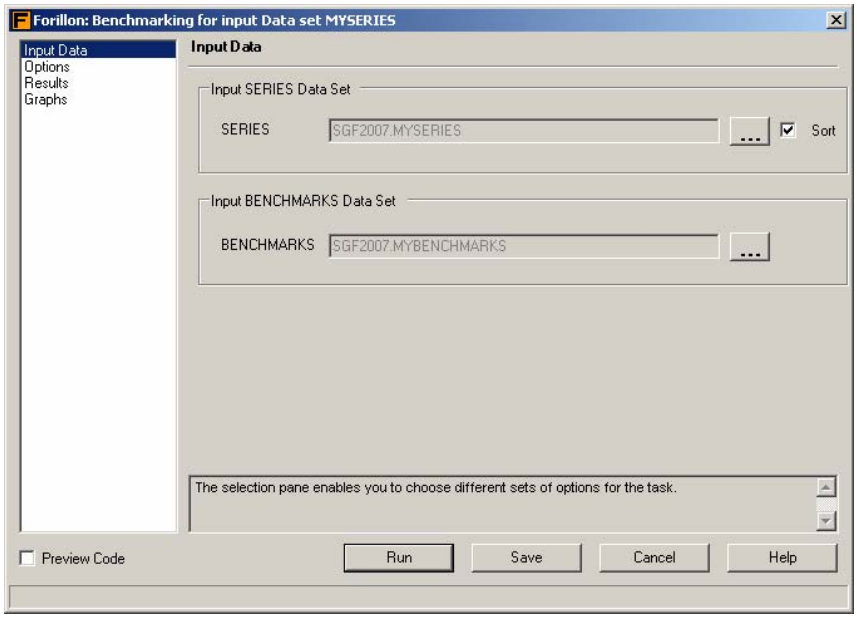

Figure 1. Input Data pane

The left side pane shown in Figure 1 allows the navigation between different sets of input parameters and selections. The bottom pane provides indications on the selection of the different parameters or input files.

Once the two input data files are provided, selection of Options in the left side pane brings you to the Options pane (Figure 2). The Options pane consists of three sections, one section for each parameter of the benchmarking methodology. You must specify values for the LAMBDA, RHO and BIAS parameters. In the example, suggested default values for λ and ρ for a quarterly series are selected. By checking the *User-provided value* box for either LAMBDA or RHO, you can provide your own values: RHO has to be a non-negative real number and suggested values for LAMBDA are real numbers between -3 and 3 (to ensure that computation limits are not exceeded). The last section concerns the bias parameters: BIASOPTION and BIAS. Valid values for the BIASOPTION are:

- BIASOPTION=1: No bias estimation. The bias used to pre-adjust the sub-annual series will be the default value or the value specified with the BIAS= option.
- BIASOPTION=2: Bias estimation for informational purposes only. The bias value is estimated, printed in the log, but not used to pre-adjust the sub-annual series. The bias used to pre-adjust the sub-annual series will be the default value or the value specified with the BIAS= option.
- BIASOPTION=3: Bias estimation. The bias value is estimated and used to pre-adjust the sub-annual series. Any value specified with the BIAS= option will not be used.

In the example, both options BIASOPTION and BIAS are used. With BIASOPTION=2, the bias will be estimated with all the available benchmarks and will be printed in the log (Figure 3), but it will not be applied. The bias that will be applied is BIAS=0.94 which is in fact the bias externally computed with only the last three years of benchmarks.

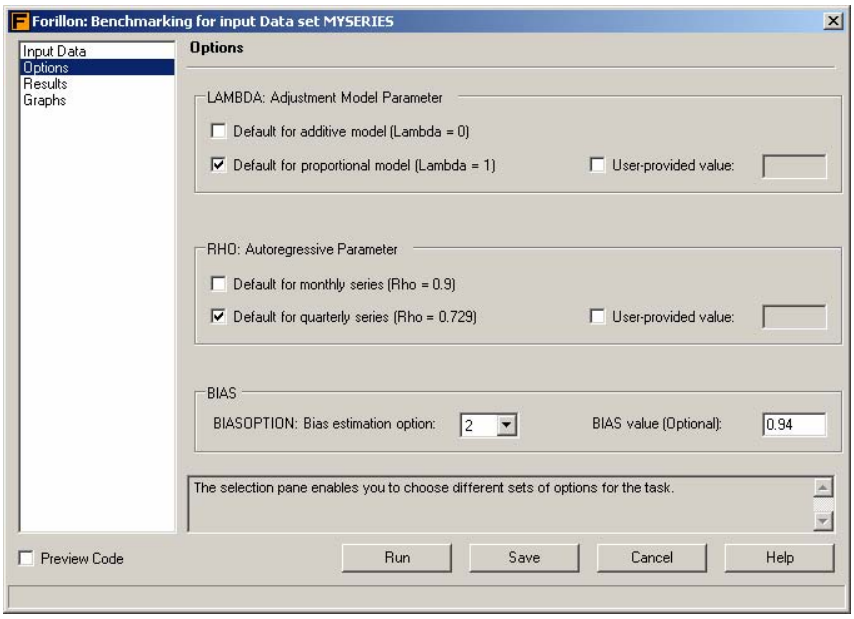

Figure 2. Options pane

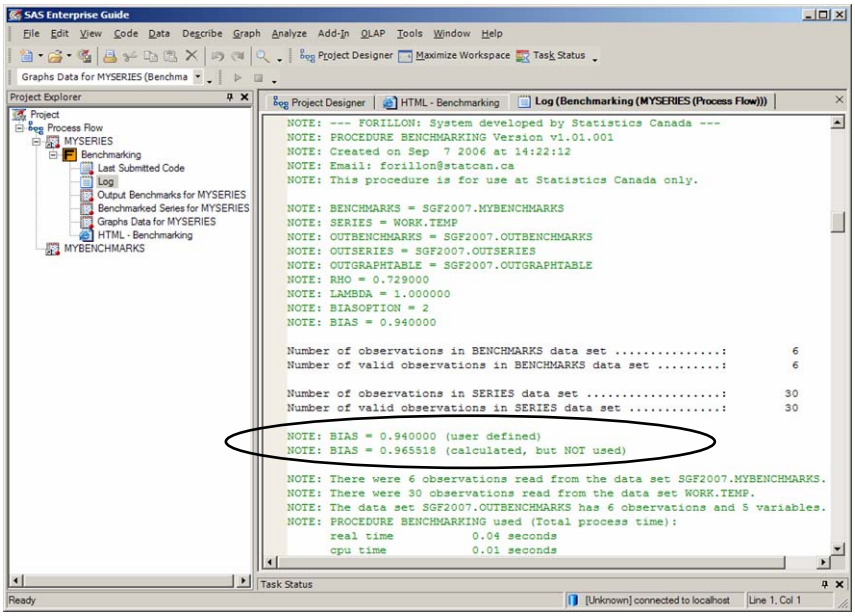

Figure 3. Some notes printed in the SAS Log by PROC BENCHMARKING

The Results pane (Figure 4) concerns the output files. The OUTBENCHMARKS file will store the benchmarks used by the procedure. This output benchmarks data set may be different from the input one if one or more benchmarks are found by the procedure to be redundant (i.e. are a linear combination of other benchmarks) and therefore are eliminated. The OUTSERIES file will contain the resulting benchmarked series. If OUTGRAPHTABLE is checked, all data necessary for the production of graphs and analytical tables will be generated and saved in a SAS data set. As for a typical SAS-supplied task, you might want to change the names of the output files instead of using the default names. This can be achieved by clicking on the corresponding *Modify* button as it has been done for OUTSERIES, OUTBENCHMARKS and OUTGRAPHTABLE files in the example (Figure 4).

The last pane (Figure 5) refers to graphs and analytical tables. To use that pane, the OUTGRAPHTABLE box must have been checked previously on the Results pane (Figure 4). The Graphs pane offers you three different types of graphs/tables: Original Scale, Adjustment Scale, and Growth Rates. Figure 5 shows the choices made for the current example.

Note that at the bottom of the graphs pane (Figure 5), you can select the start year and the end year of the graphs. This feature is very useful if you want to focus on a specific part of the series. However, the benchmarking procedure itself is always applied to the whole input series data set.

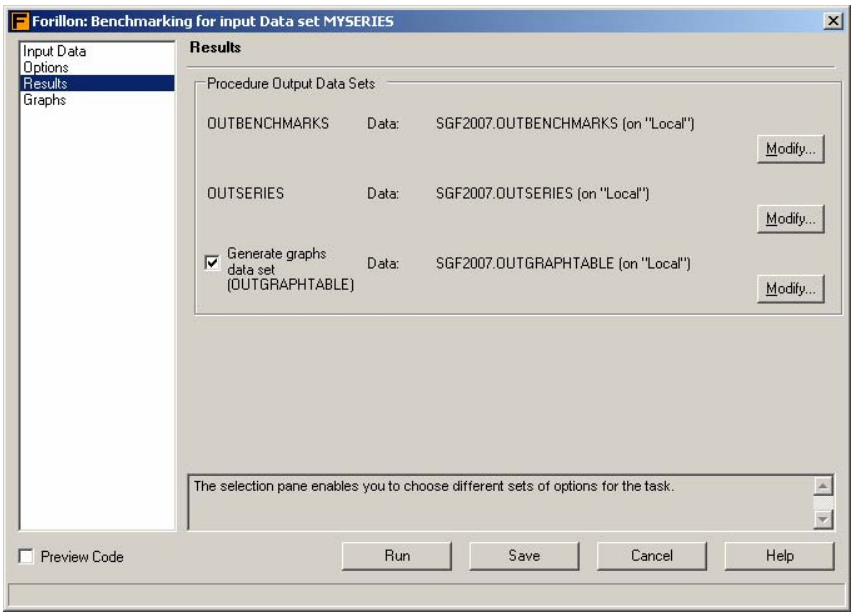

Figure 4. Results pane

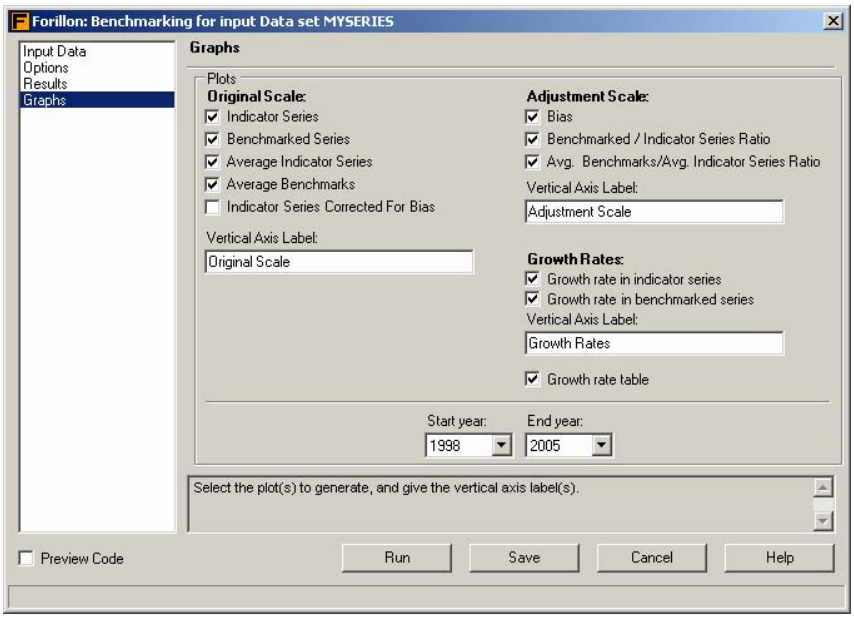

Figure 5. Graphs pane

For graphs on the original scale, with the series and parameters of the example, Figure 6 is produced. It overlays the indicator series, the resulting benchmarked series, the average indicator series and the average benchmarks series. The average series represent the annual figures from respectively the sub-annual series and the benchmarks series, divided by the number of periods covered by the benchmarks (4 for quarterly data, 12 for monthly data). They form two straight lines over each year a benchmark value is provided and give a good indication of the original discrepancies between the two sources of data. It is also possible to overlay the indicator series corrected for bias.

For graphs on the adjustment scale, with the series and parameters of the example, Figure 7 is produced. It overlays the used bias, the benchmarked series to indicator series ratio (BI ratio) and the average benchmarks to average indicator series ratio (average BI ratio). The BI ratio illustrates the movement preservation between the benchmarked and indicator series. The average BI ratio corresponds to simple pro-rating of the annual discrepancies on the sub-annual series. This ratio is the step-function seen on Figure 7. The graph shows that with pro-rating, movement is fully preserved within a year because a constant adjustment (average BI ratio) is applied to each period. However, “jumps” can be seen from one year to another. The BI ratio line shows that the benchmarking smoothes out these jumps and consequently better preserves the overall movement of the sub-annual series. The horizontal straight line at 0.94 represents the user-provided bias. With the selected parameters, the BI ratio converges to the bias for years without a benchmark. Without bias estimation, but with the same parameters, the BI ratio would converge to 1. For proportional benchmarking, as used in the example,

graphs on the adjustment scale are displayed in terms of ratio, as it would be the case for pro-rating. For additive benchmarking, the graphs would show differences instead of ratios but the interpretation stays the same.

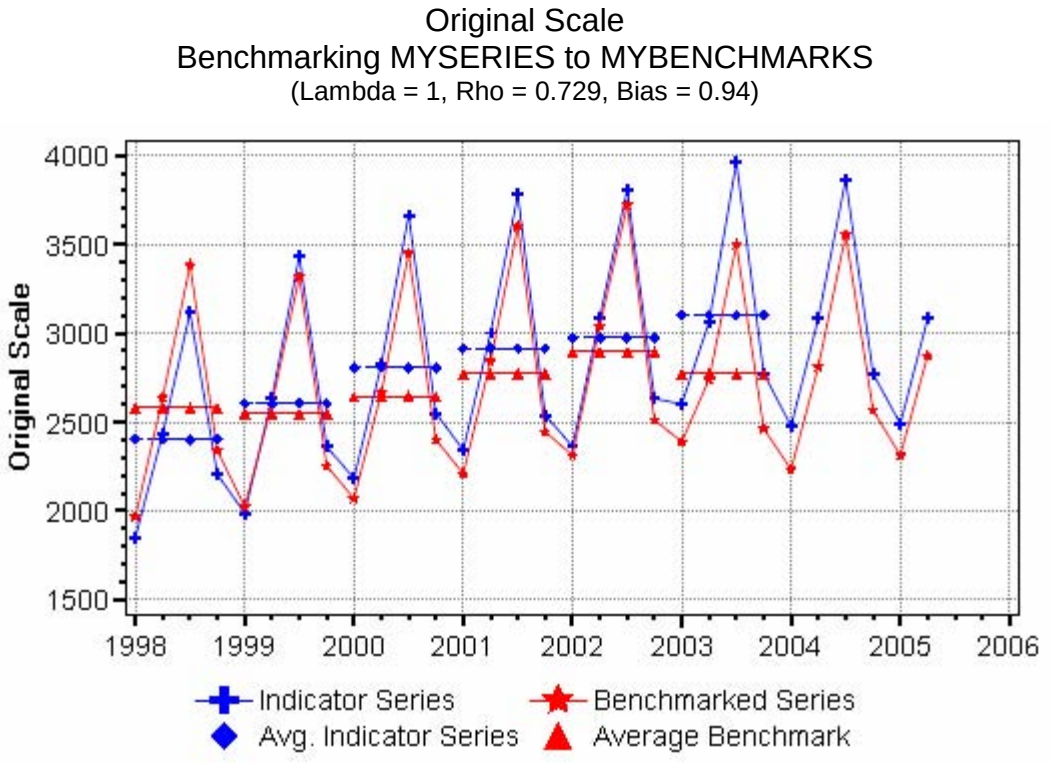

Figure 6. Original Scale graph

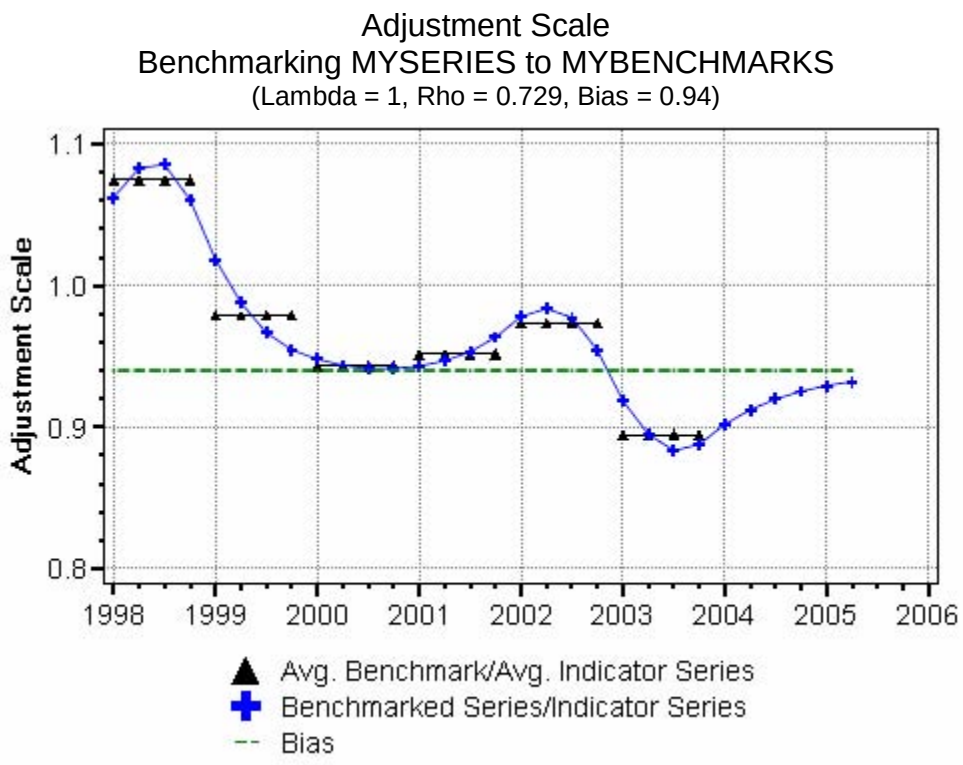

Figure 7. Adjustment Scale graph

The last type of graphs concerns growth rates. For proportional and pro-rating benchmarking, growth rates in both the indicator series and the benchmarked series are calculated in percentage, while for additive benchmarking, true growth is evaluated instead of rates. Growth rates can be presented in a bar chart or in an analytical table. They are useful to evaluate the movement preservation. Figure 8 and Figure 9 compare the growth rates of the indicator series to the growth rates of the benchmarked series for the example. The growth rates are very similar, showing that the benchmarking method insures good movement preservation from the indicator series to the benchmarked series.

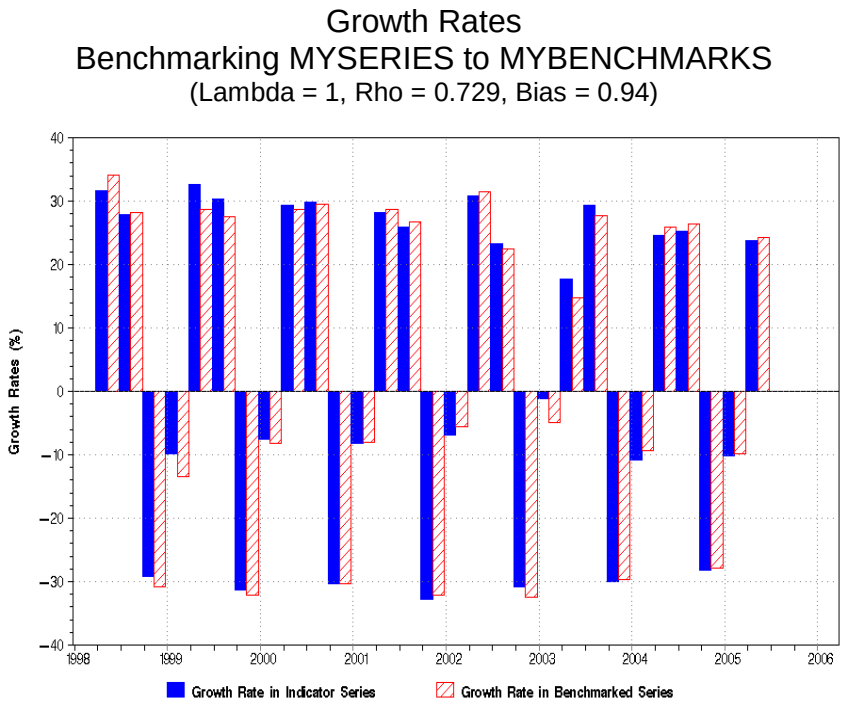

Figure 8. Growth Rates graph

Growth Rates
Benchmarking MYSERIES to MYBENCHMARKS
(Lambda = 1, Rho = 0.729, Bias = 0.94)

DATE	Indicator Series	Benchmarked Series	Growth Rate in Indicator Series (%)	Growth Rate in Benchmarked Series (%)
1998-01	1851	1966.37	.	.
1998-02	2436	2638.13	31.60	34.16
1998-03	3115	3380.88	27.87	28.15
1998-04	2205	2338.61	-29.21	-30.83
1999-01	1987	2023.58	-9.89	-13.47
1999-02	2635	2602.65	32.61	28.62
1999-03	3435	3320.35	30.36	27.58
1999-04	2361	2253.42	-31.27	-32.13
2000-01	2183	2069.45	-7.54	-8.16
2000-02	2822	2662.95	29.27	28.68
2000-03	3664	3449.09	29.84	29.52
2000-04	2550	2400.52	-30.40	-30.40
2001-01	2342	2208.68	-8.16	-7.99
2001-02	3001	2841.09	28.14	28.63

Figure 9. Growth Rates table (first portion)

For all the input parameters, the custom task will inform you by a pop-up if an invalid value has been given. A warning will also indicate if a mandatory variable is missing from one of the input files, or has a non-numeric type.

When all the required input parameters and selections are correctly provided, the Run button is enabled and has to be pushed for running the procedure (see Figure 5). There are three other buttons at the bottom of the window. The Save button can be used to save the task options and return to the SAS Enterprise Guide process flow window without running the task. The Cancel button allows you to quit the custom task window without saving the parameters and without running the task. The Help button gives access to the benchmarking procedure user guide which provides many details on each input parameter. To save the task (and the results) from one session of SAS Enterprise Guide to another, you must save the project from within the main SAS Enterprise Guide window.

That was an overview of what can be done with the custom task. Of course, the SAS code of the procedure is generated behind the scene. It can be accessed just by checking the *Preview code* box at the bottom left corner of the custom task window (see Figure 5), or by clicking "Last submitted Code" in the project explorer, once the task has run (as in Figure 10).

You might want to use the code in SAS directly, not through the custom task, and adapt it to run in batch mode, using different benchmarking selections on several series, for example. Figure 10 shows how easy it is to specify all parameters directly in the SAS code.

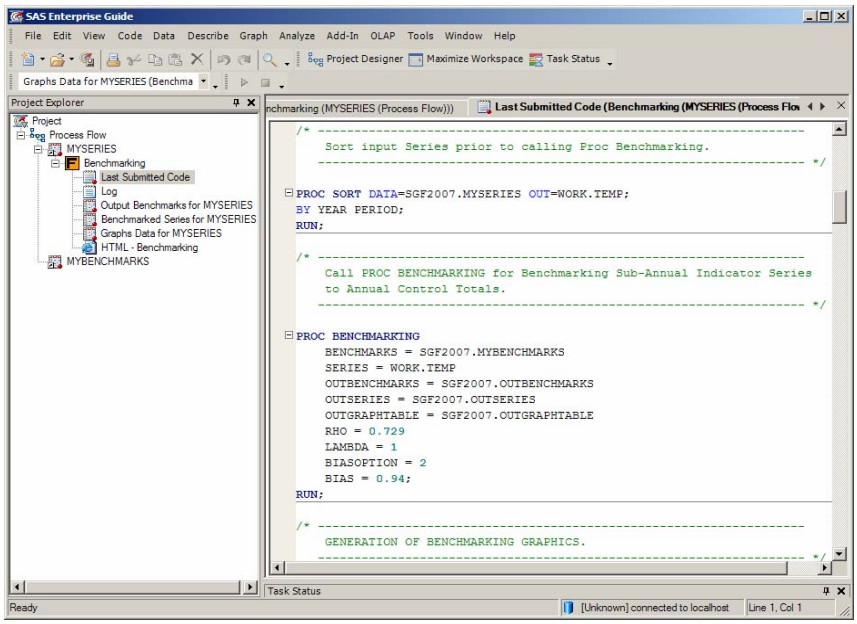

Figure 10. PROC BENCHMARKING SAS code

CONCLUSION

The custom SAS procedure PROC BENCHMARKING from Statistics Canada provides an easy way to benchmark sub-annual series to auxiliary annual benchmarks. Having a look and feel similar to that of SAS-supplied Enterprise Guide tasks, its accompanying custom task is an easy-to-use exploratory tool with graphs features to quickly visualize the original series, the benchmarked series, the growth rates, etc. If you are interested in a benchmarking system running in batch mode, the code generated by the custom task can be incorporated in a production routine as it is the case for SAS-supplied procedures.

REFERENCES

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

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 Eurostat Conference on Seasonality, Seasonal Adjustment and Their Implications for Short-Term Analysis and Forecasting, Luxembourg, May 2006.

SAS Institute Inc. (1991), *SAS/TOOLKIT® Software: Usage and Reference, Version 6, First Edition*. Cary, NC: SAS Institute Inc.

SAS Institute Inc. 2006. "Creating Custom Add-In Tasks for SAS Enterprise Guide"
http://support.sas.com/documentation/onlinedoc/guide/customtasks/SAS.Shared.AddIns.zip

CONTACT INFORMATION

Your comments and questions are valued, encouraged and should be sent to forillon@statcan.ca. You may also contact the authors at:

Edith Latendresse
Statistics Canada
150 Tunney's Pasture Driveway
R.H. Coats, 17th floor
Ottawa, Ontario
K1A 0T6
Canada

(613) 951-2066
edith.latendresse@statcan.ca

Maurice Djona
Statistics Canada
150 Tunney's Pasture Driveway
R.H. Coats, 14th floor
Ottawa, Ontario
K1A 0T6
Canada

(613) 951-5331
maurice.djona@statcan.ca

Susie Fortier
Statistics Canada
150 Tunney's Pasture Driveway
R.H. Coats, 17th floor
Ottawa, Ontario
K1A 0T6
Canada

(613) 951-4751
susie.fortier@statcan.ca