Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Growth Rates Preservation (GRP) temporal benchmarking: Drawbacks and alternative solutions

by Jacco Daalmans, Tommaso Di Fonzo, Nino Mushkudiani and Reinier Bikker

Release date: June 21, 2018

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at STATCAN.infostats-infostats.STATCAN@canada.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p} preliminary
^{r} revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E} use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2018
Vol. 44, No. 1, pp. 43-60
Statistics Canada, Catalogue No. 12-001-X

Growth Rates Preservation (GRP) temporal benchmarking: Drawbacks and alternative solutions

Jacco Daalmans, Tommaso Di Fonzo, Nino Mushkudiani and Reinier Bikker[1]

Abstract

Benchmarking monthly or quarterly series to annual data is a common practice in many National Statistical Institutes. The benchmarking problem arises when time series data for the same target variable are measured at different frequencies and there is a need to remove discrepancies between the sums of the sub-annual values and their annual benchmarks. Several benchmarking methods are available in the literature. The Growth Rates Preservation (GRP) benchmarking procedure is often considered the best method. It is often claimed that this procedure is grounded on an ideal movement preservation principle. However, we show that there are important drawbacks to GRP, relevant for practical applications, that are unknown in the literature. Alternative benchmarking models will be considered that do not suffer from some of GRP's side effects.

Key Words: Benchmarking; Growth rate preservation; Data reconciliation; Macro integration.

1 Introduction

Benchmarking monthly and quarterly series to annual data is a common practice in many National Statistical Institutes. For example, each year Statistics Netherlands aligns 12 quarterly Supply and Use Tables with the three most recent annual accounts (Eurostat, 2013, Annex 8C).

The benchmarking problem arises when time series data for the same target variable are measured at different frequencies with different levels of accuracy. One might expect that a temporal aggregation relationship between these time series is fulfilled, e.g., that four quarterly values add up to one annual value, but because of differences in data sources and processing methods, this is often not the case. Benchmarking is the process to remove such discrepancies. In this process the preliminary values are adjusted to achieve mathematical consistency between low-frequency (e.g., annual) and high-frequency (e.g., quarterly or monthly) time series.

There are two main principles of benchmarking. Firstly, low-frequency benchmarks are fixed, because these data sources describe levels and long-term trends better than high-frequency sources. Secondly, short-term movements of high-frequency time series are preserved as much as possible, as these data sources provide the only information on short-term movements.

Several benchmarking methods are available in the literature. These methods differ in the way short-term movements of high-frequency series are defined. A distinction can be made between multiplicative and additive methods. Multiplicative methods try to preserve relative changes of preliminary high-frequency time series, while additive methods aim to preserve changes in absolute terms. In this paper the focus will be solely on multiplicative variants.

1. Jacco Daalmans, Statistics Netherlands, PO Box 24500, 2490 HA Den Haag, Netherlands. E-mail: j.daalmans@cbs.nl; Tommaso Di Fonzo, Department of Statistical Sciences, University of Padua, Via Cesare Battisti, 241, 35121 Padova PD, Italy. E-mail: difonzo@stat.unipd.it; Nino Mushkudiani, Statistics Netherlands, Den Haag. E-mail: n.mushkudiani@cbs.nl; Reinier Bikker, Statistics Netherlands, Den Haag, E-mail: r.bikker@cbs.nl.

Two well-known multiplicative methods are Denton Proportionate First Differences (PFD), by Denton (1971), and Growth Rates Preservation (GRP) by Causey and Trager (1981; see also Trager, 1982 and Bozik and Otto, 1988).

In the literature it is generally agreed that GRP is grounded on the strongest theoretical foundation (Bloem, Dippelsman and Maehle, 2001, page 100). It explicitly preserves the period-to-period rates of change of the preliminary series. However, Denton PFD is more popularly used, because it is technically easier to apply. Mathematically, the Denton method deals with a standard linearly constrained quadratic optimization problem, while GRP solves a more difficult linearly constrained nonlinear problem that can be efficiently solved by an interior-point-algorithm (Di Fonzo and Marini, 2015).

From a number of simulation studies it is known that Denton PFD and GRP lead to similar or close to similar results for the large majority of cases (Harvill Hood, 2005; Titova, Findley and Monsell, 2010; Di Fonzo and Marini, 2012 and Daalmans and Di Fonzo, 2014). Therefore Denton PFD can be used as an approximation of GRP.

The aim of this paper is to demonstrate that GRP suffers from drawbacks that are, to the best of our knowledge, not described in the literature. A first drawback is that it matters whether benchmarking is applied “forward” or “backward” in time. In this context, we will present a link with the time reversibility property from index number theory. A second drawback is that undesirable results may be obtained due to singularities in the GRP objective function.

A second aim of this paper is to present alternative benchmarking methods that do satisfy time reversibility. This paper may be valuable for practitioners who apply or consider to apply benchmarking techniques.

First, in Section 2, we will give a formal description of the Denton PFD and GRP benchmarking methods. Section 3 describes the drawbacks of the GRP method. In Section 4 two new benchmarking methods are proposed that can be used as an alternative for GRP. Results of an illustrative application to real-life data are given in Section 5. Finally, Section 6 concludes this paper.

2 Temporal benchmarking methods

This section explains the Denton PFD and GRP benchmarking procedures. Because temporal aggregation constraints are the same for Denton PFD and GRP, these are described first. Thereafter, the Denton PFD and GRP benchmarking procedures are explained.

We focus on univariate variants of these methods, in which temporal consistency is the main constraint of interest. The observations that are presented in the remainder of this paper are however also valid for the multivariate case, in which multiple time-series are reconciled simultaneously and additional constraints between time-series apply (see Di Fonzo and Marini, 2011 and Bikker, Daalmans and Mushkudiani, 2013).

2.1 General notation and temporal constraints

In general, temporal aggregation constraints can be expressed as a linear system of equalities $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{x}$ is the target vector of high-frequency values, $\mathbf{b}$ is a vector of low-frequency values, and $\mathbf{A}$ is a temporal aggregation matrix converting high- into low-frequency values.

The specific form of these constraints depends on the nature of the variables involved. For flow variables, a sum of subannual values, e.g., four quarterly values, usually needs to be the same as one annual value. For stock variables, one of the subannual values, usually the first or the last, needs to be the same as the relevant annual value. For example, for quarterly/annual flow variables, assuming for the sake of simplicity that the available time span begins on the first quarter of the first year and ends on the fourth quarter of the last observed year, it is

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Denoting by $\mathbf{p}$ a vector of preliminary values, in general it is $\mathbf{Ap} \neq \mathbf{b}$, otherwise no adjustment would be needed. We look for a vector of benchmarked estimates $\mathbf{x}^*$, a particular outcome for $\mathbf{x}$, which should be "as close as possible" to the preliminary values and that satisfies $\mathbf{Ax}^* = \mathbf{b}$.

Not all sub annual periods need to be covered by a benchmark. Thus, the number of rows in $\mathbf{A}$ may be smaller than the total number of annual periods, see e.g., Dagum and Cholette (2006) for more details.

In a benchmarking operation, characteristics of the original series $\mathbf{p}$ should be considered. For example, in an economic time series framework, the preservation of the temporal dynamics (however defined) of the preliminary series is often a major interest of the practitioner.

2.2 Growth Rates Preservation (GRP) and Denton PFD

This section gives a formal description of GRP and Denton PFD.

Causey and Trager (1981; see also Monsour and Trager, 1979 and Trager, 1982) obtain the benchmarked values $x_t^*,\ t = 1, \ldots, n$ as a solution to the following optimization problem:

$$\min_{x_t} f_{\mathrm{F}}^{\mathrm{GRP}}(\mathbf{x}) \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{b}, \quad \text{where} \quad f_{\mathrm{F}}^{\mathrm{GRP}}(\mathbf{x}) = \sum_{t=2}^{n} \left(\frac{x_t}{x_{t-1}} - \frac{p_t}{p_{t-1}} \right)^2. \qquad (2.1)$$

The GRP criterion to be minimized, $f_{\mathrm{F}}^{\mathrm{GRP}}(\mathbf{x})$, explicitly relates to growth rates: it minimizes the sum of squared differences between growth rates of preliminary and benchmarked values. The subscript "F" in the minimization function stands for "Forward", later in this paper a "Backward" minimization function will be defined.

Denton (1971) proposed a benchmarking procedure grounded on the *Proportionate First Differences* (PFD) between target and original series. Cholette (1984) slightly modified the result of Denton, in order to correctly deal with the starting conditions of the problem. The PFD benchmarked estimates are thus obtained as the solution to the constrained quadratic minimization problem

$$\min_{x_t} f_{\mathrm{F}}^{\mathrm{PFD}}(\mathbf{x}) \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{b}, \quad \text{where} \quad f_{\mathrm{F}}^{\mathrm{PFD}}(\mathbf{x}) = \sum_{t=2}^{n} \left(\frac{x_t}{p_t} - \frac{x_{t-1}}{p_{t-1}} \right)^2. \tag{2.2}$$

The Denton PFD criterion to be minimized, $f_{\mathrm{F}}^{\mathrm{PFD}}(\mathbf{x})$, is a sum of squared linear terms, which is easier to deal with than the nonlinear GRP objective function.

3 Two problems with GRP benchmarking

3.1 Time reversibility

Time reversibility means that it does not matter whether a method is applied forward or backward in time. This property can be of interest in many application areas.

In physics, it means that if time would run backwards, all motions are reversed. In index number theory, time reversibility was introduced in a classical work of Fisher (1922, page 64). It is stated that "if taking 1913 as a base and going forward to 1918, we find that, on the average, prices have doubled, then, by proceeding in the reverse direction, we ought to find the 1913 price level to be half that of 1918". The motivation of this principle is that the direction of time can be considered arbitrary; it does not have any naturally preferred direction.

Time reversibility can also be applied in the context of benchmarking. It means that if we would reverse a time series, apply benchmarking, and reverse the benchmarked series back again, we get exactly the same results as for benchmarking the original series. In other words: from the benchmarked results it cannot be seen whether benchmarking has been applied forward or backward in time.

Benchmarking a reversed time series, according to GRP and Denton PFD, respectively, is equivalent to minimizing the following objective functions

$$f_{\mathrm{B}}^{\mathrm{GRP}}(\mathbf{x}) = \sum_{t=2}^{n} \left(\frac{x_{t-1}}{x_t} - \frac{p_{t-1}}{p_t} \right)^2 \tag{3.1}$$

and

$$f_{\mathrm{B}}^{\mathrm{PFD}}(\mathbf{x}) = \sum_{t=2}^{n} \left(\frac{x_{t-1}}{p_{t-1}} - \frac{x_t}{p_t} \right)^2, \tag{3.2}$$

where subscript “B” stands for backwards. These objective functions are obtained from the forward objective functions by interchanging t and $t-1$. From now on, the minimization of (3.1) or (3.2) will be called “backward benchmarking”, as opposed to standard, forward benchmarking.

As mentioned above, a benchmarking method satisfies the time reversibility property if forward and backward benchmarking lead to the same results. It can be easily seen that $f_{\mathrm{F}}^{\mathrm{GRP}}(\mathbf{x}) \neq f_{\mathrm{B}}^{\mathrm{GRP}}(\mathbf{x})$, while $f_{\mathrm{F}}^{\mathrm{PFD}}(\mathbf{x}) = f_{\mathrm{B}}^{\mathrm{PFD}}(\mathbf{x})$. From this it follows that Denton PFD satisfies the time reversibility property, but GRP does not.

More practically, in many production processes “forward” benchmarking is applied, for example for the reconciliation of the Dutch Supply and Use tables (Bikker et al., 2013). However, after a revision, revised time series may be constructed “back in time”, by using backward objective functions. It is highly undesirable that there are any differences in outcomes that can be purely attributed to a difference in “time direction”. Practitioners who are unaware of the time reversibility property, may apply forward and backward benchmarking and mistakenly assume that both methods lead to the same results.

Although it is true that any benchmarking application can be restricted to preserving forward growth rates, it is undesirable that results are affected by the irrelevant property of time direction. Therefore, any benchmarking method should preferably satisfy time reversibility. Moreover, Subsection 3.3 illustrates that a benchmarking method that is not symmetric in time may change the timing of the most important economic events, e.g., the peaks and troughs that demark the start and end of a crisis.

3.2 Singularity

A second problem of GRP is the singularity of its objective function. If x_{t-1} approaches to zero in case of forward benchmarking (or x_t for backward benchmarking) the objective function value tends to infinity. This causes several problems.

One complication is that the optimization problem becomes unstable, a small change in preliminary values can lead to a large shift in benchmarked values. Consequently, undesirably large revisions can be obtained when benchmarking updated data.

Another complication is that, since a correction of near zero values can be heavily penalised, growth rates of such values are strongly preserved. This may however come at the expense of relatively large corrections of other growth rates. On the other hand, one may argue that growth rates do not contain much information for extremely small (close-to-zero) values. Hence, growth rate preservation can be deemed inappropriate in this case. Subsection 5.3 shows a real-life example of this problem.

A third complication is that, as close-to-zero benchmarked values may cause a large objective function value, GRP methods tend to avoid such values. Consequently, irregular correction patterns can be obtained. In particular, negative benchmarked values may be obtained for a problem in which all preliminary values are positive. Consider an example in which two consecutive values are 100. Then, an adjustment of the first value from 100 to -100 is less costly in terms of GRP's objective function value than a correction from 100 to 30. The corresponding objective function values are $((100/-100)-(100/100))^2 = 4$ and $((100/30)-(100/100))^2 = 5.44.$ A value that goes from a large positive to a large negative will however usually not be considered good movement preservation. Therefore, the example also demonstrates the questionability of the use of growth rates when positive and negative values occur.

For this reason, it can be advisable to avoid negative outcomes by inclusion of non-negativity constraints, see Subsection 4.1 for more details. For Denton PFD negative values are less likely obtained. In the previous example, an adjustment from 100 to 30 is preferred to an adjustment from 100 to -100. A real-life example of this problem is shown in Subsection 5.3.

Although singularity of GRP's objective function may trigger negative benchmarked values, it is not the only cause. Denton PFD may also yield negative values. In general, there is a risk of negative benchmarked values, when the (relative) change from one benchmark to another significantly differs from the (relative) change from the underlying annualised preliminary values.

A fourth complication of GRP's singular objective function is that irregular peaks and throughs may occur in a benchmarked time series. The explanation is that in standard GRP a correction of large positive value to a close-to-zero value is less costly in terms of the objective function value than an opposite correction from close-to-zero to a large positive. That is, a correction of a growth rate g with a factor $c,$ where $c > 1,$ corresponds to a larger objective function value than a correction with $1/c,$ especially if c is large. The objective function values are $((c-1)g)^2$ and $\left(\frac{(c-1)}{c}g\right)^2$ respectively. Since large upward corrections from a close-to-zero value are relatively costly, these are avoided as much as possible. Thus, the GRP's benchmarked values move more gradually from a close-to-zero value than Denton's results do. To compensate for this, larger peaks may be necessary for the following time-periods to fulfill the temporal aggregation constraint. As benchmarking usually aims at as smooth as possible corrections over time, irregular peaks can be considered undesirable. Related to the relatively slow growth from a close to zero value is that the peaks tend to turn up later in time than for a time-symmetric method like Denton PFD. For the backward variant of GRP the opposite occurs, benchmarked time series move relatively quickly from a close to zero value, which gives rise to relatively early peaks. The example in Subsection 3.3 illustrates this problem.

3.3 Example

Below we present an example that illustrates the problems of GRP methods. In this example, a time series consisting of 15 months is reconciled with five quarterly values. The monthly series is constant: each monthly value is 10. The quarterly values are: 80, 250, 80, 400 and 100, respectively. Figure 3.1 compares the results of Denton PFD, GRPF and GRPB.

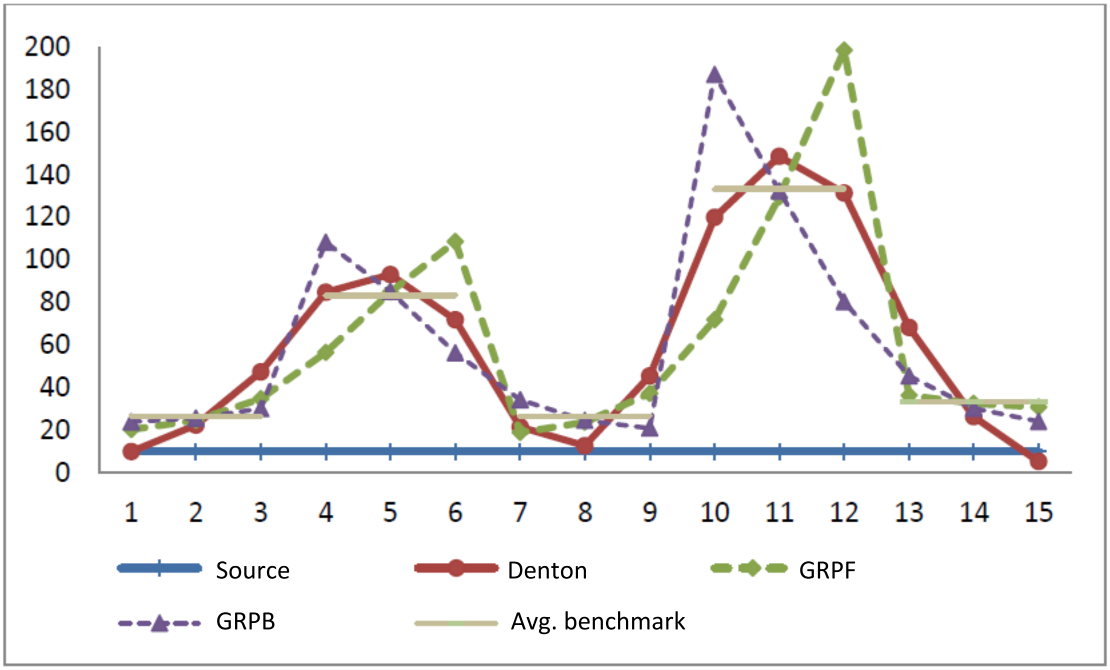

Figure 3.1 Example: Results of three benchmarking methods. "Avg. benchmark" stands for the average level of the monthly values that complies with the quarterly benchmarks and that is computed as one-third of its quarterly counterpart.

As the largest differences occur between both GRP methods, time reversibility is obviously not satisfied. The highest and lowest points appear at different months. The example clearly shows that the use of a different benchmarking method may lead to substantially different conclusions.

In accordance with Subsection 3.2, GRPF leads to relatively late peaks, i.e., at the last month of each quarter, while GRPB results in early peaks, i.e., at the first month of each quarter. Denton PFD's results are in between, peaks and troughs occur at the middle month of each quarter.

It needs however to be noted that the example cannot be considered representative for real life applications. In general, benchmarking methods are not meant to be used for reconciling large differences and for constant sub annual series. To explain the latter, a main assumption of Denton PFD is that the sub annual series provides information about short-term change. Constant series however cannot be considered very informative. Nevertheless, the problem of reconciling constant term series does occur in problems that are closely related to benchmarking, like interpolation and calenderization (see e.g., Dagum and Cholette, 2006 and Boot, Feibes and Lisman, 1967). The reason for choosing this example is purely educational. It provides good insight into properties of the different types of objective functions. The reader is referred to Subsection 5.3 for more realistic examples.

4 Alternative benchmarking techniques

In Section 3 we identified two problems with GRP methods. In this section we consider two alternative benchmarking techniques that solve the time irreversibility property.

4.1 Simultaneous growth rate preservation

Here, we propose two alternative objective functions for GRP. The first is a "time symmetric" variant of GRP, defined by

$$f_{\mathrm{S}}^{\mathrm{GRP}}(\mathbf{x}) = \frac{1}{2}\sum_{t=2}^{n}\left(\frac{x_t}{x_{t-1}} - \frac{p_t}{p_{t-1}}\right)^2 + \frac{1}{2}\sum_{t=2}^{n}\left(\frac{x_{t-1}}{x_t} - \frac{p_{t-1}}{p_t}\right)^2, \tag{4.1}$$

where subscript "S" stands for "simultaneous". The method will be called GRPS in the remainder of this paper. The GRPS objective function both preserves forward and backward growth rates. As far as the authors know this method has not been mentioned elsewhere in the literature. It can be easily seen that GRPS satisfies time reversibility: interchanging t and $t-1$ does not alter the objective function.

However, the second problem in Section 3 (singularity of objective function) is not considered. One of the consequences, negative benchmarked values, can be avoided by imposing lower bounds of zero on the benchmarked values. This can be done by including inequality constraints to an optimization problem, which is a well-established technique (e.g., Nocedal and Wright, 2006). The other problems related with singularity can however still occur.

4.2 Logarithmic growth rate preservation

Another "time symmetric" variant of GRP is given by the logarithmic form:

$$f_{\mathrm{L}}^{\mathrm{GRP}}(\mathbf{x}) = \sum_{t=2}^{n}\left[\log\left(\frac{x_t}{x_{t-1}}\right) - \log\left(\frac{p_t}{p_{t-1}}\right)\right]^2. \tag{4.2}$$

This function was firstly considered by Helfand, Monsour and Trager (1977). It is immediately verified that function (4.2) satisfies the time reversal property as well. The objective function can be considered the logarithmic version of GRP and equally well as the logarithmic version of Denton PFD. It will be denoted GRPL in the remainder of this paper, where "L" stands for "logarithmic".

Note that (4.2) can be used for strictly positive preliminary values only, and that it produces benchmarked values that are larger than zero as well. This does not seem an important limitation, as Section 3 already mentioned that growth rate preservation can be considered inappropriate for problems with positive and negative values. Nevertheless, a potential solution for time series with negative values is to add a sufficiently large constant to the series prior to benchmarking and subtract that constant from the benchmarked series. A potential drawback of this solution is that adding a constant distorts initial growth-rates. Thus, it is unclear whether preliminary growth rates are actually preserved. Further research is necessary to better understand the implications of this solution.

Although GRPL necessarily produces positive values, other problems in Section 3.2, related to a singular objective function can still occur.

4.3 Comparison

When comparing GRPS and GRPL, it can be expected that GRPL behaves more like Denton PFD. Below we will give two reasons for this.

Firstly, because of the asymptotic properties of the log function, the problem that close-to-zero values are avoided is less severe for GRPL than for GRPS. Close-to-zero values are associated with large adjustments of growth rates. Very large adjustments of growth rates are penalised less in GRPL than in GRPS, since GRPS's objective function grows faster when corrections are large.

Secondly, the first-order Taylor linearization of GRPL's objective function corresponds to Denton PFD's function, whereas the approximation of GRPS leads to a different result. The linearization of the squared terms of the objective function in the preliminary values are given by $\left(\frac{x_t}{p_t}\right)-\left(\frac{x_{t-1}}{p_{t-1}}\right)$ and $\left(\frac{p_t}{p_{t-1}}+\frac{p_{t-1}}{p_t}\right)\left\{\left(\frac{x_t}{p_t}\right)-\left(\frac{x_{t-1}}{p_{t-1}}\right)\right\}$ for GRPL and GRPS respectively.

4.4 Example

In order to explore the properties of GRPL and GRPS, we will consider the example of Subsection 3.3 again. Figure 4.1 compares results of the symmetric $f_{\mathrm{S}}^{\mathrm{GRP}}$, $f_{\mathrm{L}}^{\mathrm{GRP}}$ and $f^{\mathrm{PFD}}(\mathbf{x})$ methods.

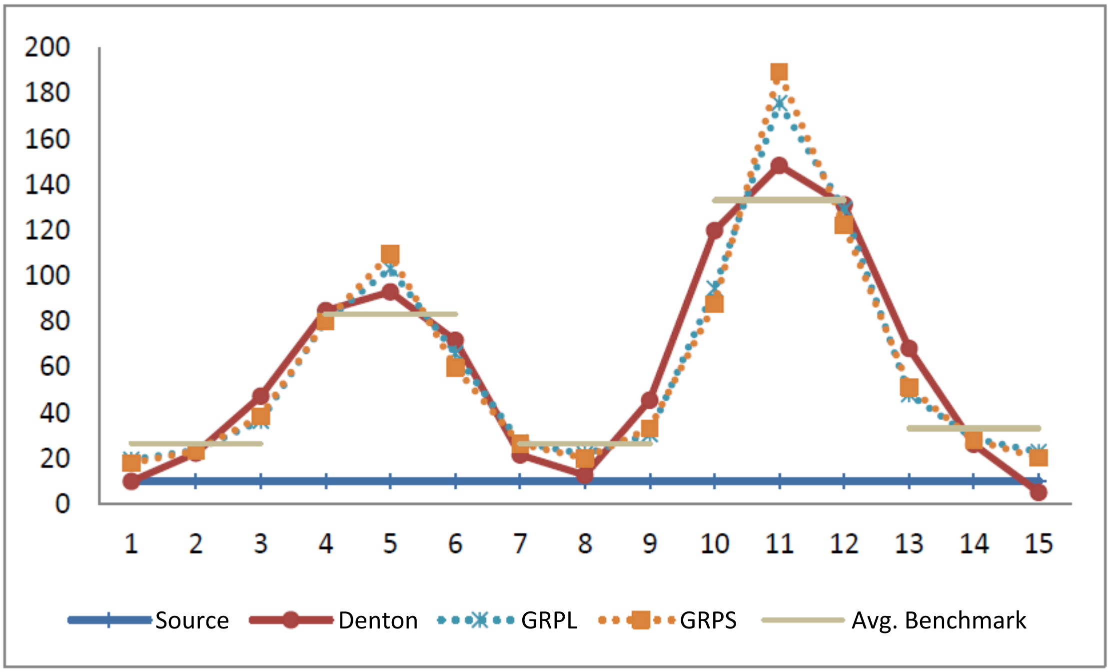

Figure 4.1 Example: results of three symmetric benchmarking methods. "Avg. Benchmark" stands for the average level of the monthly values that complies with the quarterly benchmarks and that is computed as one-third of its quarterly counterpart.

Firstly, it can be seen that the peaks and troughs occur at the same periods for all time symmetric methods.

Secondly, some of the drawbacks related to the singularity of the objective function still occur. When compared to Denton PFD, GRP methods tend to avoid close-to zero values, move away relatively slowly from low values (in both directions) and lead to irregular large peaks.

Thirdly, in accordance to Subsection 3.3, GRPL resembles Denton PFD more than GRPS, which follows from the slightly lower peaks of GRPL.

5 Empirical test

In this section an illustration exercise is conducted on real-life data, in order to find out whether or not the problems mentioned in Section 3 do occur in a realistic, practical application.

5.1 Data sets

The data set used for the illustration is obtained from quarterly and annual trade as published on the website of United Nations (UN).

The United Nations Commodity Trade Statistics Database (UN Comtrade) contains data from statistical authorities of reporting countries, or are received via partner organizations like the Organisation for Economic Co-operation and Development (OECD). The United Nations Totaltrade (UN Tottrade) data are mostly taken from the International Financial Statistics (IFS), published monthly by the International Monetary Fund (IMF). Differences between both sources emerge because of differences in data collection methods and purposes (United Nations, 2017). All data are publicly available at http://comtrade.un.org/.

We use UN Tottrade as data source for quarterly data and both UN Tottrade and UN Comtrade as sources for annual data. Both data sources include imports and exports for approximately 200 UN member states.

For our application all series were selected that include three annual totals and twelve quarterly values for 2002-2004. The variables of interest are total imports and exports. Series with quarterly or annual values smaller than 0.1 million dollars were deleted, as multiplicative benchmarking methods cannot be considered appropriate for zero or near zero values (see Subsection 3.2). Since the series are in million dollars, the cutoff value only excludes “extreme” cases and still leaves some real-life cases of singularity issues.

We end up with 238 time series for Comtrade and 253 series for Tottrade. The average year to year growth rates discrepancy between the annualized quarterly series and their benchmarks are 5.9%-point and 2.7%-point for Comtrade and Tottrade benchmarks, respectively. For the majority of series the discrepancy can be considered small. The percentage of series with a maximum discrepancy below 5%-point are 79% and 87%, respectively.

5.2 Results

Our first aim is to assess overall performance. We will compare the degree of preservation of the preliminary values and their growth rates for the various methods that are discussed in this paper.

Table 5.1 shows for the five methods the median values over all series, for the functions $f_{\mathrm{F}}^{\mathrm{GRP}}$, $f_{\mathrm{B}}^{\mathrm{GRP}}$, $f_{\mathrm{S}}^{\mathrm{GRP}}$ for forward, backward and simultaneous movement preservation and f^{Level} for preliminary value preservation. The latter function measures total squared relative adjustment, defined by

$$f^{\mathrm{Level}}(\mathbf{x}) := \sum_{t=1}^{n}\left(\frac{x_t}{p_t} - 1\right)^2. \tag{5.1}$$

Table 5.1
Median values of criteria in (2.1), (3.1), (4.1) and (5.1)

	COM data set				TOT data set			
	$f_{\mathrm{F}}^{\mathrm{GRP}}$	$f_{\mathrm{B}}^{\mathrm{GRP}}$	$f_{\mathrm{S}}^{\mathrm{GRP}}$	f^{Level}	$f_{\mathrm{F}}^{\mathrm{GRP}}$	$f_{\mathrm{B}}^{\mathrm{GRP}}$	$f_{\mathrm{S}}^{\mathrm{GRP}}$	f^{Level}
Denton PFD	0.87	0.88	0.88	26.42	0.33	0.41	0.37	2.07
GRPF	0.84	0.98	0.93	26.43	0.27	0.48	0.45	2.06
GRPB	1.00	0.82	0.91	26.47	0.48	0.28	0.45	2.07
GRPS	0.87	0.89	0.88	26.41	0.34	0.38	0.36	2.07
GRPL	0.87	0.88	0.88	26.42	0.33	0.41	0.37	2.07

The values for the COM and TOT data sets are $*10^{-2}$ and $*10^{-5}$, respectively.

It can be seen from Table 5.1 that the GRPF method, that is designed to preserve forward growth rates, results in relatively poor backward movement preservation. The opposite is also true: GRPB does not preserve forward movements very well. From these results, we can conclude that time reversibility actually matters. Table 5.1 also demonstrates that the time symmetric methods, Denton PFD, GRPS and GRPL, perform well on all measures and that difference between those methods are only marginal.

To assess forward, backward and simultaneous growth rate preservation, a relative criterion is used that compares the values of the objective functions $f_{\mathrm{F}}^{\mathrm{GRP}}(\mathbf{x})$, $f_{\mathrm{B}}^{\mathrm{GRP}}(\mathbf{x})$ and $f_{\mathrm{S}}^{\mathrm{GRP}}(\mathbf{x})$ with their optimum values, which are obtained from GRPF, GRPB and GRPS, respectively. Analogous to the standards in Di Fonzo and Marini (2012), movement preservation is consided acceptable if it lies within 10% of the optimum value. That is, if $f^{\text{ method}}(\mathbf{x}) / f^{\text{ optimum}}(\mathbf{x}) \leq 1.1,$ where f is one of the previously mentioned objective functions.

For the five methods considered, Table 5.2 shows the percentage of time series with acceptable forward, backward and simultaneous movement preservation.

Table 5.2
Percentage of time series with acceptable movement preservation

	COM data set			TOT data set		
	Forward	**Backward**	**Simult.**	**Forward**	**Backward**	**Simult.**
Denton PFD	79.4	78.6	95.8	79.4	79.4	96.0
GRPF	100.0	48.7	81.5	100.0	47.8	82.6
GRPB	47.1	100.0	76.9	44.3	100.0	75.1
GRPS	82.4	77.3	100.0	80.6	79.4	100.0
GRPL	79.8	79.0	96.6	79.4	79.4	96.0

For Denton PFD an acceptable degree of simultaneous movement preservation is found for more than 95% of all cases. Thus, one can conclude that Denton PFD can be considered as a very good approximation for the optimal GRPS method; the approximation is even better than the GRPF and GRPB methods, for which acceptable performance is found for around 80% of all cases.

So far, we focused on performance for entire time series. Below we will consider the occurrence of large and extreme reconciliation adjustments made to single values and growth rates.

To measure the adjustments made to growth rates, the absolute difference $|\, g_{it}(\mathbf{x}) - g_{it}(\mathbf{p}) \,| * 100\%$ is used, where g_{it} is a growth rate for series i and period t. Tables 5.3 and 5.4 compare the occurrence of large and extremely large adjustments to forward, backward and simultaneous growth rates.

Table 5.3
Percentage of large growth rate adjustments (> 10%-point difference)

	COM data set			TOT data set		
	Forward	**Backward**	**Simult.**	**Forward**	**Backward**	**Simult.**
Denton PFD	2.0	2.1	1.9	0.8	0.6	0.6
GRPF	1.9	2.4	2.3	0.6	0.9	0.7
GRPB	2.3	1.5	2.0	1.1	0.3	0.8
GRPS	1.9	1.9	1.8	0.8	0.6	0.6
GRPL	2.0	1.9	1.9	0.8	0.6	0.5

Table 5.4
Percentage of extreme growth rate adjustments (> 50%-point difference)

	COM data set			TOT data set		
	Forward	**Backward**	**Simult.**	**Forward**	**Backward**	**Simult.**
Denton PFD	0.3	0.2	0.4	0.1	0.1	0.1
GRPF	0.2	0.2	0.2	0.0	0.1	0.1
GRPB	0.3	0.0	0.2	0.2	0.0	0.1
GRPS	0.2	0.1	0.2	0.1	0.0	0.1
GRPL	0.2	0.1	0.2	0.1	0.0	0.1

These tables show minor differences between methods.

Small differences between methods are also in observed in Table 5.5, which shows large and extreme corrections to preliminary values, as measured by the relative criterion $\left(x_{\mathrm{it}} / p_{it}\right) * 100\%$.

Hence, one can conclude that the problems caused by singularity do not translate into more often occurring large corrections.

Table 5.5
Percentage of large adjustments to preliminary values

	COM data set			TOT data set		
	Large (>10%)	**Extreme (>100%)**	**Negative (<0%)**	**Large (>10%)**	**Extreme (>100%)**	**Negative (<0%)**
Denton PFD	13.2	1.0	0.0	5.8	0.4	0.0
GRPF	13.0	1.0	0.0	5.8	0.3	0.1
GRPB	13.1	0.9	0.0	5.6	0.3	0.0
GRPS	13.1	0.9	0.0	5.8	0.4	0.0
GRPL	13.0	0.9	0.0	5.8	0.4	0.0

Most remarkable in Table 5.5 are the negative benchmarked values obtained for GRPF in the TOT data. An example of this is illustrated in Figure 5.3.

Despite the similar results of the five benchmarking methods in Tables 5.3-5.5, there are clear differences in smoothness of reconciliation adjustments. To demonstrate this, we will use the smoothness indicator (Temurshoev, 2012).

$$\text{Smoothness} = \sum_{t=2}^{n-2} \left[\mathrm{BI}_t - \overline{\mathrm{BI}_t}\right]^2, \tag{5.2}$$

where BI_t is the so-called benchmark-to-indicator ratio, i.e., x_t / p_t and $\overline{\mathrm{BI}_t}$ is the 5-terms moving average $\frac{1}{5}\sum_{k=t-2}^{k=t+2} \mathrm{BI}_k$.

According to this indicator, we find in Table 5.6 that the smoothest results are obtained for Denton PFD and GRPL. Conversely, the asymmetric GRPF and GRPB methods yield the most irregular adjustments. It follows that the time-symmetric method GRPS, but most so GRPL, suffers less from singularity than the asymmetric methods GRPF and GRPB do. These results most clearly illustrate the problems with the singularity of GRP's objective function that were described in Subsection 3.2.

Table 5.6
Smoothness indicator values (5.2), summed over all series

	COM data set	TOT data set
Denton PFD	3.4	0.3
GRPF	9.8	39.0
GRPB	8.2	2.9
GRPS	4.3	1.1
GRPL	3.3	0.5

5.3 Examples

Below we show two examples to demonstrate that the problems in Section 3 do occur in a real-life application.

The first example, in Figures 5.1 and 5.2, illustrates that non-symmetric GRP methods may change the timing of the most important economic events. When considering the first nine quarters, the two highest values occur at different time periods. GRPF's peak periods are at quarters 6 and 7 and those of GRPB are at quarters 5 and 6. Closely related to this, is that GRPF moves away relatively slowly from the close-to-zero values at quarters 1-4.

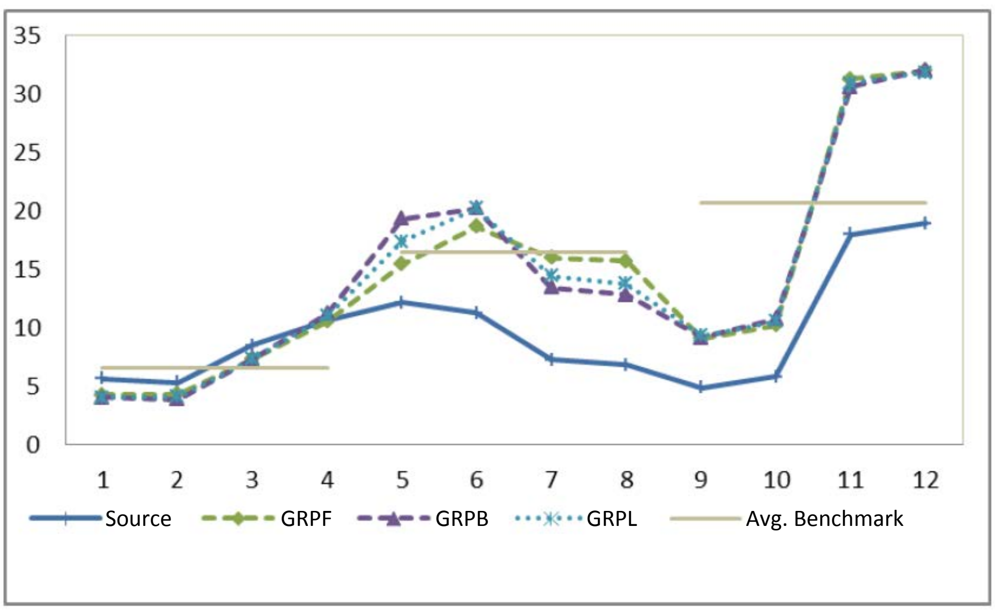

Figure 5.1 Exports Burundi, Comdata, 2002-2004, millions of US dollar. "Avg. Benchmark" stands for the average level of the quarterly data that complies with the annual benchmarks and that is computed as one-fourth of its annual counterpart.

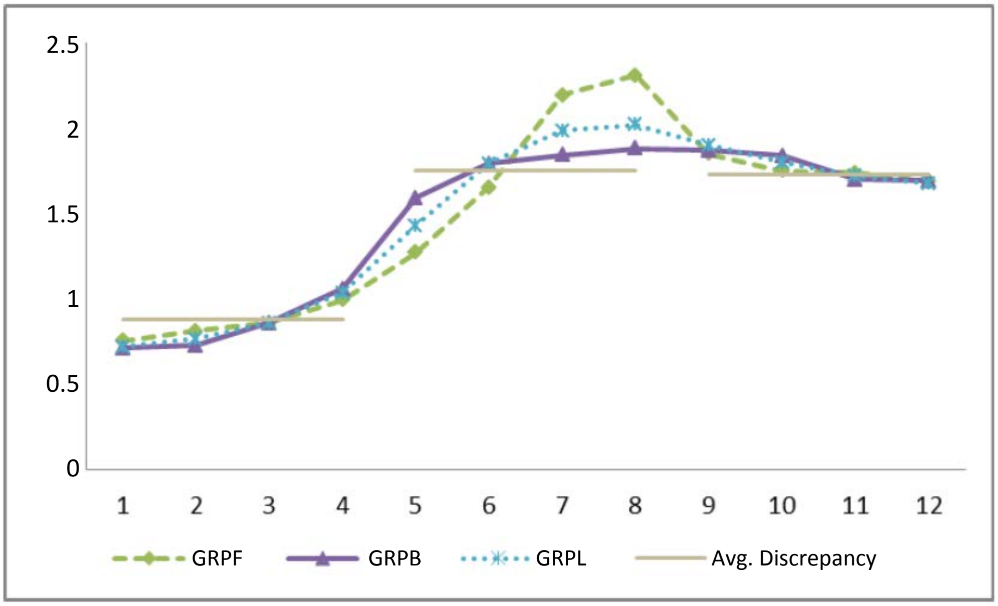

Figure 5.2 Benchmark to Indicator ratios, Exports Burundi, 2002-2004. "Avg. Discrepancy" stands for the annual BI-ratio, i.e., the ratio of an annual benchmark and the sum of the underlying quarterly indicators.

The second example illustrates the complications of a singular objective function. As shown in Figure 5.4, GRPF closely preserves growth rates of the quarters 6-10. This comes however at the expense of an irregular peak in quarter 5 and negative benchmarked values in the quarters 11 and 12.

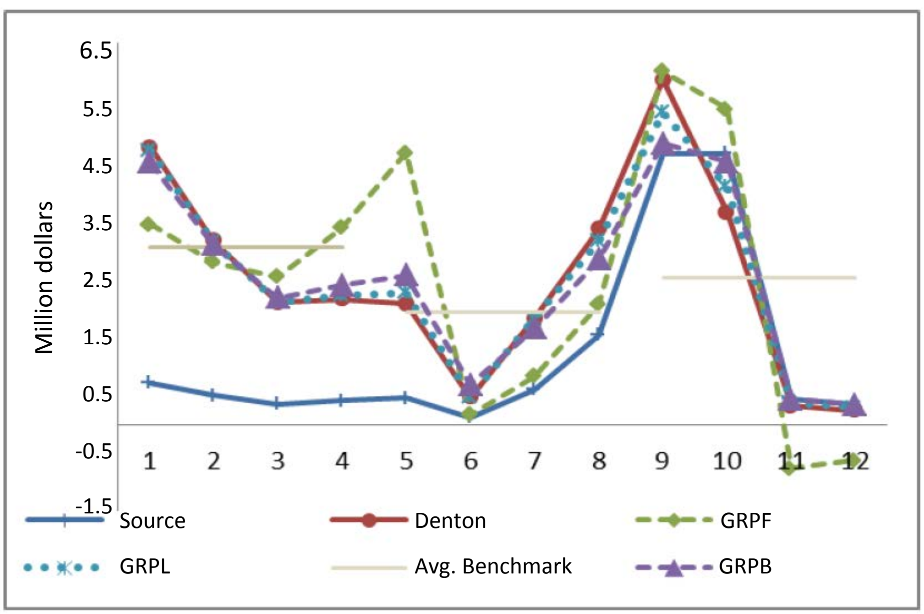

Figure 5.3 Exports Gambia, Totdata, 2002-2004, millions of US dollar. "Avg. Benchmark" stands for the average level of the quarterly data that complies with the annual benchmarks and that is computed as one-fourth of its annual counterpart.

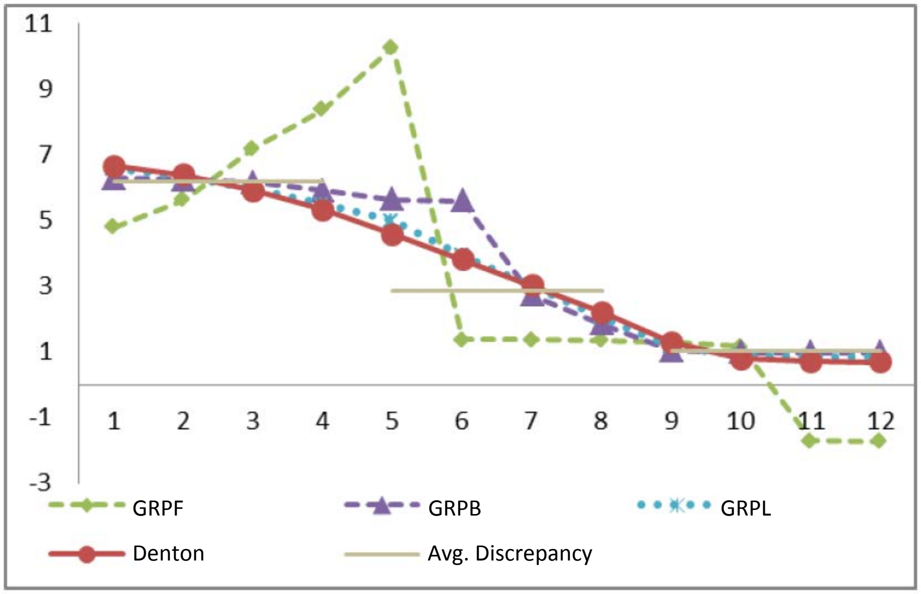

Figure 5.4 Exports Gambia, Totdata, 2002-2004, benchmark to indicator ratio. “Avg. Discrepancy” stands for the annual BI-ratio, i.e., the ratio of an annual benchmark and the sum of the underlying quarterly indicators.

6 Conclusions

Two well-known multiplicative benchmarking methods are Denton Proportionate First Differences (PFD) and Growth Rates Preservation (GRP). It is generally agreed that GRP has the strongest theoretical foundation. It better preserves initial growth rates than Denton PFD. However, from a technical point of view, Denton is the easiest method to apply. Because of this, and because Denton PFD is often a good approximation of GRP, Denton PFD is more popularly applied.

In this paper two drawbacks of GRP are demonstrated that, to the best knowledge of the authors, have not been mentioned elsewhere.

The first drawback is that GRP does not satisfy the time reversibility property. According to this property it should not matter for the results whether forward or backward growth rates are preserved. That is, benchmarking an original time series, $t = 1,\ldots,n,$ or a “reversed” time series, $t = n,\ldots,1$ should lead to the same benchmarked series. Since direction of time is irrelevant for any benchmarking application, any benchmarking method should preferably satisfy time reversibility. Moreover, a benchmarking method that does not satisfy time reversibility may yield entirely difficult conclusions on the timing of economic events depending on the chosen time direction. For these reasons forward and backward GRP methods should preferably be discouraged.

In this paper two alternative GRP methods are presented that do satisfy time reversibility. The first alternative, a new GRPS method, preserves both forward and backward growth rates. The other alternative, an existing GRPL method, preserves logarithms of the forward growth rates.

A second drawback of all GRP methods in this paper are the singularities in its objective functions. Complications of this are: avoidance of close to zero outcomes, irregular peaks in results and unnecessary negative values in benchmarked results.

These problems actually occurred in an illustrative application on real-life data. Although unnecessary negative values only occasionally occurred, reconciliation adjustments are much more irregular than for Denton PFD. Since smoothness of reconciliation adjustments (BI ratios) is often the main interest of benchmarking, asymmetric GRP methods can be discouraged for many applications.

While the literature considers Denton PFD "a good approximation" of the ideal GRP method, our main conclusion is that Denton PFD is even more appropriate than standard GRP for many applications. Denton is computationally easier to apply, it does not suffer from the problems related to time irreversibility and a singular objective function. Furthermore, the approximation of Denton PFD's results is even more close for the time-symmetric versions of GRP than for standard GRP.

However, when growth rate preservation is the key point of interest, a time-symmetric version of GRP can also be a good choice, most in particular GRPL. Time symmetric methods preserve growth rates slightly better than Denton PFD, satisfy time reversibility and suffer less severe from the drawbacks of a singular objective function than standard GRP.

Acknowledgements

The authors would like to thank three reviewers and the associate editor for useful comments that greatly contributed to improving the paper.

References

Bikker, R.P., Daalmans, J.A. and Mushkudiani, N. (2013). Benchmarking large accounting frameworks: A generalized multivariate model. *Economic Systems Research*, 25, 390-408.

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

Boot, J.C.G., Feibes, W. and Lisman, J.H.C. (1967). Further methods of derivation of quarterly figures from annual data. *Applied Statistics*, 16, 65-75.

Bozik, J.E., and Otto, M.C. (1988). Benchmarking: Evaluating methods that preserve month-to-month changes. Bureau of the Census - Statistical Research Division, RR-88/07, URL: http://www.census.gov/srd/papers/pdf/rr88-07.pdf.

Causey, B., and Trager, M.L. (1981). Derivation of Solution to the Benchmarking Problem: Trend Revision. Unpublished research notes, U.S. Census Bureau, Washington D.C. Available as an appendix in Bozik and Otto (1988).

Cholette, P.A. (1984). Adjusting sub-annual series to yearly benchmarks. *Survey Methodology*, 10, 1, 35-49. Paper available at https://www150.statcan.gc.ca/n1/pub/12-001-x/1984001/article/14348-eng.pdf.

Daalmans, J.A., and Di Fonzo, T. (2014). Denton PFD and GRP benchmarking are friends. An empirical evaluation on Dutch Supply and Use Tables. Paper presented to the IIOA-conference, Lisboa 14-18 July.

Dagum, E.B., and Cholette, P. (2006). *Benchmarking, Temporal Distribution and Reconciliation Methods for Time Series Data*. New York: Springer-Verlag.

Denton, F. (1971). Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization. *Journal of the American Statistical Association*, 66, 99-102.

Di Fonzo, T., and Marini, M. (2011). Simultaneous and two-step reconciliation of systems of time series: Methodological and practical issues. *Journal of the Royal Statistical Society C*, 60, 143-164.

Di Fonzo, T., and Marini, M. (2012). Benchmarking time series according to a growth rates preservation principle. *Journal of Economic and Social Measurement*, 37, 225-252.

Di Fonzo, T., and Marini, M. (2015). Reconciliation of systems of time series according to a growth rates preservation principle. *Statistical Methods & Applications*, 24, 651-669.

Eurostat (2013). *Handbook on Quarterly National Accounts – 2013 Edition*. http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-13-004.

Fisher, I. (1922). *The Making of Index Numbers*, Houghton-Mifflin, Boston.

Harvill Hood, C.C. (2005). An empirical comparison of methods for benchmarking seasonally adjusted series to annual totals. Paper presented at the Workshop on Frontiers in Benchmarking Techniques and their Applications to Official Statistics, Luxembourg, 7-8 April 2005.

Helfand, S.D., Monsour, N.J. and Trager, M.L. (1977). Historical Revision of Current Business Survey Estimates. *Proceedings of the Business and Economic Statistics Section*, American Statistical Association, 246-250.

Monsour, N.J., and Trager, M.L. (1979). Revision and benchmarking of business time series. *Proceedings of the Business and Economic Statistics Section*, American Statistical Association, 333-337.

Nocedal, J., and Wright, S. (2006). *Numerical Optimization, 2nd Edition*. New York: Springer.

Temurshoev, U. (2012). *Entropy-Based Benchmarking Methods*. University of Groningen. http://www.ggdc.net/publications/memorandum/gd122.pdf.

Titova, N., Findley, D. and Monsell, B.C. (2010). Comparing the Causey-Trager method to the multiplicative Cholette-Dagum regression-based method of benchmarking sub-annual data to annual benchmarks. *Joint Statistical Meetings Proceedings, Business and Economic Statistics Section*, 3007-3021.

Trager, M.L. (1982). Derivation of Solution to the Benchmarking Problem: Relative Revision. Unpublished research notes, U.S. Census Bureau, Washington D.C. Available as an appendix in Bozik and Otto (1988).

United Nations (2017). *2016 International Trade Statistics Yearbook. Volume I Trade by Countries*. New York: United Nations. https://comtrade.un.org/pb/downloads/2016/ITSY2016VolI.pdf.