ADJUSTING SUB-ANNUAL SERIES TO YEARLY BENCHMARKS

Pierre A. Cholette[1]

This paper proposes a modification to the method of Denton (1971) for adjusting sub-annual series to yearly totals. These totals originate from more reliable sources and constitute annual benchmarks. The benchmarked series derived according to the modified method is more parallel to the unbenchmarked series than this is the case with the original method. An additive and a proportional variant of the method are presented. These can easily be adapted for flow, stock and index series. Also presented are a few recommendations about the preliminary benchmarking of current data and the management of "historical" estimates of the series.

1. INTRODUCTION

In many cases, the statistician obtains sub-annual data of a series from one source of data (such as a sample survey); and, the corresponding annual benchmark values from another more reliable source of data (such as a census). The annual sums of the observed sub-annual values are generally not equal to the annual benchmark values. Such sub-annual series require adjustment to annual benchmarks, that is benchmarking.

The solution proposed by Denton (1971) (and generalized by Fernandez in 1981) consists of finding a sub-annual series which would display the movement of the available sub-annual series as much as possible and whose annual sums (or averages) would match the more reliable annual benchmarks. The level of the resulting series would then be given by the annual benchmarks, whereas its movement would be dictated by the original sub-annual series. In other words, the adjusted or benchmarked series should run as parallel as possible to the original, while still satisfying the annual benchmarks. This paper suggests a modification to Denton's specification which makes the original and the adjusted series even more parallel.

We follow the model of Ehrenberg (1982) for the presentation of scientific

[1] Pierre A. Cholette, Time Series Research and Analysis, Statistics Canada.

papers. The reader will be exposed to the illustrations and results first; and the methodological details, afterwards.

2. ILLUSTRATION OF THE RESULTS

Figure 1 shows the corrections $(x_t - z_t)$ made to the original series z_t according to the additive solution (with first differences) of Denton and according to the corresponding solution proposed in this paper. Since the corrections are to be added to the original sub-annual series z_t, the adjusted series x_t will be completely parallel to the original series, if and only if the corrections are constant. In the figure, this happens only for the corrections derived under the method proposed in this paper.

Figure 1 presented a trivial and ideal case which allowed the solution of constant corrections: All the average annual discrepancies, the differences between the annual benchmarks and the annual totals of the original series (divided by the number of months per year), were constant. Figure 2 displays a more realistic case, where the five average annual discrepancies vary about 200. As in the first example, the corrections derived by the herein proposed method are much more constant, especially in the first year.

As explained below, Denton's method does not only minimizes the change in the corrections (to make them as constant as possible) but also the size of the first correction. This can be seen both in Figures 1 and 2, where the first corrections are close to zero. The alternative solution, on the other hand, only minimizes the change in the corrections. Graphically this consists of fitting a curve through the average annual discrepancies, which is as flat as possible and which spans the same annual surfaces as the average annual discrepancies.

3. KEEPING THE ORIGINAL AND THE BENCHMARKED SERIES PARALLEL

Resuming the additive first difference formulation of Denton as well as his notation, the desired series x_t minimizes the following objective function

$$p(x) = \sum_{t=1}^{n} (\Delta x_t - \Delta z_t)^2 = \sum_{t=1}^{n} (\Delta(x_t - z_t))^2, \; x_0 = z_0, \qquad (1)$$

where z_t stands for the original sub-annual series at time t. This function is minimized subject to the equality constraints between the annual sums of the values obtained and the available benchmarks y_i:

$$\sum_{t=(i-1)k+1}^{ik} x_t = y_i, \; i = 1, 2, \ldots, m. \tag{2}$$

where k is the number of "months" per year.

Denton justifies hypothesis $x_0 = z_0$ claiming that it is legitimate to assume the equality of the last fitted and observed values prior to the estimation interval. Objective function (1) would then mean that the adjusted series x_t should have the same slope as the original series z_t; and therefore, that the slope of the differences between the two series should be minimized (subject to the constraints). However, after substituting $x_0 = z_0$, objective function (1) can be rewritten as:

$$p(x) = (x_1 - z_1)^2 + \sum_{t=2}^{n} (\Delta(x_t - z_t))^2. \tag{3}$$

This transformation emphasizes that the assumption $x_0 = z_0$ implies minimizing the size of the first correction. As illustrated in Figures 1 and 2, minimizing the first correction pulls the correction curve towards zero at the start of the series. This produces a wave in the first year which is transmitted to the other years. This wave in the corrections prevents, by definition, the maximum parallelism between the observed and adjusted series.

The specification proposed here simply refrains from postulating $x_0 = z_0$ and yields the following objective function

$$p(x) = \sum_{t=2}^{n} (\Delta(x_t - z_t))^2, \tag{4}$$

subject to the same constraints of equation (2).

In linear algebra, the constrained objective function is written

$$\underline{u}(\underline{x}, \underline{q}) = (\underline{x} - \underline{z})'\underline{A}\,(\underline{x} - \underline{z}) - 2\,\underline{q}'(\underline{y} - \underline{B}'\underline{x}), \tag{5}$$

where the vectors and matrices involved are:

$$\underset{n\times1}{\underline{x}} = \begin{vmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{vmatrix}, \quad \underset{n\times1}{\underline{z}} = \begin{vmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ \cdot \\ z_n \end{vmatrix}, \quad \underset{m\times1}{\underline{y}} = \begin{vmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_m \end{vmatrix}, \quad \underset{m\times1}{\underline{q}} = \begin{vmatrix} q_1 \\ q_2 \\ \cdot \\ \cdot \\ \cdot \\ q_m \end{vmatrix}, \tag{6}$$

$$\underset{n\times n}{\underline{A}} = \underline{D}'\underline{D}, \quad \underset{(n-1)\times n}{\underline{D}} = \begin{vmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{vmatrix}, \tag{7}$$

$$\underset{n\times m}{\underline{B}} = \begin{vmatrix} \underline{j} & 0 & \cdots \\ \underline{0} & \underline{j} & \cdots \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & \cdot & \end{vmatrix}, \quad \underset{k\times1}{\underline{j}} = \begin{vmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \end{vmatrix}, \ (n = km). \tag{8}$$

Vector $\underline{q}$ contains the Lagrangian multipliers. Variables n (= mk), m and k respectively stand for the number of observations and of years in the series and the number of months per year.

The normal equations associated with objective function (5) are

$$\begin{aligned} d\underline{u}/d\underline{x} &= (\underline{A} + \underline{A}')(\underline{x} - \underline{z}) + 2\,\underline{B}\,\underline{q} = \underline{0} \\ d\underline{u}/d\underline{q} &= 2(\underline{B}'\underline{x} - \underline{y}) = \underline{0} \end{aligned} \tag{9}$$

and yield solution

$$\begin{vmatrix} \underline{x} \\ \underline{q} \end{vmatrix} = \begin{vmatrix} \underline{A} & \underline{B} \\ \underline{B}' & \underline{0} \end{vmatrix}^{-1} \begin{vmatrix} \underline{A} & \underline{0} \\ \underline{0} & \underline{I} \end{vmatrix} \begin{vmatrix} \underline{z} \\ \underline{y} \end{vmatrix} = \underset{(n+m)\times(n+m)}{\underline{W}} \begin{vmatrix} \underline{z} \\ \underline{y} \end{vmatrix}. \tag{10}$$

Substituting identity y = B'z + r, where r contains the m annual discrepancies, gives

$$\begin{bmatrix} \underline{x} \\ \underline{q} \end{bmatrix} = \begin{bmatrix} \underline{A} & \underline{B} \\ \underline{B}' & \underline{0} \end{bmatrix}^{-1} \begin{bmatrix} \underline{A} & \underline{0} \\ \underline{B}' & \underline{I} \end{bmatrix} \begin{bmatrix} \underline{z} \\ \underline{r} \end{bmatrix} = \begin{bmatrix} \underset{n\times n}{\underline{I}} & \underset{n\times m}{\underline{W}_x} \\ \underset{m\times n}{\underline{0}} & \underset{m\times m}{\underline{W}_1} \end{bmatrix} \begin{bmatrix} \underline{z} \\ \underline{r} \end{bmatrix} \rightarrow \underline{x} = \underline{z} + \underline{W}_x \underline{r}. \qquad (11)$$

This reformulation of the solution reduces computing time in the application of the calculated weights compared to formulation (10). Also note that once the weights W_x are obtained, they can be used for any number of series having the same number of observations. Furthermore, we recommend (Cholette, 1978, section 6; 1979, 4.3) to compute W_x for a 5-year interval and to use it in a moving average manner (moving one year at the time) for series of 5 years and more. Apart from saving on calculations, this procedure generates only two revisions in the estimates (ceteris paribus) when new years of observations are added to the series.

Denton solves the inversion in equation (10) by parts. This is impossible here since matrix A is singular. The overall matrix however is not singular and can be inverted.

In fact, the method developped herein uses the solution proposed by Boot, Feibes and Lisman (1967) to interpolate between annual data in the absence of sub-annual information. Solution (11) exactly consists in interpolating between the annual discrepancies with the method of these authors and in adding the resulting estimates (the corrections) to the original sub-annual series.

4. PROPORTIONAL VARIANT

The proportional method now presented in this section is also a variant of Denton's proportional method, from which $x_0 = z_0$ was removed. As in Section 2, the objective function still minimizes the sum of the squared differences between the slopes of the original and desired sub-annual series (z_t and x_t). Each term in the sum is weighted however by the value of the corresponding sub-annual observation:

$$p(x) = \sum_{t=2}^{n} (\Delta(x_t - z_t)/z_t)^2 = \sum_{t=2}^{n} (\Delta(x_t/z_t))^2. \qquad (12)$$

This variant is suitable for series with strong seasonality, when it is thought that seasonal trough months cannot account for the annual discrepancy as much as seasonal peak months: The size of the corrections are proportional to the level of each observation, as illustrated in Figure 3. The low observations get smaller corrections than the seasonally higher observations, although the minimized proportional corrections x_t/z_t are as flat as permitted by the annual discrepancies. Note that with the proportional variant all observations must be positive and that all the adjusted values will also be positive.

It can also be shown (Cholette, 1978, Section 3; 1979, 3) that the proportional variant is a linear approximation of the strongly non-linear growth rate preservation method (Smith, 1977: Helfand et al., 1978), which would have the following objective function:

$$p(x) = \sum_{t=2}^{n} (x_t/x_{t-1} - z_t/z_{t-1})^2. \qquad (13)$$

The approximation is exact in situations of constant annual proportional discrepancies on the estimation interval.

In linear algebra, the constrained objective function associated to the proportional method is

$$\underline{u}(\underline{x}, \underline{g}) = (\underline{x} - \underline{z})' \underline{Z}^{-1}\underline{A}\, \underline{Z}^{-1}(\underline{x} - \underline{z}) - 2\, \underline{g}'(\underline{y} - \underline{B}'\underline{x}), \qquad (14)$$

where Z^{-1} is a diagonal matrix with elements $1/z_1$, $1/z_2$, ... The solution has the same structure as the additive variant ($Z^{-1}A\ Z^{-1}$ replacing A in (11)) and writes:

$$\begin{vmatrix} \underline{x} \\ \underline{g} \end{vmatrix} = \begin{vmatrix} \underline{Z}^{-1}\underline{A}\ \underline{Z}^{-1} & \underline{B} \\ \underline{B}' & \underline{0} \end{vmatrix}^{-1} \begin{vmatrix} \underline{Z}^{-1}\underline{A}\ \underline{Z}^{-1} & \underline{0} \\ \underline{B}' & \underline{I} \end{vmatrix} \begin{vmatrix} \underline{z} \\ \underline{r} \end{vmatrix} = \begin{vmatrix} \underline{I} & \underline{W_x} \\ \underline{0} & \underline{W_1} \end{vmatrix} \begin{vmatrix} \underline{z} \\ \underline{r} \end{vmatrix}. \qquad (15)$$

Unlike the weights in the additive variant however, weights W_x of the proportional solution must be computed for each series and even for each application interval of a given series.

5. STOCK AND INDEX SERIES

The additive and proportional variants of the method presented above are designed for flow series, whose annual values correspond to the sum of the sub-annual values. The solutions can very easily be adapted for stock series, whose annual values are associated to only one sub-annual value (usually that of the last month); and for index series, whose annual values correspond to the average of the sub-annual values. For a quarterly stock series, for instance, one merely has to redefine the component vector $\underline{j}$ of matrix $\underline{B}$ as

$$\underset{1\times 4}{\underline{j}'} = [0\ 0\ 0\ 1];$$

and, for monthly index series as

$$\underset{1\times 12}{\underline{j}'} = [1/12 \quad 1/12 \quad .... \quad 1/12].$$

6. DISCUSSION

6.1 Historical Data

There is a lot of confusion regarding the interpretation of assumption $x_0 = z_0$ of Denton. In that respect, the author writes: "It is assumed that no adjustments are to be made to the original series for years outside the range from year 1 to m, inclusive." (p. 100, above equation (3.2)).

If these years are left untouched because they never had any benchmarks, the solution proposed by Denton is defendable: No corrections result for years -1 and 0; and small and gradually introduced corrections, at the start of year 1. (Remember that $x_0 = z_0$ implies minimizing the first correction.) The resulting adjusted series is then continuous as illustrated in Figure 4 by curve ADEB.

However, if the first years are left untouched because they were already benchmarked and are now considered "historical", we do not agree with assumption $x_0 = z_0$. Indeed, this assumption will generally produce a discontinuity between years 0 and 1, as shown in Figure 4 by curve A'CDEB. Years -1 and 0 have already received corrections of magnitude around CD, whereas the start of year 1 receives corrections which are as small as possible.

In order to "freeze" the historical data after a certain number of years, two solutions are possible. First, one can explicitely specify the freezing constraint in the objective function which becomes

$$p(x) = ((x_1 - z_1) - (x_0 - z_0))^2 + \sum_{t=2}^{n} (\Delta (x_t - z_t))^2, \qquad (16)$$

where $(x_0 - z_0)$ is known and equal to the last correction used for historical year 0. This correction is generally not equal to zero (Cholette, 1979b, 1983). This specification amounts to determining the starting point of the correction curve.

Second, a less specific but equally effective solution consists of applying the methodology already proposed in this paper (additive or proportional versions) as a moving average, which moves one year at the time. With a 5-year estimation interval, for instance, the estimates automatically become final after two years of revision; and, after one year, in the case of a 3-year interval (Cholette, 1978, section 6 a; 1979, 4.3). The resulting benchmarked series is continuous, as illustrated in Figure 4 by curve A'CB.

6.2 Implementation

The practioners of benchmarking have a tendency to feed to the benchmarking programme the already benchmarked years of data followed by one year of unbenchmarked data (all accompanied by their benchmarks). For methodologists, it is obvious that one must always submit the unbenchmarked data (with the yearly benchmarks). Feeding benchmarked data will generally induce an artificial seasonal movement in the resulting benchmarked series (Cholette, 1978, Section 6b).

6.3 Preliminary Benchmarking of Current Data

A final comment is in order. During a current (uncompleted) year, one cannot calculate growth rates, for instance, between the benchmarked segment of the series (AB) and the unbenchmarked segment (CD). Doing so usually produces a discontinuity BC between the two segments AB and CD as illustrated in Figure 5 by curve ABCD.

Two solutions are then possible. One, the inter-temporal comparisons are based only on the unbenchmarked data. Two, the current data are preliminarily benchmarked by repeating the last available correction BC for the current year. (Note that including the incomplete current year in the objective function (4) (or 12) would yield identical preliminarily benchmarked values.) One can then compare the benchmarked segment AB with the preliminarily benchmarked segment BE as illustrated in Figure 5 by curve ABE. We favour this second alternative.

6.4 Relation with Other Methods

The Denton (1971) benchmarking method, the modified Denton method (presented in this paper), the methods of Glejser (1966), of Boot, Feibes and Lisman (1967), of Lisman and Sandee (1964), and of Bassie (1939) could be reffered to as univariate methods. No series other than that considered and its annual benchmarks enter the benchmarking process. On the contrary, the methods by Friedman (1962), by Chow and Lin (1971), by Somermeyer, Jansen and Louter (1976) and by Wilcox (1983) are multivariate. Auxiliary series are used in the computation of the desired series.

For instance, Chow and Lin (1971) proposed a method to obtain the desired sub-annual series from yearly totals and from related series. The movement of the resulting series is as much as possible similar to the movements of the related series (and the series obtained satisfies the annual constraints). Fernandez (1981) observes that the Chow and Lin method can produce movement discontinuities between the years. He then proposes a synthesis of the Chow-Lin and of the Denton methods. The combined method elimates the inter-annual discontinuities, but still relies on the hypothesis $x_0 = z_0$. As illustrated above, this hypothesis often introduces spurious fluctuations in the calculated series. We would think that it should be possible to refrain from the hypothesis in the case of Fernandez as in the case of Denton.

7. SUMMARY AND CONCLUSIONS

Denton (1971) intended to keep the original and benchmarked series as parallel as made possible by the annual discrepancies. This paper suggested a modification to the benchmarking method which makes the original and benchmarked series more parallel than is the case with the original method. This improvement holds both for the additive and the proportional variants of the method. We suspect that the generalized multivariate method by Fernandez could be improved in the same direction.

The method proposed can very easily be adapted for flow, stock as well as index series.

Before making intertemporal comparisons between the benchmarked and current data, it is essential to preliminarily benchmark the current data (in the manner proposed).

The suggested 5-year moving average implementation of the method will automatically "freeze" the past estimates after two years of revision.

REFERENCES

[1] Baldwin, A. (1978), "New Benchmarking Algorithms using Quadratic Minimization." National Product Division, Statistics Canada, Research Paper.

[2] Bassie, B.L. (1939), "Interpolation Formulae for the Adjustment of Index Numbers," Proceedings of the Annual Meetings of the American Statistical Association

[3] Boot, J.C.G., Feibes, W., Lisman, J.H.C. (1967), "Further Methods of Derivation of Quarterly Figures from Annual Data," Applied Statistics, Vol. 16, No. 1, pp.65-75.

[4] Cholette, P.A. (1978), "A Comparison and Assessment Various Adjustment Methods of Sub-Annual Series to Yearly Benchmarks," Time Series Research and Analysis, Statistics Canada, Research Paper 78-03-001B.

[5] Cholette, P.A. (1979a), "Adjustment Methods of Sub-Annual Series to Yearly Benchmarks," Proceedings of the Computer Science and Statistics, 12th Annual Symposium on the Interface, J.F. Gentleman Ed., University of Waterloo, pp. 358-36.

[6] Cholette, P.A. (1979b), "A Note on 'Freezing' Past Estimates when Benchmarking," Time Series Research and Analysis, Statistics Canada, Research Paper 79-06-002E.

[7] Cholette, P.A. (1982), "Minimum Quadratic Adjustment Program (MQAP-I) of Series to Annual Totals - Users Manual," Time Series Rsearch and Analysis, Statistics Canada, 82-11-003B.

[8] Cholette, P.A. (1983), "Benchmarking Series with Bi-Annual Benchmarks when Knowing the Ending Point," Time Series Research and Analysis, Statistics Canada, Research Paper 83-05-002B.

[9] Chow, G.C., Lin, An-loh (1971), "Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series," Review of Economics and Statistics, Vol. 53, No. 4, pp. 372-375.

[10] Dagum, E.B. (1977), "Comparison of Various Interpolation Procedures for Benchmarking Economic Time Series," Time Series Research and Analysis, Statistics Canada, Research Paper 77-05-006E.

[11] Denton, F.T. (1971), "Adjustment on Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization," J.A.S.A., Vol. 66, No. 333, pp. 99-102.

[12] Fernandez, R.B. (1981), "A Methodological Note on the Estimation of Time Series," Review of Economic and Statistics, Vol. 63, pp. 471-476.

[13] Friedman, M. (1962), "The Interpolation of Time Series by Related Series," J.A.S.A., Vol. 57, No. 300, pp. 729-757.

[14] Glejser, H. (1966), "Une méthode d'évaluation de données mensuelles à partir d'indices trimestriels ou annuels," Cahiers Economiques de Bruxelles, No. 19, 1er trimestre, pp. 45-64.

[15] Helfand, S.D. Monsour, N.J, Trager, M.L. (1978), "Historical Revision of Current Business Survey Estimates," U.S. Bureau of the Census, (Research Paper).

[16] Huot, G. (1975), "Quadratic Minimization of Montlhly Estimates to Annual Totals," Time Series Research and Analysis, Statistics Canada, Research Paper 75-11 010E.

[17] Lisman. J.H.C., Sandee, J. (1964), "Derivation of Quarterly Figures from Annual Data," Applied Statistics, Vol. 13, No. 2, pp. 87-90.

[18] Smith, P. (1977), "Alternative Method for Step Adjustment," Current Economic Analysis Division, Statistics Canada, Research Paper.

[19] Somermeyer, W.H, Jansen, R., Lauter, A.S. (1976), "Estimating Quarterly Values from Annually Known Variables in Quarterly Relationships," J.A.S.A, Vol. 71, No 355, pp. 588-595.

[20] Wilcox, J.A. (1983), "Disaggregating Data Using Related Series," Journal of Business and Economic Statistics, Vol. 1, No 3, pp. 187-191.

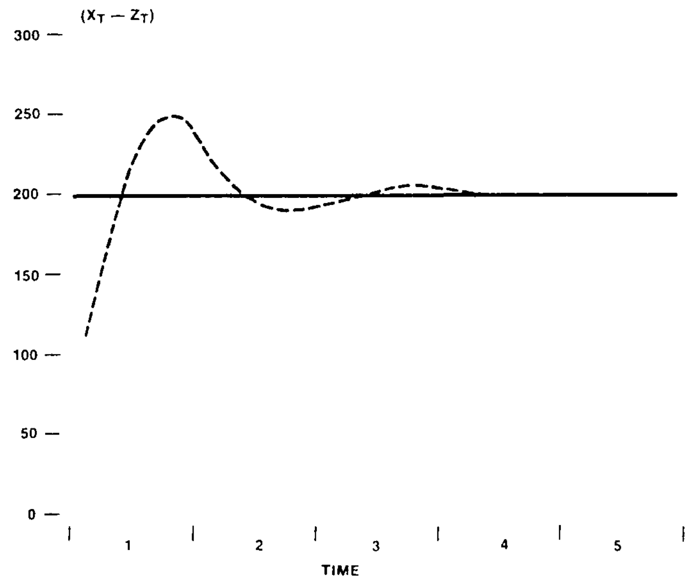

Figure 1: Corrections $(x_t - z_t)$ made to the unbenchmarked series according to Denton's method (dashed line) and according to the method proposed in this paper (solid) in an ideal situation of constant annual discrepancies.

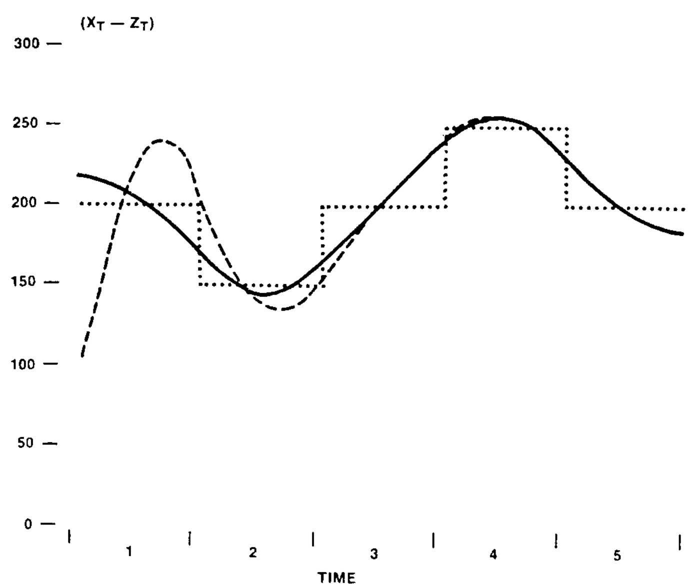

Figure 2: Corrections $(x_t - z_t)$ made to the unbenchmarked series according to Denton's method (dashed line) and according to the benchmarking method proposed in this paper (solid) in a situation of variable average annual discrepancies (dotted).

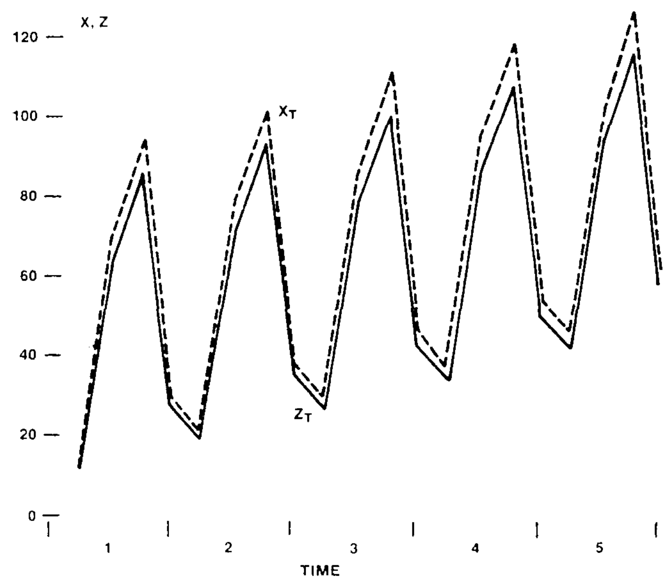

Figure 3: Original series (solid curve) and benchmarked series (dashed) according to the proportional variant of the benchmarking method proposed in this paper (in a situation of constant annual proportional discrepancies).

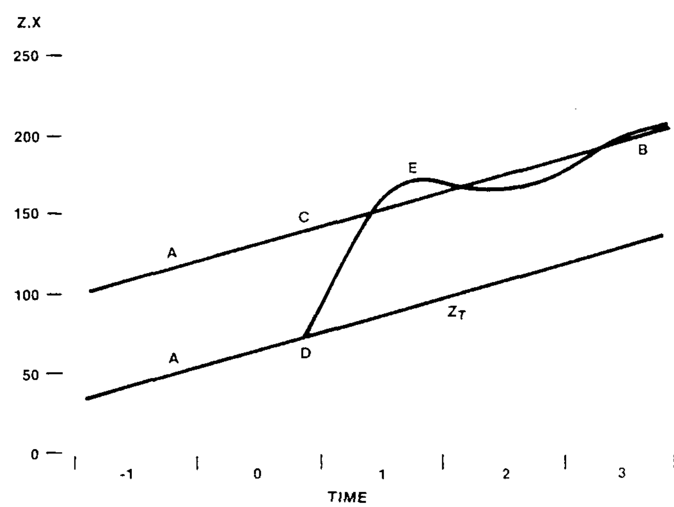

Figure 4: Benchmarked series according to Denton's method, when there are no benchmarks for year -1 and 0 (curve ADEB) and when there are benchmarks and year -1 and 0 were already benchmarked (A'CDEB): and according to the method proposed in this paper, applied in a moving average manner, when there are benchmarks for years -1 and 0 (A'B).

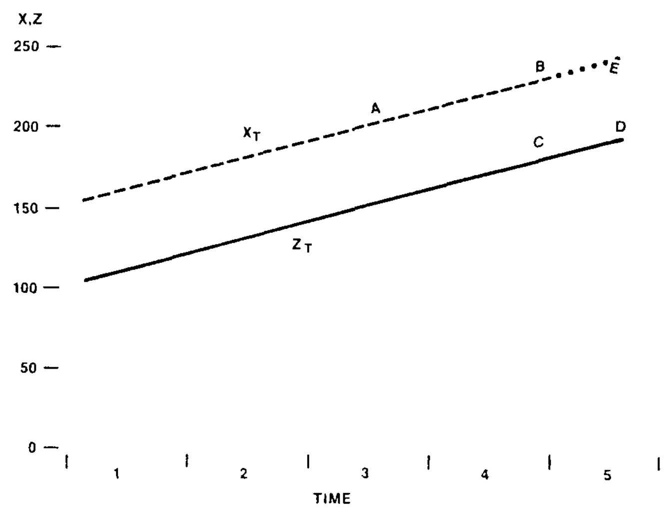

Figure 5: Continuity between the benchmarked series (dashed curve) and the preliminarily benchmarked series (dotted) and discontinuity BC between the benchmarked (dashed) and the unbenchmarked (solid) series.