# On Economic Data Reconciliation Using Some Benchmarking Models

**Isaac O. Ajao[1]; Femi J. Ayoola[2]; Joseph O. Iyaniwura[3]**

[1]Department of Mathematics and Statistics,
The Federal Polytechnic, Ado-Ekiti,
Ado-Ekiti, Nigeria.
e-mail: isaacoluwaseyiajao@gmail.com

[2.3]Department of Statistics,
University of Ibadan,
Ibadan, Nigeria.
e-mail: [2]ayoolafemi@yahoo.com; [3]jo_iyaniwura@yahoo.com

***Abstract*—Most high frequency series collected by statistical agencies are either monthly or quarterly, and may be subject to survey error. Annual values, otherwise known as benchmarks, on the other hand are more reliable, therefore there is need for adjustment, correction and reconciliation by statistical agencies in order to arrive at consistent values. Benchmarking is the process of adjusting the quarterly series to be consistent with the annual values. In this paper, various reconciliation procedures comprising Denton's additive level difference (ALD), additive first difference (AFD), and proportional first difference (PFD) were employed in order to rid of the discrepancies often noticed between quarterly and annual data values. Using economic data from IMF (2015), the results obtained are tested using some descriptive statistics, charts, and Benchmarking-to-Indicator (BI) ratios. It was discovered that PFD technique of reconciliation in solving benchmarking problem is better.**

***Keywords:*** *Reconciliation, Benchmarking, Additive methods, Proportional methods, BI ratio*

## I. Introduction

Benchmarking deals with the problem of combining a series of high-frequency data (e.g., quarterly data) with a series of less frequent data (e.g., annual data) for a certain variable into a consistent time series. The problem arises when the two series show different levels and movements and need to be made temporally consistent. Because low-frequency data are usually more comprehensive and accurate than high-frequency ones; the high frequency series is benchmarked to the low-frequency data. A common problem faced by official statistical agencies is the adjustment of monthly or quarterly time series which have been obtained from sample surveys to make them consistent with more accurate values obtained from other sources. These values can be aggregates or individual values at arbitrary points along the series (Durbin and Quenneville, 1997). The sources can be censuses, more accurate sample surveys, administrative data or combination of these. This adjustment process is called benchmarking and the more accurate values are called benchmarks. Typically, benchmarks are either yearly totals or values observed at a particular time-point each year. For simplicity, we assume that the data are monthly and the benchmarks are annual, though the broader interpretation should be borne in mind. In time series literatures, this problem, involving either one or many time series, is generally known as temporal disaggregation (Ajao *et al.*, 2015; Friedman, 1962; Chow and Lin, 1971; Ginsburgh,1973; Bournay and Laroque, 1979; Fernandez, 1981; Rossi, 1982; Litterman,1983; Al-Osh, 1989; Di Fonzo, 1990, 2002, 2003; Guerrero, 1990, 2003, 2005; Wei and Stram, 1990; Guerrero and Martinez, 1995; Salazar, Smith, and Weale, 2004; Proietti, 1998, 2005; Cabrer and Pavia, 1999; Guerrero and Nieto, 1999; Harvey and Chung, 2000; Hodgess and Wei, 2000; Santos Silva and Cardoso, 2001; Casals, Jerez, and Sotoca, 2005; Mitchell, Smith, Weale, Wright,and Salazar, 2005) or benchmarking of time series (Denton, 1971; Helfand, Monsour, and Trager, 1977; Cholette, 1984, 1987, 1988; Bozik and Otto, 1988; Hillmer and Trabelsi, 1987; Taillon, 1988; Laniel and Fyfe, 1989;

Trabelsi and Hillmer, 1990; Chen, Cholette, and Dagum, 1997; Chen and Dagum, 1997; Durbin and Quenneville, 1997; Dagum, Cholette, and Chen, 1998; Di Fonzo and Marini, 2005; Quenneville and Rancourt, 2005), where the latter definition (benchmarking) naturally matches with the philosophy of the data reconciliation procedures.

## II. Research Methodology

### 1.1 Formulation of the Problem and a General Approach to its solution

Assume that we are concerned with the intra-annual time periods of which there are $k$ per year, $k$ being an integer. Let the time series of interest cover $m$ years and consist of $n = mk$ values. The original values are represented in column-vector form by $s = [s_1\ s_2\ .\ .\ .\ s_n]'$. Assume also we have, from a different source, a set of $m$ annual totals represented by $a = [a_1\ a_2 \ldots a_n]'$. The problem is to adjust the original vector $z$ to obtain a new vector $\theta = [\theta_1\ \theta_2 \ldots \theta_n]'$ by a method which (a) minimizes the distortion of the original series, in some sense, and (b) satisfies the condition that the $k$ values of the new series within each year sum to the given annual total for that year. More formally, we specify a penalty function, $p(\theta, s)$, and express the problem as that of choosing $\theta$ so as to minimize $p(\theta, s)$, subject to:

$$\sum_{(T-1)k+1}^{Tk} \theta = a_T \qquad \text{(for} \quad T=1,2,\ldots,m) \qquad \ldots (1)$$

Consider the class of penalty functions represented by $(\theta - s)' A (\theta - s)$, a quadratic form in the differences between the original and adjusted time-series values, $A$ being a symmetric–$(n \times n)$ non-singular matrix to be specified later. We set up a Lagrangian expression and write:

$$u = (\theta - s)' A (\theta - s) - 2\lambda' (a - B'\theta) \qquad \ldots (2)$$

where

$$\lambda = [\lambda_1\ \lambda_2\ .\ .\ .\ \lambda_m]$$

and

$$B = \begin{bmatrix} j & 0 & . & . & . & 0 \\ 0 & j & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & j \end{bmatrix} \qquad \ldots (3)$$

$j$ being a $k$-dimensional column vector in which each element is unity and ***0*** being a $k$- dimensional null column vector. $B$ is $n\ X\ m$. The penalty minimizing solution is obtained by taking partial derivatives of $u$ with respect to the elements:–$x$ and $\lambda$, and then equating the resultant to zero, and solving for $x$ and $\lambda$. For convenience, we write $r = (a - B'\theta)$ for the vector of discrepancies between the two sets of annual totals and express the solution in the form:

$$\begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} A & B \\ B' & 0 \end{bmatrix} \times \begin{bmatrix} A & 0 \\ B' & I \end{bmatrix} \times \begin{bmatrix} z \\ r \end{bmatrix} \qquad \ldots (4)$$

Where $I$ is the (m×m) identity matrix and 0 is the (m×m) null matrix. (It is assumed, of course, that the second-order conditions necessary for the solution to be a minimum are satisfied.) Using a well-known result for deriving the inverse of a partitioned matrix, the solution for $\theta$ is then found to be $\theta = s + Cr$, where $C = A^{-1}B(B' A^{-1} B)^{-1}$. Thus the adjusted values are equal to the original values plus linear combinations of the discrepancies between the two sets of annual totals.

### 2.2 The Original and Modified Additive First Difference Variants of the Denton Method

The underlying model of the Denton benchmarking method is the following:

$$s_t = \theta_t + e_t \qquad \ldots (5)$$

$$a_m = \sum_{t=t_{1,m}}^{t_{L,m}} j_{m,t} \theta_t \qquad \ldots (6)$$

Where $s_t$ and $\theta_t$ are respectively the observed sub-annual series and the benchmarked series. The behaviour of the error is specified by an objective function. The objective function of the original additive first difference Denton method is:

$$\min(\theta_1 - s_1)^2 + \sum_{t=2}^{T}[(\theta_t - s_t) - (\theta_{t-1} - s_{t-1})]^2 \quad \ldots(7)$$

$$\min(\theta_1 - s_1)^2 + \sum_{t=2}^{T}[(\theta_t - \theta_{t-1}) - (s_t - s_{t-1})]^2 \quad \ldots(8)$$

Denton imposes the constraint $(\theta_0 = s_0)$. This Denton initial condition forces the benchmarked series to equal the original series at time zero and results in the minimization of the first correction $(\theta_1 - s_1)$. Without the Denton initial condition (first term), eqn. (8) specifies that the difference between the benchmarked and the original series $(\theta_t - s_t)$ must be as constant as possible through time. The modified first difference variant precisely omits this first term to solve the short-coming in the original model.

In both the original and modified variants, an objective function is minimized subject to the benchmarking constraints (6). The objective function can be written in matrix algebra as:

$$f(\theta,\gamma) = (\theta - s)' D' D(\theta - s) - 2\lambda'(a - J\theta) \quad \ldots(9)$$

$$= \theta' D' D\theta - 2\theta' D' Ds + s' D' Ds - 2\lambda' J\theta \quad \ldots(10)$$

Where $a$ stands for the benchmarks, $s$ and $\theta$ respectively denote the observed original series and benchmarked series, and $\gamma$ contains the Lagrange multipliers associated with the linear constraints $a - J\theta = 0$. Matrix $J$ is the temporal sum operator.

The necessary conditions for optimization require that the derivative of the objective function (10) with respect to the parameters be equal to zero:

$$\frac{\partial(f(\theta,\gamma))}{\partial\theta} = 2D' D\theta - 2D' Ds + 2J'\gamma = 0 \quad \ldots(11)$$

$$\frac{\partial(f(\theta,\gamma))}{\partial\theta} = 2J\theta - 2a = 0 \quad \ldots(12)$$

The sufficient condition for a minimum is that the matrix of the second order derivatives $\frac{\partial(f(\theta,\gamma))}{\partial\theta^2}$ be positive definite. Matrix $A = D' D$ is indeed positive definite because it is of the form $A = B' B$

$$D' D\theta + J'\gamma = D' Ds \quad \ldots(13)$$

$$J\theta = a \equiv Js + (a - Js) \quad \ldots(14)$$

Where $a$is replaced by $Js + (a - Js)$ which provides a more convenient result. Eqn. (14) may be expressed as:

$$\begin{bmatrix} D'D & J' \\ J' & 0 \end{bmatrix} \times \begin{bmatrix} \theta \\ \gamma \end{bmatrix} = \begin{bmatrix} D'D & 0_{T\times M} \\ J' & I_M \end{bmatrix} \times \begin{bmatrix} s \\ (a - Js) \end{bmatrix} \quad \ldots(15)$$

Eqn. (15) re-arranged we have

$$\begin{bmatrix} \theta \\ \gamma \end{bmatrix} = \begin{bmatrix} D'D & J' \\ J' & 0 \end{bmatrix}^{-1} \times \begin{bmatrix} D'D & 0_{T\times M} \\ J' & I_M \end{bmatrix} \times \begin{bmatrix} s \\ (a - Js) \end{bmatrix} \quad \ldots(16)$$

The inversion by parts in (16) leads to the following alternative

$$\hat{\theta} = s(D'D)^{-1} J' \left(J(D'D)^{-1} J'\right)^{-1}(a - Js) \quad \ldots(17)$$

**2.3 The Benchmark-to-indicator (BI) ratio framework**

This is for converting individual indicator series into estimates of individual quarterly variables. To understand the relationship and reconciliation between the corresponding annual and quarterly data, it is useful to observe the ratio of the annual benchmark to the sum of the four quarters of the indicator (the annual BI ratio). Movements in the observed annual BI ratio show inconsistencies between the long-term movements in the indicator and in the annual data. The relationship between the annual data and the quarterly indicator can be assessed by looking at the movements of the annual BI ratio, namely the ratio of the annual benchmark to the sum of the four quarters of the indicator. The annual BI ratio can be expressed as:

$$\frac{A_n}{\sum_{t=4n-3}^{4n} I_t} \qquad \text{for } n = 1, \ldots, y$$

where
$A_n$ is the annual target variable for year $n$
$\sum_{t=4n-3}^{4n} I_t$ is the annual sum of the quarterly observations of the indicator $I$ for the same year $n$,
$y$ is the last available year

When the BI ratio changes over time, it signals different patterns between the indicator and the annual data; instead, a constant annual BI ratio mean that the two variables present the same rates of change. As a result, movements in the annual BI ratio can help identify the quality of the indicator series in tracking the movements of the annual variables over the years.

## III. ANALYSIS AND RESULTS

Table 1:Annual Import data

| Year | Annual data |
|---|---|
| 2009 | 944 |
| 2010 | 957 |
| 2011 | 980 |
| 2012 | 1031 |
| 2013 | 1088 |
| 2014 | 1136 |
| 2015 | 1127 |

Source: IMF, 2015

Table 2: Quarterly data

| Quarters | Indicators | Quarters | Indicators |
|---|---|---|---|
| 2009Q1 | 77.09 | 2013Q1 | 86.71 |
| 2009Q2 | 86.16 | 2013Q2 | 96.76 |
| 2009Q3 | 82.02 | 2013Q3 | 95.27 |
| 2009Q4 | 88.72 | 2013Q4 | 103.22 |
| 2010Q1 | 75.86 | 2014Q1 | 93.12 |
| 2010Q2 | 87.4 | 2014Q2 | 101.63 |
| 2010Q3 | 84.41 | 2014Q3 | 98.54 |
| 2010Q4 | 89.77 | 2014Q4 | 106.71 |
| 2011Q1 | 76.79 | 2015Q1 | 94.35 |
| 2011Q2 | 88.89 | 2015Q2 | 101.74 |
| 2011Q3 | 85.49 | 2015Q3 | 97.4 |
| 2011Q4 | 92.86 | 2015Q4 | 104.2 |
| 2012Q1 | 82.78 | | |
| 2012Q2 | 91.35 | | |
| 2012Q3 | 89.98 | | |
| 2012Q4 | 96.8 | | |

Source: IMF, 2015

By implementing equation (17)

$$\hat{\theta} = s\left(D^{'}D\right)^{-1} J^{'}\left(J\left(D^{'}D\right)^{-1} J^{'}\right)^{-1}\left(a - Js\right)$$

Using MATLAB 7.10, the following results are obtained and summarized in table 3

Table 3: Assessment of the Discrepancies using the BI ratios

| | ALD | AFD | PFD |
|---|---|---|---|
| Mean | 2.8475 | 2.8467 | 2.8414 |
| Median | 2.8344 | 2.8304 | 2.8408 |
| Maximum | 3.0703 | 3.0416 | 2.8574 |
| Minimum | 2.71 | 2.7256 | 2.8247 |
| Std. Dev. | 0.1095 | 0.0972 | **0.0101** |
| Standard Error | 0.0207 | 0.0184 | **0.0019** |
| Skewness | 0.6325 | 0.5791 | **0.0619** |
| Kurtosis | -0.622 | -0.8212 | -1.0962 |
| Range | 0.3603 | 0.316 | **0.0327** |

The result of the descriptive statistics of the BI ratios in table 6 shows that the PFD has the least values in Standard deviation, standard error, skewness, and range. The standard deviations and standard errors for PFD BI ratios are 0.0101 and 0.0019 respectively, lower than values from ALD and AFD, this means the estimated benchmarks obtained through PFD is more stable than the other two methods. These tell us that PFD method of benchmarking performs better than ALD and AFD in term stability of the obtained data.

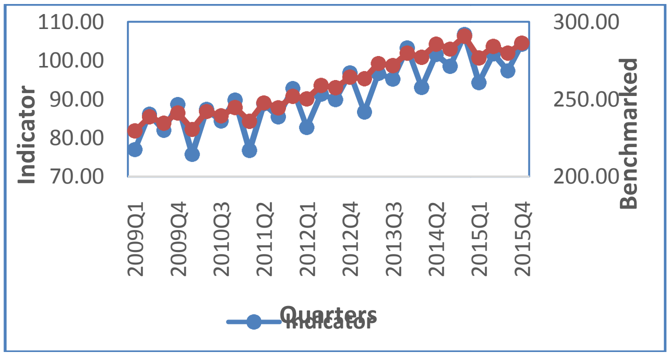


Fig. 1: The indicator and the estimated Benchmarked series using ALD method

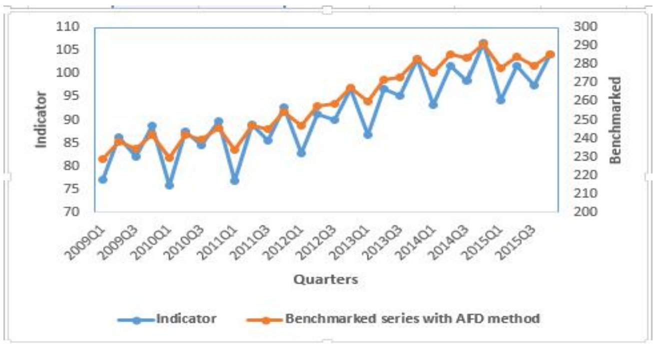


Fig. 2: The Indicator and the estimated Benchmarked series using AFD method

The two graphs (Figs. 1 and 2) show that there are some inconsistencies in the between the indicator and the benchmarked series produced by the ALD and AFD methods. It seems to be higher in the series obtained using the ALD method.

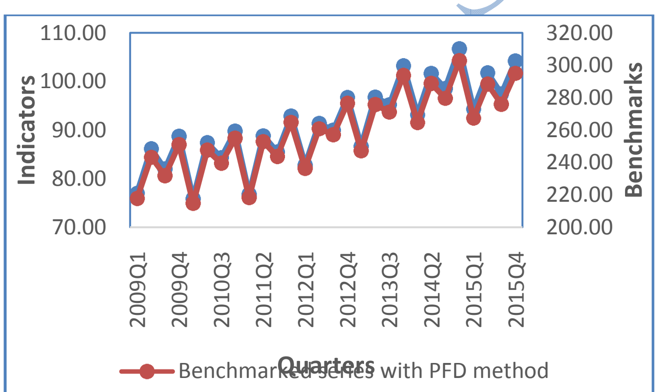


Fig. 3: The Indicator and the estimated Benchmarked series using PFD method

From Fig. 3, PFD method seems to be better with the level of consistency observed between the indicator and the benchmarked series obtained both at the backward and forward series

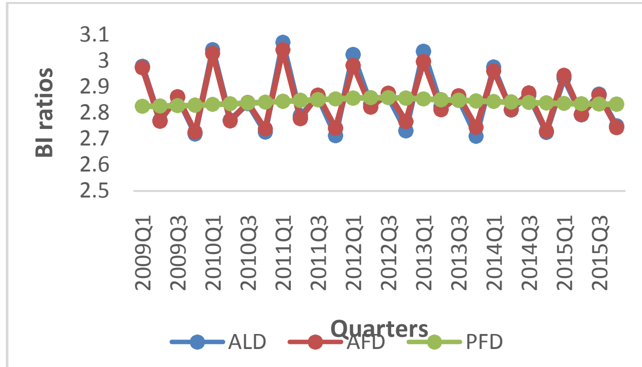


Fig. 4: Benchmark-to-Indicator ratio for three methods

There is step problem (as fig.4 reveals) in the ALD and AFD methods, but PFD method smoothens the series, and inconsistence noticed in the series is reduced.

## IV. Discussions

From the tables and graphs above, it can be noticed that the discrepancies are obvious between the indicator and benchmarked series in ALD and AFD methods, the methods cannot reconcile the indicator with the benchmarked effectively. On the other hand, the PFD method gives a better result, in the sense that the benchmarked series derived by the method is more consistent that the first two methods. It can be noticed that the PFD method outperforms the other methods by its low values in standard error, standard deviation, range, and skewness (Table 3). Also the graphs (Figs. 3 and 4)show more stability and less inconsistency for benchmarked series produced by PFD methods.

## V. Conclusion

It can be concluded that the Denton Proportion First Difference (PFD) method reconciles better the movements in the indicator series. It is therefore recommended that users of economic data with benchmarking problem should employ the use of the Denton PFD. Further research can be carried out investigations into using various auto correlation coefficients to control the inconsistencies introduced by errors.

## References

[1] Ajao, I. O., Ayoola, F. J. and Iyaniwura, J. O. (2015): Temporal Disaggregation Methods in Flow Variables of Economic Data: Comparison Study. *International Journal of Statistics and Probability*, vol. 5, no. 1, 36-45

[2] Al-Osh, M. (1989): A dynamic linear model approach for disaggregating time series data, *Journal of Forecasting*, 8, 85–96.

[3] Bournay, J. and Laroque, G. (1979): Reflexionssur la m´ethoded'´elaboration des comptestrimestriels," Annales de l'INSEE, 36, 3–30.

[4] Cabrer, B. and Pavıa, J. (1999): Estimating quarterly time series in fulfilling annual and quarterly constraints. *International Advances in Economic Research*, 5, 339–349.

[5] Casals, J., Jerez, M. and Sotoca, S. (2005): Empirical modeling of time series sampled at different frequencies, paper presented at the workshop on Frontiers in benchmarking techniques and their application to official statistics, Luxembourg, 7-8 April 2005.

[6] Dagum, E., Cholette, P.and Chen, Z. (1998): A unified view of signal extraction, benchmarking, interpolation and extrapolation of time series *International Statistical Review*, 66, 245–269.

[7] Denton, F.T. (1971): Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization, *Journal of the American Statistical Association*, Vol. 66, No. 333, 365-377

[8] Di Fonzo, T. and Marini, M. (2003): Benchmarking systems of seasonally adjusted time series according to Denton's movement preservation principle, Dipartimento di ScienzeStatistiche, Universita di Padova, working paper 2003.9.

[9] Durbin, J. and Quenneville, B. (1997): Benchmarking of state-space models, *International Statistical Review*, 65, 23-48

[10] Friedman, M. (1962): The interpolation of time series by related series, *Journal of the American Statistical Association*, 57, 729–757.

[11] Ginsburgh, V. (1973): A further note on the derivation of quarterly figures consistent with annual data, *Applied Statistics*, 22, 368–374.

[12] Harvey, A. and Chung, C. (2000): Estimating the underlying change in unemployment in the UK, *Journal of the Royal Statistical Society*, 163, 303–328.

[13] Hodgess, E. and Wei, W. (2000): Temporal disaggregation of stationary bivariate time series, *Linear Algebra and its Applications*, 321, 175–196.

[14] Litterman, R. (1983): A random walk, Markov model for the distribution of time series, *Journal of Business and Economic Statistics*, 1, 169–173.

[15] Proietti, T. (1998): Distribution and interpolation revisited: a structural approach, *Statistica*, 58, 411–432.

[16] Quenneville, B. and Rancourt,E. (2005): Simple methods to restore the additivity of a system of time series, paper presented at the workshop on Frontiers in benchmarking techniques and their application to official statistics, Luxembourg, 7-8 April 2005.

[17] Rossi, N. (1982): A note on the estimation of disaggregate time series when the aggregate is known, *The Review of Economics and Statistics*, 64, 695–696.

[18] Salazar, E., Smith, R. and Weale, M. (2004): Interpolation using a Dynamic Regression Model: Specification and Monte Carlo Properties, NIESR Discussion Paper n. 126.

[19] Trabelsi, A. and Hillmer, S. (1990): Benchmarking time series with reliable benchmarks, *Applied Statistics*, 39, 367 – 379.

[20] Wei, W. and Stram, D. (1990): Disaggregation of time series models, *Journal of the Royal Statistical Society,* 52, 453 – 467.