An Improved Approach To Modelling and Evaluation of Economic data with Irregular Benchmarks

BY

ISAAC OLUWASEYI AJAO

B.Sc, M.Sc. (Statistics, Ilorin), M.Phil. (Statistics, Ibadan)

A Thesis in the Department of STATISTICS
Submitted to the Faculty of Science in partial fulfilment of
the requirement for the Award of the Degree of

DOCTOR OF PHILOSOPHY (PH.D STATISTICS)

of the

UNIVERSITY OF IBADAN

April 6, 2021

Certification

I certify that this work was carried out by **Mr Isaac Oluwaseyi AJAO** in the Department of Statistics, University of Ibadan, Ibadan, Nigeria

. . . .

Supervisor
Prof. J. O. Iyaniwura
B.Sc, M.Sc (Ibadan), Ph.D (Exeter)
Department of Statistics,
University of Ibadan, Nigeria

Dedication

This project is dedicated to the Lord Jesus Christ and my dearest spouse, Mrs. Josephine Oke Ajao.

. . . but the people that do know their God shall be strong, and do exploits.

Acknowledgments

My utmost sincere thanks go to the Lord Jesus Christ who preserved me through from the beginning to the end of my programme. My gratitude goes to my supervisor in person of Prof. J. O. Iyaniwura for his time and experience rendered to me throughout my research years, thank you for always being there sir.

Special thanks to my two co-supervisors Prof. O. I. Shittu and Dr F. J. Ayoola for their unceasing encouragements, perusing, additions and subtractions throughout the duration of the project. I must not but make mention of some staff members of the Department of Statistics, University of Ibadan, great scholars like Dr K. O. Obisesan, Prof. E. Olubusoye, Dr O. Alaba, Dr C. Udomboso, Dr O. S. Yaya, Dr Akanbi and others whose invaluable supports really assisted me in this research. I also remember the inestimable roles by international agencies such as the National Bureau of Statistics (NBS), the Central Bank of Nigeria (CBN), the International Monetary Fund (IMF) and the Statistics Canada for releasing their materials.

Thanks are also due to my parents, friends and colleagues. Ps. C. A. Ajao, Ps. H. O. Ajao, Dr O. S. Obafemi, Dr I. I. Raji and my H.O.D., Department of Mathematics and Statistics, The Federal Polytechnic, Ado-Ekiti, Mr. R. O. Iyiola for their timely support and concern.
My sincere and profound gratitude goes to my wonderful and loving wife, Mrs. Josephine Ajao, for her prayers and moral support all the time, which have brought this project into success. May the Lord God reward you abundantly. (Amen).

I pray that God will move you all forward in Jesus name (Amen).

Isaac Oluwaseyi Ajao

Abstract

Benchmarking deals with problem of combining a series of high-frequency data with a series of low-frequency data to form a single consistent time series. Various benchmarking methods in literature lack some observations (necessary for development of the eventual new series), at the beginning and end of the original series, which pose missing values challenge to the methods. Hence, there is need for an improved approach that will better capture these missing values. Therefore, the study was designed to develop an Autocorrelated Indicator Benchmarking Model (AIBM) that fills the value gaps without affecting the movement and pattern of the original series.

Two equations: $s_t = \sum_{h=1}^{H} r_{th} B_h + \theta_t + e_t$ and $a_m = \sum_{t=t_{1m}}^{t_{2m}} j_{mt}\theta_t + \varepsilon_m$ from the generalised least squares regression models were used to develop the new model, where s_t is the high-frequency series, r the regressors, h minimum value of the regressors, H the maximum value of the regressors, and B the bias values. The time effect is $\sum_{h=1}^{H} r_{th} B_h$. The benchmarked values θ_t, satisfied the annual constraints. The autocorrelated error, low-frequency series, the coverage fractions, and the non-autocorrelated error, are e_t, a_m, j_{mt}, and ε_m, respectively. The i^{th} and j^{th} values in the high frequency series are t_{1m} and t_{2m}, respectively. The model was validated with simulated data and real life data on the Nigeria's Gross Domestic Product (1975 to 2013) obtained from the Nigeria Bureau of Statistics annual report. The performance of the proposed model was evaluated based on autocorrelation coefficients (ρ) values compared with the existing models such as, Proportional Balanced Difference (PBD), Proportional Order One Difference (POOD), Additive Order Two Difference (AOTD), Proportional Order Two Difference (POTD), and Bias Adjusted (BADJ), using the Coefficient of Variation (CV) of the obtained growth rates. Minimum CV value will give a preferred model.

The developed AIBM was given as

$$\hat{\theta} = s - V_e J' V_d^{-1} Js + V_e J' V_d^{-1} JR \operatorname{var}(\hat{\beta}) R' J' V_d^{-1} Js - R \operatorname{var}(\hat{\beta}) R' J' V_d^{-1} Js - V_e J' \left(V_d^{-1} J V_e J' - I\right) V_\varepsilon^{-1} a$$
$$- \left(R \operatorname{var}(\hat{\beta}) R' J'\right)\left(V_d^{-1} J V_e J' - I\right) V_\varepsilon^{-1} a + \left(V_e J' V_d^{-1} JR \operatorname{var}(\hat{\beta}) R' J'\right)\left(V_d^{-1} J V_e J' - I\right) V_\varepsilon^{-1} a$$

, where $\hat{\theta}$ is the matrix of the benchmarked estimates. The covariance matrices of the survey, low frequency, and high frequency errors are V_e, V_d, and V_ε, respectively. Also the estimates of bias parameters and regressors are $\hat{\beta}$ and R, respectively. For simulated data, the CV values of growth rates from PLD, PFD, ASD, PSD, BADJ, and AIBM at $\rho = 0.729$ were -29.620, -14.033, -24.353, -13.160, -19.591, -29.486; at $\rho = 0.900$ were -29.620, -14.033, -24.353, -13.160, -19.632, -29.606; at $\rho = 0.990$ were -4.402, -4.987, -4.371, -4.954, -7.137, -4.402; and at $\rho = 0.999$ were -4.402, -4.987, -4.371, -4.994, -7.309, -4.402, respectively. For real life data, the CV values at $\rho = 0.729$ were 3.195, 3.196, 3.198, 3.200, 1.582, 1.318; at $\rho = 0.900$ were 3.195, 3.196, 3.198, 3.200, 1.582, 1.318; at $\rho = 0.990$ were 3.195, 3.196, 3.198, 3.200, 1.582, 1.121; and at $\rho = 0.999$ were 3.195, 3.196, 3.198, 3.200, 1.582, 1.105, respectively. The AIBM has the minimum CV in the growth rates, indicating its strength over the existing models.

The autocorrelated indicator benchmarking model captured missing values at the beginning and end of the original series, while preserving the properties of the series. The model is therefore recommended for handling irregular data.

Keywords: Benchmarked estimates, Autocorrelated indicators, regressors, Coefficient of variation
Word count: 497

Table of Contents

List of Tables

List of figures

Glossary of main terms

1. Benchmark: the series of less frequent data, which is considered more reliable.

2. Benchmarked series: the resultant series from benchmarking.

3. Benchmarking: the procedure for combining a series of high-frequency data (the indicator series) with a series of less frequent data (the benchmarks) for a certain variable into a consistent time series.

4. BI ratio: benchmark to indicator ratio.

5. Binding benchmark: A benchmark with dispersion value equal to zero; that is, the benchmarked series will fully match the benchmark.

6. Coefficient of variation: the standard deviation of the observation divided by the value of the observation.

7. Flow variable: variable measured by summing over a reference period, e.g. sales per month.

8. Index variable: variable measured as an average over a reference period, e.g. the index of industrial production.

9. Indicator series: the series to be benchmarked, used interchangeably with "original series".

10. Interpolation similar to benchmarking, but the indicator series is not directly related to the series of benchmarks.

11. Non-binding benchmark: A benchmark with a positive dispersion value;

that is, the benchmarked series will not exactly match the benchmark.

12. Original series: the series to be benchmarked.
13. Proportional Denton Method: benchmarking method which preserves the growth rate of the original series in an optimal manner.
14. Reference period: set of consecutive months covered by a data point, e.g. April 1, 2000 to June 30, 2000.
15. Stock variable: variable measured as of a particular date, e.g. inventories.
16. PRM: Prorating Method
17. PBD: Proportional Balanced Difference
18. ABD: Additive Balanced Difference
19. POOD: Proportional Order One Difference
20. AOTD: Additive Order Two Difference
21. POTD: Proportional Order Two Difference
22. BADJ: Bias Adjusted
23. AIBM: Autocorrelated Indicator Benchmarking Model

CHAPTER ONE

INTRODUCTION

1.1 Background to study

The term Benchmarking can be explained as the situation where there are two locations of information for a similar focused variable, with various frequencies, and is making sure that adjustment of irregularities between the various evaluations is done, for example quarterly and yearly data from various roots. It is expected that three-month accounts from various countries are predictable with the yearly accounts, therefore a reasonable perspective on financial advancements is introduced contrasts in developmental rates among sub-annual and yearly GDP would befuddle and bother clients. Benchmarking is commonly done reflectively as yearly benchmark information are accessible at some point after sub-annual information.

It is a way of forcing a high frequency data such as weekly, monthly, or quarterly referred to as the indicator series to obey the series of the lower frequency. The purpose of this is to come up with an hybrid series that will be better and reliable. The low frequency series (annual) are the benchmarks while the high frequency are the indicator series which may be monthly or quarterly. The problem of benchmarking arises when the benchmark to indicator ratio is greater than 1. It is generally assumed that the ratio be equal to 1, because the total sum of the indicator series is expected to be equal to the benchmarks.

Obtaining high frequency data is not a challenge for users of economic data, but it's reliability in usage. The low frequency data obtained by organisations are sponsored by the state or federal government of many nations, so in most cases they are reliable. However, problem occurs when the user or researcher discovers that the high frequency data is not consistent with the annual from these organisations. Effort therefore is needed to be put in place in making sure that evaluation of such data sets is performed in order to reconcile the series gotten from two sources. This can be week to week, month to month, quarter to quarter, benchmarked to new series. In Nigeria, the data collected by the Central Bank of Nigeria and the National Bureau of Statistics is highly reliable because it has the financial strength to make use of professionals in all walks of life to obtain their data. Therefore the same data collected from organisation may have benchmarking problem because the total from the first source may not be consistent with the later.

In our effort to benchmark data sets from two sources, the dimension of such data sets have to change. The number of observations in the low frequency and high frequency series have to reconcile to produce the dimension of the indicator series which will be used as the representative of the data from two sources, this is the benchmarked series and can be used for future research because of its reliability. Benchmarking is therefore an improvement on both the low frequency and high frequency data.

Temporal disaggregation and calendarisation are connected systems which depend on comparative methodological standards and rules as those of benchmarking. Temporal distribution is breaking benchmark series into more high frequency series. Calendarisation on the other hand is aggregating the data into days of week, months and quarters in the year

Benchmarking provides solution to the challenge of bringing together a series high frequency (for example, month to month) and a series of low frequency (for example, quarter to quarter) in order to have a time series data that will be suitable for other uses. When a conspicuous irregularity between two arrangements is seen, the low recurrence information is generally expected to give a dependable information. There are two central approaches to manage benchmarking of time arrangement: an essentially mathematical system and a verifiable showing approach. The mathematical procedure encompasses the gathering of least-square minimisation strategies (Denton, 1971). This benchmarking strategy relies upon an improvement preservation decide that is for the most part used by government real associations and public banks all over. Factual demonstrating approaches incorporate autoregressive integrated moving average (ARIMA) techniques suggested by Hillmer and Trabelsi (1987), regression models (Cholette and Chhab, 1991), and models of state space (Durbin and Quenneville, 1997).

Benchmarking is a significant issue looked by the measurable offices. For an objective financial variable, two wellsprings of information, for instance, an annual authoritative information and a quarterly rehashed overview information, might be accessible. At the point when errors between the series of high recurrence (for example quarterly) information and a series of low recurrence information emerge, the last is normally accepted to give progressively reliable data. The issue of modifying the month to month or quarterly time series to make them predictable with the sub-annual or annual aggregates is known as benchmarking.

Changing the estimations of a series of values of a variable at successive times that is seen over shifting time into qualities that spread schedule time is also known as calendarisation, for example, day, week, month, quarter and year. For instance, a specific number of travelers take open transportation consistently, and these numbers are summed to give an aggregate for a given revealing period. Calendarisation includes temporal conveyance of the detailed qualities into, state, every day occurrences and accumulation of the subsequent day by day insertions into the ideal recurrence, month to month, quarterly or yearly.

The temporal circulation or disaggregation process, more often than not, includes an everyday indicator which is utilised as an intermediary to speak to the day by day dimension of movement that can differ attributable to regularity, exchanging day or other date-book impacts, for example, open occasions. Without such data, the everyday dimension of movement is thought to be consistent. Given the day by day indicator, the subsequent stage is to benchmark it to the revealed qualities. The detailed qualities give solid data on the general dimension and long haul development. Benchmarking smoothens the everyday development of the indicator series while at the same time accomplishing the dimension of the announced qualities. Benchmarking gives smooth day by day esteems that signify the revealed qualities; thus, the announced qualities are the aggregates of every day estimates and this compares to their temporal circulation. Benchmarking is an obliged streamlining issue that can be made to fathom in numbers or utilising proper models of state space.

1.1.1 Defining Benchmarking using the idea of Interpolation

Interpolation is the process of introducing new set of observations into a series following a specified pattern and distribution. It is a method or operation of finding from a given terms of a series, as of numbers of observations, other intermediate terms in conformity with the laws of the series. Here the intervals into which the observations introduced are called the benchmarks, and the series obtained for interpolation are the benchmarked series.

1.1.2 The Benchmarking Problem

The benchmarked series to be evaluated is denoted as $d_{r,u}$ where $r = 1, \ldots, y$ signifies the annual period, and u indicates the quarters. It can be obtained from the set of data that follows, the high frequency series: $c_{r,u}$; and the yearly benchmarks, are:

$$D_r = \sum_{u=1}^{4} d_{r,u} \text{ for continuous values}$$

or

$$D_r = \frac{1}{4}\sum_{u=1}^{4} d_{r,u} \text{ for gauged data}$$

Benchmarking is aimed at reconciling data sets generated by two different sources. Assessing $d_{r,u}$ over the range for which $c_{r,u}$ is accessible is important. Evaluating $d_{r,u}$ past the period for which $c_{r,u}$ accessible is increasingly stable. For this situation anticipating and benchmarking strategies should be joined. Determining $c_{r,u}$ then benchmarking the all-inclusive series onto the benchmarks D_r is a reasonable arrangement. Truth be told in view of the overall practicality of 12-month and 3-month information, amid the present year y, the qualities $c_{y,1}$, $c_{y,2}$, $c_{y,3}$ will commonly be accessible before the yearly benchmark D_y , or C_y is known. So, the last couple of information focuses in the fractional year toward the finish of the series $d_{y,1}$, $d_{y,2}$, $d_{y,3}$ are evaluated without an yearly sum - a significant part of any technique for benchmarking is the manner by which it manages this issue.

1.2 Aim and Objectives

In this research, the aim is to take further the work of Denton (1971), and Dagum and Cholette (2006) by developing an improved approach to modelling and evaluation of economic data with irregular benchamrks, that is, the Autocorrelated Indicator Benchmarking Model (AIBM) to solve benchmarking problem. The definite objectives are to:

i. Improve on the order one and order two proportional and additive benchmarking models using restrained minimised quadratic polynomial

ii. Introduce various autocorrelation coefficients into the new models in order achieve a robust hybrid series

iii. Regulate the sub-yearly data dimension to obtain a consistent one

iv. Write a MATLAB algorithm to provide solution to challenge of benchmarking

1.3 Statement of the problem

Across the globe, one of the problems faced by government agencies that collect and publish economic data is that of adjusting low frequency data such as annual series to a high frequency data such as quarterly or monthly series. This general problem in economic data is known as benchmarking. The purpose is to revise a sub-annual series so that it is forced to sum for a certain time period to a given figure (a benchmark) or forced at a certain time to be equal to a given figure (a benchmark).

Many occasions, researchers get high frequency data from a particular source (for example, using questionnaire); and the relating yearly totals from quite a different source, (for example, a registration). The yearly totals of the high frequency data are commonly not having same pattern with the yearly sums. Such high frequency series expect to sum to yearly totals, when this fails we talk about benchmarking.

Numerous issues regularly experienced in the readiness of economic data is that of altering month to month or quarterly series got from one source to make them agree with annual sums obtained from other data source. It is assumed that our interest is on the high frequency series which has q values every year, it should be noted that q is a whole number. It is also assumed that the low frequency data has u years and contains $y = uq$ observations. The first series are in matrix form is denoted by $\mathbf{c} = [\mathbf{c_1} \ \mathbf{c_2} \ . \ . \ . \ \mathbf{c_y}]'$. It is expected that another series $\mathbf{m}$ yearly values which come from another source is denoted by $\mathbf{d} = [\mathbf{d_1} \ \mathbf{d_2} \ . \ . \ . \ \mathbf{d_u}]'$ The challenge available for research is to adjust the set of values $\mathbf{c}$ in order to obtain another set of values $\delta = [\delta_\mathbf{1} \ \delta_\mathbf{2} \ . \ . \ . \ \delta_\mathbf{y}]'$ by minimisation with order-two benchmarking models.

1.4 Justification for the study

Various methods of benchmarking have been considered and utilised by users of economic data and other researchers, yet objective function using 1st and 2nd order differences models, and introduction of various levels of autocorrelation coefficients as ways of reducing inconsistencies in the series have not been done.

1.5 Source of Data

The economic datasets used for this research were obtained from the National Bureau of Statistics (NBS) bulletin, covering a period of 1975 to 2013

1.6 Scope of the Study

This study is aimed at providing solution for benchmarking challenge by making use of restrained order-two minimisation with numerical approach and autoregressive models of many orders by making use of the economic data sets obtained from the National Bureau of Statistics (NBS).

CHAPTER TWO

REVIEW OF LITERATURE

2.1 Introduction

In this chapter, efforts are made to review some previous works of various authors. It exposes different methods used by various researchers on adjusting high frequency data to annual benchmarks. A summary on various literature in regards to the ways benchmarking problem is resolved is contained in this chapter. This part also contains literature review of both theoretical and empirical.

2.2 Theoretical Literature Review

It is stressed by Aadland (2000) that the circulation and act of introducing observations in time series have been liable to different information changes. Tests using Monte Carlo method are performed, which propose that inability to represent these information changes may prompt a lot of blunders in estimation.

The vast majority of the information acquired by factual offices must be balanced, redressed or some way or another handled by analysts so as to land at helpful, steady and publishable qualities. For instance, the administration organizations that gather and distribute high frequency countries' accounts time series must deliver sub-yearly information that simultaneously consent to the significant annual figures and fulfill bookkeeping requirements (Eurostat, 1999). This sort of issue emerges additionally when an arrangement of equally time-spaced data is regularly balanced utilising an approach with the aim of keeping the pattern of the real data intact after some adjustment has been made (Di Fonzo and Marini, 2003).

To stay away from ventures between back to back years, benchmarking systems de- pendent on some development protection standards are suggested. Cholette (1984) suggested a generally utilised benchmarking technique, with the change for the beginning condition. This method selects a target function (or first-order Proportional Differences) of the benchmarked series, and position it as conceivable to the high frequency data. From now on, this technique will be alluded to as the fundamental Denton first-order difference technique. Under the given requirements, the Denton first-order difference system is gone for safeguarding at the best the developments in the indicator series.

This challenge is laid out in the manual of the International Monetary Fund (IMF)'s high frequency countries' Accounts: Conceptions, Sources of data, and Data gathering (Bloem et al., 2001). To maintain a strategic distance from a conceivable predisposition and brighten the nature of the appraisals in calculation of functions outside the range of values, the Countries Account guideline suggests an upgraded variant of the Denton's First-order Proportional Difference arrangement.

For the month to month studies, the inspiration for calendarisation comes to a limited extent from the way that the information that are accounted for may not speak to a similar time span as the reference time frame. The time range of a unit revealing under a run of the mill four weeks– four weeks– five weeks example will shift from six per cent to sixteen per cent when differentiated and the time length of a veritable month to month design (thirty or thirty-one days excepting February).

In order to make the data good and presentable for usage, it is desirable to turn the calendar intervals to minimums such as days, weeks, months, and years. Considering the 12-months survey, the time interval is likely to be closer to the comparison time chosen.

Using the particular instance from four weeks–four weeks–five weeks, Cholette and Chhab (1991) have proposed a technique to change over information alluding to a fluctuating number of days into calendar values. Their technique expect that a day by day indicator series is accessible. In their paper, every day high frequency data is built from seven day by day loads speaking to the overall significance of the seven days of the week. Next, the everyday indicator series is benchmarked to the announced stream review with a suitable variation of the Denton (1971) benchmarking method modified by Cholette (1984).

Cholette and Chhab (1991) talked about why their technique gives an improvement over different strategies, for example, professional rating the detailed qualities as indicated by either the quantity of days or the total of the everyday loads in the announced periods. For instance, with customizing as indicated by the quantity of days, the everyday esteems are assessed by the normal of the benchmarks. Such every day interpolations are consistent inside each announcing period and in this way unexpectedly change higher than ever between two benchmarks.

The assumption of steady day by day esteems is clearly abused when the dimension of the benchmarks differs starting with one benchmark then onto the next; along these lines, theoretically, consistent day by day interpolations are at most impolite daily values that can be adjusted positively. Researchers in benchmarking can refer to the comments in Cholette and Chhab (1991) for further studies. To acquire every day evaluates through standard benchmarking techniques, for example, the changed, Denton relative first-contrast strategy, it is important to tackle a direct arrangement of conditions Quenneville and Fortier (2012). The techniques requires reversal of substantial networks which may raise challenges of productivity.

Asogu (1997) examined some non-parametric distribution free and robust algorithms for interpolating annual series into quarterly series and applied the procedure to disaggregate money supply, imports and exports data directly.

Bozik and Otto (1988) gave two reports from the results of their research. The first talks about comparison of three alternatives for benchmarking of monthly series to yearly series. The three objective function are minimisation of changes in sum of squares with respect to adjusted data and relative monthly changes. It was found out that there is no practical difference between seasonally adjusted and unadjusted methods. It was also observed that unadjusted method obtained lowers the median absolute change in the levels. In comparing the methods, the relative method is identical to unadjusted method in the second decimal places.

One of the many researchers that worked on interpolation of economic data methods was Ajayi (1978). Using parametric linear regression approach in which he attempted to estimate quarterly GDP. Since no primary quarterly data are available for current GDP, interpolated quarterly series are obtained by regressing annual export income on time.

The bookkeeping requirements associating the series are not satisfied if an arrangement of time series is occasionally changed. In order to deal with the problem, an accounting constraint of classical univariate benchmarking using Denton (1971) method was employed. This method is based on principle of movement preservation. Since autocorrelation is present between the variables, it is therefore necessary to deal with the whole set of temporal aggregation relationship. In addition, calculations involving matrices will be needed when implementing this method in real life applications.

Many economic data are only available on sum totals of high frequency series. Whenever the researcher needs disaggregated data, he now faces the challenge of benchmarking. A data based method by Guerrero (1990) was formulated in order to handle the problem, this gives estimator of broken series. The method needs a first approximate value of the data in order to effectuate the rule set by the sum totals. Some data were employed in implementing the developed method to test its adequacy, and comparison with other existing methods were done.

Ajani (1978) attempted to estimate quarterly GDP series, was aborted ostensibly because of the non-availability of related series needed as inputs to construct the series.

A template developed by Moauro and Savio (2004) was on multivariable data for fragmented series, the series was given in a particular number of observations, but later into higher frequency series. The developed method uses the independent time series model calculated by the Kalman filter. The method is flexible and allows all kinds fragmented series of raw and seasonally adjusted time series.

The use of methods of state space to give solution to the problem of disaggregated data is encouraged. One of such is by dynamic regression models which is one of the most popular methods for economic data. Some of them are Chow-Lin, Fernandez and Litterman. The work contributes to existing literature in these ways: (i) it focuses on the concise beginning of the various models, revealing that the outcome is important to the features of the maximum likelihood estimates and formulating all around models of autoregressive distributed lag. (ii) it exhibits the function of diagnostics histories in summarizing the quality of the fragmented estimates. (iii) it gives an adequate overview of the Litterman model, stressing the challenges mostly encountered in practice when the model is being estimated.

Proietti and Moauro (2006) makes use of the United States economy by using time series with various frequencies such as monthly and quarterly data. The dynamic factor model was considered by the authors. A problem of temporal sum totals with non-linear constraint emerged as a result of quarterly time series that were included. The contribution of the paper was the provision of monthly estimates of quarterly indicators.

Proietti and Moauro (2008) carried out an approximate calculation of some countries' accounts which generates quite a lot of challenges that are of vital consequences for current economic assessment. In many nations in Europe, national accounts are obtained using the method of disaggregation of the original yearly totals by making use of the indicator series. The most important advantage of this method is to allow temporal fragmentation and adjustment be done at the same time. The only issue is that identifying and separating effects of seasons and calendar from aggregate data often lead to controversy.

Numerical algorithm developed by Causy (1981) which was later reviewed by Trager (1982) became widely known as the Causy-Trager method. Monthly changes that have been subjected to certain benchmarks constraints are minimised iteractively using steepest descent. The work was not published but their notes are written as appendixes in the reports by Bozik and Otto (1988).

Rodŕiguez, Rodŕiguez, and Dávila (2003) proposed a method that gives quarterly series from annual series. However, this method and some others have not been carefully tested to confirm which is recommended for use with Monte Carlo simulation, the authors compared the methods which use only data from annual series. Results from the methods, when analysed as functions of attributes, for example, number of years of the real annual series suggest which of the methods to employ.

Santos and Cardosol (2001) made an application of the Chow and Lin (1971) method for data fragmentation making use of dynamic models. Flexibility is added consider- ably to the basic approach especially when the series are stationary or not stationary. Somermeyer, Jansen and Louter (1976) method assume that the data used should be weighted moving average of yearly observations. The weights have to be obtained through quadratic programming. Quarterly series functions for the U.S. and the Netherlands were estimated using this multivariate method on the basis of annual income. Considering the parameter approximate values and sizes and randomness of the values from the calculated and the observed differences, the method of Feibes, Boot and Lisman is not as reliable as the multivariate method.

Wei and Stram (1990) worked on model that disaggregate aggregate series. The model is also used to detect autocovariance structure. Assuming the aggregates of time series be nonoverlapping totals is fragmented consecutive values, if an aggregate ARIMA (p, d, r) model, and if there is no periodicity of a particular order that is not opened. For a disaggregate model whose autocovariances of a given aggregate model to exist, the order must be odd and the real roots of the autoregressive model must also be positive.

Abad and Quilis (2005) presented a set of computer programs purposely prepared to execute disaggregation of economic time series with various techniques: singles de- pendent variable without indicator series, many dependent variables with indicator series and constraints. The algorithm has two important parts: a MATLAB code and an interface in microsoft Excel. This is used in compiling the national accounts in Spain and to execute special and full analysis which certifies the data and model reliability.

The European Union (EU) method of benchmarking aggregates of national accounts is basically is temporal fragmentation of series. What Eurostat normally does to compile quarterly accounts using the Chow and Lin benchmarking method supported by Denton multivariate technique. The paper reveals the summary of the technique used in bringing together the aggregates of the European national accounts with the intention of benchmarking issues such as quality and contents of the indicator series, influence various of seasonal adjustment. In addition, the method using benchmarking technique measures the possible effects of knots brought about by the changes in method (Barchellan, 2005)

Bruno et al (2005) developed a technique to combine two approaches, that is, temporal disaggregation and Bridge models. The model is used to analyse the GDP movements in quarters, but related series was used to confirm the reliability of the method. The output of the data (GDP) from the countries reveals that the information gathered from the two indicator series in the bridge models could be helpful in selecting the necessary related needed in the procedure of temporal disaggregation.

Ciammola et al (2005) compares various techniques of disaggregation of annual series in the presence of an indicator series, using an experiment of Monte Carlo. Estimating the autoregressive model brought about by the Chow and Lin solution is the first objective of the model. The following are the three methods of estimation that were considered in their paper: The Chow and Lin method (1971) making use of the autocorrelation and autoregressive parameter, the log-likelihood maximization of the residual sum of squares, developed by

Barbone, Bodo and Visco, (1981).

Di Fonzo and Marini (2005) did some adjustments and correlations on certain economic data. This is necessary for the series to be made ready for research purposes. This government agencies across the nations gather sub annual national accounts time series that must agrees with the annual totals and validate the accounting constraints (Eurostat, 1999).

Guerrero and Neto (1990) provided a procedure for eliminating discrepancies using various methods reconciliation of data sets. Discrepancies normally comes in whenever temporally and contemporaneously sum totals series are available for research purposes. One dependent variable benchmarking method by Denton (1971) using movement preservation technique and a data-oriented benchmarking method are explored in the paper. It's procedure exploits the autoregressive attributes of the initial series for adjustment. Researchers normally make use of both simulated and real life data to evaluate their performance.

Generally, the procedure for handling benchmarking is basically principle of movement preservation, and it has become known by various government statistical bodies globally. Similar models of state space by Durbin and Quenneville (1997) were also developed. It is a well known fact that binding benchmarking is the most widely used agencies to provide way out to hurdles of differences found in real life data. The method of unbinding is widely used for data obtained through survey, however although it's source must be known and reliable. It is established that since the series are adjusted every month or quarter, filtering process may bring some problems. Therefore, as a result of adjustment, the accounting constraints may be violated.

When variables that occur in regular intervals have discrete observations, they are referred to as unobserved integrals of flow functions. There exists more than two expressions involving one or two variables from flows. The initial step is to reduce or completely delete the continous movement in the measures of central tendency and dispersion. The second step is to ascertain the stationarity of the variable even after series of activities of adjustments have been carried out on it. The data points need transformation in order to become a flow variable. One important fact is that individually separate and distinct data contains more dependable information about low frequency data. On the other hand, related series and combined analysis introduces differences through high frequency data. A vital advantage of this system is that integration removes the differences embedded in the high frequency data. As a result of this, obtaining measures of association in such series having similar intervals of frequency will be impossible. (Gudmundason, 2005)

One major problem experienced in the handling economic time series data is that many of the data sets is only available in the form of annual sum totals. For instance, it is easier seeing monthly or quarterly figures than having daily or weekly data. A lot of research had been done had been done on univariate time series, but little in the case of multivariate (Rossie, (1982), Di Fonzo, (1990)). This paper works on problem of temporal disaggregation considering the case of the multivariate. This implies the challenge of m¿1 high frequency series by making use of the relevant low frequency data. In both cases, the aggregation constraints must be fulfilled. A univariate polynomial technique was developed by Ahmon (1988) using an econometric software. It is to convert low to frequency series to high frequency by interpolation.

Using a non-linear method also referred to as Causey-Trager technique, benchmarking of monthly and quarterly series are done every 5 years at the U.S. Census Bureau. The X-12 ARIMA program for seasonal adjustment together with the reviewed Den- ton procedure are employed to seasonally adjust the series to annual totals, though it results to some distortions in the benchmarked series obtained from the technique.

Apart from using the reviewed Denton method, the Census Bureau in Canada pro- posed various methods of benchmarking as better alternatives for benchmarking while using the X-12 ARIMA the procedure is carried out using a sample time series to test the features of the benchmarks from X-12 ARIMA method, the procedure from Census Bureau in Canada, and the Causy-Trager method.

The goals of the study are to investigate some of the features of different procedures of benchmarking at the Census Bureau and to verify settings for procedure of regression analysis in the agency. Results show that some discrepancies were noticed in monthly series, and small consistent outcome in smooth benchmark factors using both the method of regression when $\lambda = 1$ and Causy-Trager (Hood, 2005).

Keogh and Jennings (2005) look into how cells in arrays can be rearranged and calibrated to give marginal sums using an accepted input-output table having avail- able information on the marginal totals. Other researchers developed a method for detecting associations between sub annual and annual data using the method of a dynamic regression. The model may be nonlinear but can be used to carry out interpolation between one observation and the other on a variable which occur only on annual series. Discussions on its implementation are made and monthly GDP was were used as an example for European area.

In their paper, Moauro and Proietti (2005) explains the challenges encountered when obtaining estimates of national accounts for quarters, adjustment of seasons, and effects of calendar components gotten by breaking annual totals to sub-annuals, using related monthly indicators. The paper suggests and applies a method that depends on obtaining bivariate basic time series at a low frequency. It is discovered that monthly values give better estimation of components of the normal calendar. One important feature of this approach is that data adjustment and temporal frag- mentation can be done at the same time. This proposed method joins with the suggestions from the Eurostat on national accounts seasonal adjustment.

Kalman filters provide a reliable way out to many challenges of time series. In Belgium, the Research and Development section of its national bank created a platform that gives a comprehensive framework for a single dependent variable Kalman filters. This procedure is used in solving fragmentation of normal series involving single dependent variable time series using various methods, without necessarily depending on related series. The computer programs written by Palate (2005) goes the same way with the work of Proietti (2004)

According to Koopman (2001), the computer programs written should be on the foundation of Kalman filter. In the paper, comprehensive formulae for filtering, calculation of probability and smoothening are provided. Effects of predicting variables are also given vital attention which can be introduced to the state vector or obtained through the Kalman filter augmentation.

Quilis (2005) carried out an overview of many benchmarking methods which were used to gather the national accounts of Spain, the work contains some stages which are: one-response variable and multi-response variable disaggregation are employed to generate data. Another one is using ARIMA model as used in TRAMO-SEATS computer programs. The last stage is adjusting seasonal data and transforming it in order to have consistent sub-annual and annual national accounts. This stage is implemented using multi-response procedure for temporal disaggregation. The implementation of the method is achieved through the use of quarterly supply use tables, generally referred to as the QSU model.

Bayes method was proposed by Rojo and Sanz (2005), it is to obtain sub annual series which will be consistent with annual series with the helps of indicators. The new structure requires setting of continuous variables. The result is usually higher probability of prior which smoothens the series better. Provision was made for a prior error covariance matrix, this is necessary in order to account for other data, which are said to be independent apriori. The posterior covariance matrix was also made available by the method. Obviously, the distribution of the prior and the distribution of the likelihood function are from the family of normal -gamma. The paper was concluded with evaluation of the proposed method. For the technique to compare well with the classic method, many errors schemes are simulated, and as a result, many indicators are obtained. The outcome series are therefore compared with the classic counterpart.

Smith and Hidiroglou (2005) paper proposes macro level technique in carrying out benchmarking of high frequency to low frequency series. In efforts to reduce distortions between high frequency data totals and low frequency data, another procedure is the macro level technique. The low frequency series is used as auxiliary data in estimating the survey values, hence the adjustment of the high frequency is done in order to have consistency in the estimates of the annual and the sub annual.

Trasbelsi and Hedhili (2005) made presentation of various procedures that deal with the challenge of disaggregation of economic data with the availability of economic data with the availability of related series and without the presence of related series. It can be inferred from the two methods that the first method share the discrepancies between the annual sums and the sub annual data in such a way as to maintain consistency. However, the second method which uses only aggregate values provides nothing to add to the economic interpretation.

2.3 Empirical Literature Review

In this section, the review of empirical literature is made. This is necessary in order to drive home the importance of the various benchmarking models available in the literature and their applications.

2.3.1 Techniques for solving benchmarking problems

Basically there are two major ways:

i. Numerical computations technique,

ii. Modeling technique.

There is difference between the numerical computation technique and the modelling technique. The difference is that the numerical does not assign any model for data that occur at regular interval of time. The numerical technique uses the well-known minimisation of the least squares suggested by Helfand, Monsour, and Trager (1977), and the technique formulated by Ginsburgh (1973). The modelling technique uses Autoregressive Integrated Moving Average given by Hillmer and Trabelsi (1987), State Space models suggested by Durbin and Quenneville (1997), and many models of regression given by Laniel and Fyfe (1990), and Cholette and Dagum (1994)). Chow and Lin (1971) also came up with an approach of the least-squares for inserting values of a function between the values already known, appropriation, and extrapolation of data that occur in regular time interval or binding time series.

2.4 Plan of the Problem and a methodology dependent on Quadratic minimisation to its solution

Assume that we want to get rid of the challenge about the within-year time data of which there are q every year, q being a whole number. Let the time data have u years also, have $y = uq$ values. The first qualities are spoken to in section vector structure by $c = [c_1 \ c_2 \ . \ . \ . \ c_y]'$ Additionally we have, from an alternate source, a lot of u annual aggregates spoken to by $d = [d_1 \ d_2 \ . \ . \ . \ d_u]'$. The issue is to modify the first vector h to acquire another vector $\delta = [\delta_1 \ \delta_2 \ . \ . \ . \ \delta_y]'$ using a technique which (i)reduces inconsistencies in the observation, and (ii) validates the rule that stipulates it that the addition of q observations of the new series must be equal to the total sum of the 12-month series under consideration, $p(\delta, c)$, the goal is to select δ in order to

minimize $p(\delta, c)$

subject to:

$$\sum_{(V-1)q+1}^{Vq} \delta = d_V \qquad \text{where V = 1, 2, . . .,} u \tag{2.1}$$

Making use of the penalty function denoted by $(\delta - c)' B(\delta - c)$, a quadratic structure in the contrasts between the first and balanced values of time series, it should be noted that B is a symmetric y X y non-singular matrix. A Lagrangian expression then results:

$$w = (\delta - c)'B(\delta - c) - 2\varphi'(\alpha - G'\delta) \tag{2.2}$$

Where

$$\varphi = [\varphi_1 \; \varphi_2 \ldots \varphi_n] \tag{2.3}$$

and G is a y X u matrix, y is the number of observations in the sub-annual series while u is the number of observations in the annual (benchmarks) series.

$$G = \begin{bmatrix} g & 0 & . & . & . & 0 \\ 0 & g & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & g \end{bmatrix}$$

g is a q dimensional - column vector in which every component is 1 and 0 being a q column vector and G is y X u. The objective function is achieved by taking an incomplete or subordinates of w as for the components of e and φ, equating them to zero, and tackling. For ease of calculation, we compose $t = (\alpha - G'\delta)$ for the vector of disparities between the two arrangements of yearly sums and express the arrangement in the structure:

$$\begin{bmatrix} e \\ \phi \end{bmatrix} = \begin{bmatrix} B & G \\ G' & 0 \end{bmatrix} . \begin{bmatrix} B & 0 \\ G' & I \end{bmatrix} . \begin{bmatrix} e \\ t \end{bmatrix}$$

Where I is the uXu character lattice and 0 is the uXu invalid framework. (It is accepted, obviously, that the second-request conditions important for the answer for be a base are fulfilled.) Using an outstanding outcome for determining the backwards of a divided framework, the answer for δ is then observed to be $\delta = c + Ft$,

where $F = B^{-1}G\,(G'B^{-1}G\,)^{-1}$. Consequently the balanced qualities are equivalent to the first qualities in addition to straight blends of the inconsistencies between the two arrangements of annual sums.

2.4.1 Using Lagranges Multipliers for the principle of Constrained

Optimisation

In a design that is optimal in nature, values for a particular set of y design variables such as, $(\delta_1, \delta_2, \ldots, \delta_y)$, should be made available in such a way that it will minimise a penalty function of the desired designed variables, by making sure that a collection of similar u unequal requirements are fulfilled. Constrained optimisation issues are commonly communicated as:

$$min_{\delta_1,\delta_2,\ldots,\delta_y}\ K = f(\delta_1, \delta_2, \ldots, \delta_n) \tag{2.4}$$

Such that

$$p_1(\delta_1, \delta_2, \ldots, \delta_y) \leq 0$$

$$p_2(\delta_1, \delta_2, \ldots, \delta_y) \leq 0$$

$$\ldots\ldots\ldots\ldots\ldots$$

$$p_u(\delta_1, \delta_2, \ldots, \delta_y) \leq 0$$

In a case where the penalty function is quadratic naturally with respect to the design variables, issue of optimisation will have a solution that is unique to itself. It is not every problem of optimisation that is not cumbersome to handle, many method of optimisation need advanced approaches. One of them is Lagranges multipliers, it is applied to solve to benchmarking introduced by Denton (1971)

The methods of Lagrange multipliers entail the adjustment of the penalty function by summing up the observations. The penalty function $K = f\ (\delta)$ is enlarged by the equations of the constraints using a collection of multiplicative Lagrange multipliers that is non-negative in nature, $\varphi_g \geq 0$. The enlarged objective function, $K_B(\delta)$, depends on the values of y design variables and u Lagranges multipliers

$$K_B(\delta_1, \delta_2, \ldots, \delta_y, \varphi_1, \varphi_2 \ldots, \varphi_u) = f(\delta_1, \delta_2, \ldots, \delta_{y,}) + \sum \varphi_j p_g\ (\delta_1, \delta_2, \ldots, \delta_{y,})$$

For the problem above, $y = 1$ and $u = 1$, so

$$K_B(\delta, \varphi) = \frac{1}{2} q\delta^2 + \varphi(1 - \delta) \tag{2.5}$$

The Lagrange multiplier, φ, effectively modifies (enlarging) the objective function from a level of quadratic $\frac{1}{2}q\delta^2$ to a different level of quadratic $\frac{1}{2}q\delta^2 - \varphi\delta + \varphi l$ with the goal that the base of the adjusted quadratic fulfills the requirement ($\delta \geq l$).

2.4.2 Solutions by Denton

Let B be an identity matrix, which implies that we are limiting the square sums of the contrasts that exists between the original and the amended observations. For this case, $F = 1/qG$, an objective function is suggested that is limited by disseminating the disparity for every year that will give yearly totals among the q time frames inside the year. As noted previously, this outcomes, by and large, in a deceptive advance or irregularity between the last time of one year and the main time of the following. Unmistakably, $B = I$ is probably going to be an awful detail.

A better alternative is utilising an objective function that depends on the differences that exist between the first and the adjusted series:

$$f(\delta, c) = \sum_{v=1}^{y}(\Delta\delta_y - \Delta c_v)^2 \quad = \sum_{v=1}^{y}[\Delta(\delta_t - c_t)]^2 \tag{2.6}$$

2.5 Method proposed by Denton with it's variants

Denton (1971) came up with a benchmarking technique that depends on the guide- lines of preservation of the movement in the series obtained from a more reliable source. As indicated by this standard, the new benchmarked series δ_υ is expected to follow the pattern in the real series from the reliable source, for instance, data from a government agency. One of the reasons why this holds is that the within-year transition of the existing series c_υ remains the sole accessible series. Denton suggested various meanings for preservation of movement each comparing to a technique.

The first Denton technique has two noteworthy deficiencies:

i. The technique presents a transient development toward the start of the series, which crushes the expressed guideline of movement protection and

ii. Involves a certain figure of the following disparity toward the finish of the series, based on the last two disparities as it were.

The first restriction was comprehended by Helfand et al. (1977) in their multiplicative order one difference variation. Cholette (1979) gave a solution to the challenge in the variants. This results brought an altered variation for every variation of Denton's strategy. Consistent autoregressive errors and parameter with regression- solved benchmarking model explain the second constraints, this of course play out a leftover change in accordance with the fulfilled imperatives. The study catches the normal yearly error and in this way gives a figure of the following disparity that depends on recorded conduct.

2.6 Denton Method: The original First Difference

The fundamental method of benchmarking proposed by Denton is as follows:

$$c_v = \delta_v + e_v$$

$$d_u = \sum_{v=v1,u}^{vL,u} j_{u,v}\delta_v \qquad (2.7)$$

where c_v and δ_v are the original within-year series and the computed benchmarked series. The penalty function determines the pattern of the error:

$$\min(\delta_1 - c_1)^2 + \sum_{v=2}^{V}[(\delta_v - c_u) - (\delta_{v-1} - c_{u-1})]^2$$

$$\min(\delta_1 - c_1)^2 + \sum_{v=2}^{V}[(\delta_v - c_{u-1}) - (\delta_v - c_{u-1})]^2 \qquad (2.8)$$

Inasmuch the Denton method obliges the constraints (δ_0 $-$ c_0). so the benchmarked series is being forced by the first condition through Denton's method to equal the indicator series when the time starts at zero and brings out a result which minimizes the first amendment ($\delta_1 - c_1$). Assuming there is no first condition, (2.8) then makes it known that the result of subtracting the benchmarked from the first series ($\delta_v - c_v$) must be the same throughout the period of time. The first term is omitted by the adjusted variant of the first difference in order to provide solution to the deficiencies described in the preceding section:

2.6.1 Operators of Temporal Additions

U by V is the dimension of the matrix K_n which stands for the operators of temporal addition. It contains the following proportions of coverage:

$$K_n = \begin{bmatrix} g_{11} & g_{12} & \cdot & \cdot & \cdot & g_{1V} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ g_{U1} & g_{U2} & \cdot & \cdot & \cdot & g_{UV} \end{bmatrix}$$

where $g_{uv} = 0$ for $v < v_{1u}$ or $v > v_{Lu}$, and g_{uv} The proportions of coverage is domiciled in u and v in K. Taking for instance, for a 3-month sub-yearly continuous series commencing from March 1, 2015 - November 30, 2018, together with the benchmarks, beginning from March 1, 2015 and ends in March 31, 2017, then the matrix K is:

$$K_1 = \begin{bmatrix} a_1 & a_2 & a_3 & a_{41} & r & r & r & r & r & r & r & r & r & r & r \\ r & r & r & r & b_1 & b_2 & b_3 & b_4 & r & r & r & r & r & r & r \\ r & r & r & r & r & r & r & r & c_1 & c_2 & c_3 & c_4 & r & r & r \end{bmatrix}$$

$Let\ r = 0,\ a,\ b,\ c = 1$ Row one contains the first benchmark, row two has the second benchmark and so on. Column one of matrix K refers to quarter one of 2015, column two to quarter two of 2015...; Column five refers to quarter one of 2016 and so on. Peradventure there is stock series rather than flow series where the values are expected to be measured on the last day of quarter four, the resulting matrix K_2 is:

$$K_2 = \begin{bmatrix} r & r & r & a & r & r & r & r & r & r & r & r & r & r & r \\ r & r & r & r & r & r & r & b & r & r & r & r & r & r & r \\ r & r & r & r & r & r & r & r & r & r & r & c & r & r & r \end{bmatrix}$$

In case we have sub-yearly series commencing from March 1, 2015 and lasts November 30, 2018 and benchmarks following the calendar year beginning in March 1, 2015 and stopping in March 31, 2017, therefore the matrix K_3 is:

$$K_3 = \begin{bmatrix} a & r & r & r & r & r & r & r & r & r & r & r \\ r & a & r & r & r & r & r & r & r & r & r & r \\ r & r & a & r & r & r & r & r & r & r & r & r \end{bmatrix}$$

It should be noted that **a** represents 1 by 12 row vectors that contains only ones, r denotes 1 by 12 row vectors that contain only zeros, assuming that Kn is a quantity having magnitude without direction. Using the accompanying discrete series, the first eleven values of the first set of values will equal zeros, while the twelveth value be equal to one.

Matrix Kn having the dimension $U \times V$ contain benchmarking that are not regularly spaced and inconsistent periods. For instance, quarterly benchmarks may cover some duration of time υ yearly benchmarks may cover some duration υ, quarterly may cover some, and it may happen that some time duration υ may not be taken of.

If the spaces among the benchmarks are regular (for example all annual) then υ will come up as multiple of u it will make the same duration when one benchmark covers all sub-yearly periods, V. The Kronecker products then explains the matrix K:

$$Kn = I_U \otimes 1 \tag{2.9}$$

For instance, if we are interested in a series that is sub-yearly like quarterly in nature having three annual benchmarks (U = 3), then the product of the Kronecker produces this:

$$K_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

It should be noted that a is a vector of row having dimension of 1 by 4 which contains only ones, and r a vector of row having a dimension of 1 by 4 which contains only zeros. Notwithstanding, the occurrence of this case is not common practically, the current year. At the same time, it is easier for analyses that are theory based.

Considering the original and the adjusted Denton's method, we minimize a penalty function (2.7) which depends on the requirements of the benchmarking matrix algebra can be used to present the penalty function:

$$f(\delta, \eta) = (\delta - c)' D' D(\delta - c) - 2\eta' (d - K\delta) \tag{2.10}$$

$$= \delta' E' E\delta - 2\delta' E' c + c' E' Ec - 2\eta' d + 2\eta' K\delta \tag{2.11}$$

Note that c represents the benchmarks, the original and benchmarking series are represented by c and δ respectively. While the Lagranges multipliers which is related to the linear requirements $c - K\delta = 0$ are contained in η. The operators of temporal sum is matrix K. For the case of flow series, the $k = I_U \otimes 1$ and matrix K are equivalent. **1** represents a row vector of order four for quarterly or order twelve for monthly series. For the case of stock series, **1** represents a vector containing zeros within the first three columns for a 4-month series or first eleven columns for monthly series. A common difference operator which is not seasonal is the matrix E.

All the versions of Denton's techniques, both the adjusted and the first version rely on the difference operator that have been taken in penalty function (2.11). The two first difference operators are now considered.

$$E_0^{\,1}\left[V \text{ x } V\right] = \begin{bmatrix} 1 & 0 & 0 & . & . & . \\ -1 & 1 & 0 & . & . & . \\ 0 & -1 & 1 & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \end{bmatrix}$$

$$E^1\left[(V-1) \text{ x } V\right] = \begin{bmatrix} -1 & 1 & 0 & 0 & . & . & . \\ 0 & -1 & 1 & 0 & . & . & . \\ 0 & 0 & -1 & 1 & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \end{bmatrix}$$

The difference observed in the original version of the Denton's first constraints ($\delta_0 = c_0$) can be obtained if $E = E'$; and the difference in the adjusted version without the first requirement. Eq. 2.6

The major requirements needed in order to achieve maximum efficiency demand that the results of mathematical differentiation of the penalty function (2.11) on the condition that the quantities be equal zero:

$$\frac{\partial(f(\delta,\eta))}{\partial\delta} = 2E'E\delta - 2E'Ec + 2K'\delta = 0 \tag{2.12}$$

$$\frac{\partial(f(\delta,\eta))}{\partial\eta} = 2K\delta - 2c = 0 \tag{2.13}$$

Having positive definite for the second order derivatives $\frac{\partial(f(\delta,\eta))}{\partial\delta^2}$ happens to be one of the requirements for a minimum. Matrix $B = E'E$ is actually positive definite since it is of this structure $B = M'M$:

$$E'E\delta + K'\eta = E'Ec \tag{2.14}$$

$$J\theta = d \equiv Kc + (d - Kc) \tag{2.15}$$

where $Kc + (d - Kc)$ replaces d that supplies an easier outcome. Eq. (2.15) may be shown as:

$$\begin{bmatrix} E'E & K' \\ K & O \end{bmatrix}\begin{bmatrix} \delta \\ \eta \end{bmatrix} = \begin{bmatrix} E'E & O_{VXU} \\ K & I_U \end{bmatrix}\begin{bmatrix} c \\ (d - Kc) \end{bmatrix}$$

Re-arranged we have:

$$\begin{bmatrix} \delta \\ \eta \end{bmatrix} = \begin{bmatrix} E'E & K' \\ K & O \end{bmatrix}^{-1}\begin{bmatrix} E'E & O_{VXU} \\ K & I_U \end{bmatrix}\begin{bmatrix} c \\ (d - Kc) \end{bmatrix}$$

Taking the example of the first version of Denton's technique, the inversion of the matrix is obtained through integration by parts. Clearly $E = Eo^1$ of Eq. (2.13), does not have inverse. As good as the method used is, it does not follow natural order in the matrix above but gives this solution:

$$\hat{\delta} = c(E' E)^{-1}K' (K(E' E)^{-1}K')^{-1}(d - Kc) \tag{2.16}$$

2.6.1 Various versions of techniques by Denton

The utilization of the Denton's technique with the accompanying penalty function as a fraction of the variance that exist in the low frequency when compared with benchmarked series:

$$min\ (\delta, c, B) = (\delta - c)'\mathbf{B}(\delta - c) \tag{2.17}$$

The type of the matrix B characterizes the accompanying variations of the first Denton technique:

i. *Additive Balanced Difference* (ABD) if $\mathbf{B} = \mathbf{I}$, (2.18)

ii. *Additive Order One Difference* (AOOD) if $\mathbf{B} = \mathbf{E}'\mathbf{E}$; (2.19)

iii. *Additive Order Two Difference* (AOTD) if $\mathbf{B} = (\mathbf{E^2})'\mathbf{E^2}$; (2.20)

iv. *Proportional Balanced Difference* (PBD) if $\mathbf{B} = \hat{\mathbf{c}}^{-1}\hat{\mathbf{c}}^{-1}$ (2.21)

v. *Proportional Order One Difference* (POOD) if $\mathbf{B} = \hat{\mathbf{c}}^{-1}\mathbf{E}'\mathbf{E}\hat{\mathbf{c}}^{-1}$ (2.22)

vi. *Proportional Order Two Difference* (POTD) if $\mathbf{B} = \hat{\mathbf{c}}^{-1}(\mathbf{E^2})'\mathbf{E^2}\hat{\mathbf{c}}^{-1}$ (2.23)

AOOD and POOD penalty functions remain the only variants used by most researchers when it comes to practical applications. The AOOD and POOD possess the accompanying structures:

$$AOOD = \min(\delta_1 - c_1)^2 + \sum_{v=2}^{V}[(\delta_1 - c_1) - (\delta_{v-1} - c_{v-1})]^2 \tag{2.24}$$

$$POOD = \min\left(\frac{\delta_1}{v_1}\right)^2 + \sum_{v=2}^{T}\left(\frac{\delta_1}{v_1}\right) - \left(\frac{\delta_{t-1}}{c_{v-1}}\right)^2 \tag{2.25}$$

It can be seen that having Denton's technique with AOOD, the problem of benchmarking goes down to make it traceable to reducing the sum of the squares of time to time level changes of the benchmarked and the indicator series on the condition that the benchmarking requirement is hold.

$$\mathbf{Xc} = \mathbf{d} \tag{2.26}$$

The problem of benchmarking comes when $\mathbf{Xc} = \mathbf{d}$, that is, the sub-annual series $\mathbf{c}$ not having consistence pattern with the annual series $\mathbf{d}$. As a result we need to adjust the c in order to have a new high frequency series δ in such a way that (2.26) will be valid. On the other hand, the first difference Denton's approach in (2.25), is gotten by taking the differentials of the series sum squares with respect to time change in the proportions which is dependent on constraints of benchmarking in (2.26).

The first conditions $\delta_0 = c_0$ was imposed by Denton for the method of additive first difference $\delta_0 = c_0$ together with $\delta_{-1} = c_{-1}$ for the approach of the proportional first difference. The meaning of these conditions is that there cannot be any adjustment to the first series especially when such an adjustment is not within the range. The argument is that a movement that does not last is introduced into the series at the beginning through the first conditions, which is at variance with the law of preservation of movement (Helfand et al. 1977, Dagum and Cholette 2006). The adjusted procedure of the Denton's techniques solve this challenge by omitting the first terms in (2.24) and (2.25)

The versions of the adjusted Denton's techniques look alike with the various versions listed above having little disparities. In the case of AOOD and POOD versions, we use E in matrix $E^1\left[(V-1)\,\mathrm{x}\,V\right]$ instead of matrix E in $E_0^1\left[V\,\mathrm{x}\,V\right]$ but in the case of AOTD and POTD versions we use matrix E_c instead of E^2 which can be represented as:

$$E = [(T-2) \text{ x } T] = \begin{bmatrix} 1 & -2 & 1 & 0 & . & . & . \\ 0 & 1 & -2 & 1 & . & . & . \\ 0 & 0 & 1 & -2 & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \end{bmatrix}$$

Whenever the popular AOOD and POTD techniques are used in real life situation, it is essential to check what the benchmarked series δ_v equal or close to. For example, considering a situation when requirements (2.26) for benchmarking are not used. The condition for maximum efficiency for Denton's AFD technique is:

$$\frac{\partial(f(AOOD))}{\partial \delta_v} = 0 \qquad \text{which gives}$$

$$\delta_v^{\ *} = c_v + \frac{\left(\delta^*_{\ v+1} - c_{v+1}\right) + \left(\delta^*_{\ v-1} - c_{v-1}\right)}{2}, v \neq 1, N \tag{2.27}$$

that is, the discrepancies between the obtained high frequency and the original high frequency values $\delta_v^* - c_v$, is the same as the mean of the two adjacent level discrepancies. (2.26) When the requirement of benchmarking according to (2.27) are not met, it means that the issue should include them, (in a setting of Lagrangian). On the other hand, the POOD, in the same vein , the situation also rely on the strength of non-observance of the benchmarking requirement in term of keeping consistency, and this will be close to the solution in (2.27). Similarly, the POOD technique, the order one condition $\frac{\partial(f(POOD))}{\partial \delta_v} = 0$, not minding (2.26), gives:

$$\partial_v^* = \frac{c_v}{2} + \left(\frac{\partial^*_{v+1}}{c_{v+1}} + \frac{\partial^*_{v-1}}{c_{v-1}}\right), v \neq 1, N \tag{2.28}$$

Which means $\frac{\theta_t^*}{s_t}$, the ratio of the adjusted high frequency values (benchmarked) and the original sub-annual values equals the arithmetic mean of the duo adjacent ratios. So whenever the requirements of benchmarking (2.26) are considered, the answer of the POOD technique will be very close to the one revealed in (2.28). These values depicts the following results:

i. The versions of Denton's AOOD and POTD can be summarily referred to as smoothening techniques. For instance, using (2.27) it can be said that when non-negative, as a result the calculated δ_v will be non-negative (it cannot be zero). According to the literature, researchers claimed that the initial and the earlier versions of Boot et al. (1967) technique are smoothening methods. These are applied in obtaining sub-annual series when benchmark series are not available (Eurostat 1999).

ii. It can be observed that the first value depicts the fractional difference which is a version of the Denton's technique, this version preserves values from c to δ better than the versions of his methods (2.28). This fact in a way exposes why many benchmarking researchers preferred the POOD Denton technique over others versions.

iii If the original sub-annual series disappears easily and accept both the non-negative and negative observations, and we are taking benchmarking requirements into consideration, then (2.27) and (2.28) means that there is going to be an unwelcome inconsistent outcome. In order words, we will have $\delta_v^* < 0$ with $c_v > 0$. Therefore, the Denton's techniques AOOD and POOD do not validate both the preservation sign principle and preservation movement principle.

2.7 Background of the Denton methods

What Denton (1971) provided as solution to the issue of benchmarking which was generalised by Fernardez (1981) comprises of looking for an indicator series that will show the movement and such series is expected to produce yearly sums that will give reliable and effective annual benchmarks. The yearly values will result to yearly benchmarks, its pattern in term of movement will be determined solely by the indicator series. Moreover, the modified or benchmarked data will run in parallel with the original indicator series, this will still hold when the benchmarking constraints are met.

The following expression reveals the modification ($\delta_v - c_v$) effected on the original series c_t using the Denton's additive first difference. Bringing back the AOOD method of solving benchmarking problem the expected series c_v minimizes the following penalty function:

$$p(\delta) = \sum_{v=1}^{y} (\Delta\delta_v - \Delta c_v)^2 = \sum_{v=1}^{y} \left(\Delta(\delta_v - c_v)\right)^2, \quad \Delta\delta_0 = \Delta c_0 \tag{2.29}$$

where c_v represents the original high frequency series at time t. The resulting function can now be minimised if subjected to the benchmarking requirements that exist between the year totals of the values obtained and the benchmarks d_i that are available:

$$\sum_{v=(i-1)k+1}^{tk} \delta_v = d_{i,} \qquad i = 1,\ 2,\ .\ .\ .\ ,u \tag{2.30}$$

Note that q represents each three months per quarter in a year. The hypothesis $\Delta\delta_0 = \Delta c_0$ is validated by Denton, establishing the fact it is right to assume that the last fitted and observed values are equal. The penalty function (2.29) is now giving same slope for adjusted series otherwise known as the benchmarked δ and the original series c_v also known as the indicator series. As a result, the slope obtained from the differences from the two series is now minimised, taking into consideration the benchmarking constraints. On substitution $\Delta\delta_0 = \Delta c_0$, the penalty function is now:

$$p(\delta) = (\delta_1 - \delta c_1)^2 + \sum_{v=2}^{q} (\Delta(\delta_v - c_v))^2 \tag{2.31}$$

The change explains that the condition $\Delta\delta_0 = \Delta c_0$ implies the initial correction has to be minimised. Minimisation of the initial correction reduces the adjustment line to zero direction at the beginning of the series. The aftermath is that a wave will be produced in the initial year that will be transferred to parallelism between the original and the benchmarked series. The outcome here differs a little bit from

$\Delta\delta_0 = \Delta c_0$ and as a result the following penalty function:

$$(\Delta(\delta_v - c_v))^2 \tag{2.32}$$

obeying the rule in (2.30)

As we have it in linear algebra, the penalty has constraints is given by:

$$\underline{u}(\underline{\delta}, \underline{\eta}) = (\underline{\delta} - \underline{c})' B (\underline{\delta} - \underline{c}) - 2\underline{\eta}'(\underline{d} - \underline{X}'\underline{\delta}) \tag{2.33}$$

The following vectors and matrices are part of it:

$$\underline{\delta}_{y\,\mathrm{x}\,1} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ . \\ . \\ . \\ \delta_y \end{bmatrix}, \quad \underline{c}_{y\,\mathrm{x}\,1} = \begin{bmatrix} c_1 \\ c_2 \\ . \\ . \\ . \\ c_y \end{bmatrix}, \quad \underline{d}_{u\,\mathrm{x}\,1} = \begin{bmatrix} d_1 \\ d_2 \\ . \\ . \\ . \\ d_u \end{bmatrix}, \quad \underline{q}_{u\,\mathrm{x}\,1} = \begin{bmatrix} \eta_1 \\ \eta_2 \\ . \\ . \\ . \\ \eta_n \end{bmatrix},$$

$$\underline{B}_{y\,\mathrm{x}\,\mathrm{y}} = \underline{E}'\,\underline{E}. \quad \underline{E}_{(y\text{-}1)\,\mathrm{x}\,y} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \end{bmatrix}$$

$$\underline{X}_{q\,\mathrm{x}\,1} = \begin{bmatrix} \underline{j} & 0 & . & . & . \\ 0 & \underline{j} & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{bmatrix}, \quad \underline{j}_{q\,\mathrm{x}\,u} = \begin{bmatrix} 1 \\ 1 \\ . \\ . \\ . \end{bmatrix}, \quad (y = qu) \qquad (2.34)$$

Vector η has the multipliers of the Lagrange. $y(= uq)$, where y represents the number of observations, and u stands for the number of years, while q is the quarters number in a year. The corresponding equations using the penalty function are:

$$\frac{\partial u}{\partial \delta} = (\underline{B} + \underline{B}')(\underline{\delta} + \underline{c}) + 2\underline{G}\underline{\eta}' = 0$$

$$\frac{\partial u}{\partial \eta} = 2\left(\underline{G}'\underline{\eta}' - \underline{d}\right) = 0 \qquad (2.35)$$

which results into:

$$\begin{bmatrix} \underline{\theta} \\ \underline{\eta} \end{bmatrix} = \begin{bmatrix} \underline{B} & \underline{G} \\ \underline{G}' & \underline{0} \end{bmatrix}^{-1} \begin{bmatrix} \underline{B} & \underline{0} \\ \underline{0} & \underline{I} \end{bmatrix} \begin{bmatrix} \underline{c} \\ \underline{d} \end{bmatrix} = \underline{W}_{(y+u)\mathrm{x}(y+u)} \begin{bmatrix} \underline{c} \\ \underline{d} \end{bmatrix} \qquad (2.36)$$

replacing identity $d = G'+t$, where t contains the yearly differences:

$$\begin{bmatrix} \underline{\delta} \\ \underline{\eta} \end{bmatrix} = \begin{bmatrix} \underline{B} & \underline{G} \\ \underline{G}' & \underline{0} \end{bmatrix}^{-1} \begin{bmatrix} \underline{B} & \underline{0} \\ \underline{0} & \underline{I} \end{bmatrix} \begin{bmatrix} \underline{c} \\ \underline{a} \end{bmatrix} = \begin{bmatrix} \underline{I}_{((nxn))} & \underline{W}_{x\,(yxu)} \\ \underline{0}_{(uxy)} & \underline{W}_{1\,(uxu)} \end{bmatrix} \begin{bmatrix} \underline{c} \\ \underline{d} \end{bmatrix} \Rightarrow \underline{\delta} = \underline{c} + \underline{Z}_w \underline{v} \tag{2.37}$$

The redevelopment of the solution makes the computing period reduced when applying the computed weights Z_w if compared to the solution (2.36). It can be observed that the weights which could be used for different kind of series using similar observation number. The inversion in (2.36) solved by parts is carried out by Denton.

The developed technique makes use of the provision by Boot, Feibes, and Lisman (1967) is to carry out interpolation between yearly data when high frequency data is not available. The provision in (2.37) strictly has interpolation between the yearly discrepancies with the techniques produced by the authors and in summing up the obtained estimates with the original high frequency series.

2.7.1 Proportional Variant

The proportional techniques discussed here is a version of the Denton's proportional technique, the $\delta_0 = c_0$ was no longer included. Just like as we have it during the previous sub-topic, the penalty function differentiates the totals of the dissimilarities after squaring within the interval of the estimates of parameter of the yearly sums and the high frequency series (that is, c_v and δ_v). Each of the terms in the sum measured by the estimate of the corresponding high frequency value:

$$p(\delta) = \sum_{v=2}^{y} \left(\frac{\Delta(\delta_v - c_v)}{c_v} \right)^2 = \sum_{v=2}^{y} \left(\Delta\left(\frac{\delta_v}{c_v} \right) \right)^2 \tag{2.38}$$

This version is appropriate for series having strong seasonal effects, if it is assumed that quarters having seasonal effect cannot account for the discrepancies that occur in the annual data. The level of each observation is proportional to the size of the corrections made. It is also established that considering proportional version, all values and all the modified values will be positive.

According to (Cholette, 1978, 1979) the proportional version approximates the non-linear preservation method growth rate, this is corroborated by (Smith and Hidiroglou, 2005), and the resulting penalty function is:

$$p(\delta)=\sum_{v=2}^{y}\left(\frac{\delta_v}{\delta_{v-1}}-\frac{c_v}{c_{v-1}}\right)^2 \tag{2.39}$$

Whenever the yearly differences of the proportion on the interval of estimation is constant, the approximation is exact. The penalty function which is constrained in linear algebra that has to do with proportional technique is:

$$\underline{u}(\underline{\delta},\underline{\eta})=(\underline{\delta}-\underline{c})'\underline{C}^{-1}\underline{B}\underline{C}^{-1}(\underline{\delta}-\underline{c})-2\underline{\eta}'(\underline{d}\quad-\underline{G}'\underline{\delta}) \tag{2.40}$$

Where $\underline{C}^{-1}$ represents a diagonal matrix having $\frac{1}{c_1},\frac{1}{c_2},\ldots.$ the solution from this technique has the same pattern as the additive version ($\underline{C}^{-1}\underline{B}\ \underline{C}^{-1}$substituting B in (2.37) and:

$$\begin{bmatrix}\underline{\delta}\\ \underline{\eta}\end{bmatrix}=\begin{bmatrix}\underline{C}^{-1}\underline{B}\underline{C}^{-1} & \underline{G}\\ \underline{G}' & \underline{0}\end{bmatrix}^{-1}\begin{bmatrix}\underline{C}^{-1}\underline{B}\underline{C}^{-1} & \underline{0}\\ \underline{G}' & \underline{I}\end{bmatrix}\begin{bmatrix}\underline{d}\\ \underline{d}\end{bmatrix}=\begin{bmatrix}\underline{I} & \underline{Z_w}\\ \underline{0} & \underline{X_1}\end{bmatrix}\begin{bmatrix}\underline{d}\\ \underline{d}\end{bmatrix} \tag{2.41}$$

The quantities Z_x in the proportional solution will be calculated for each data series, also for every interval of application in a particular series, that is not so in the quantity considering the additive versions

2.8 The Cholette and Dagum Technique

A Benchmarking technique suggested by Cholette and Dagum (1994) was based on generalised least squares, regression model. The model considers (i) bias in the high frequency series and (ii) heteroscedacity in the first series. The technique proposed by Denton can also be referred to as a special case of the Cholette regression model. The technique of benchmarking by Cholette and Dagum (1994) has its root on just two expressions. it is assumed that there are no missing observations in the two series, and that each yearly total is supplied by the yearly quarterly values:

$$c=d_v+\delta_v+e_v \quad \text{for } v = 1, 2, \ldots, q \tag{2.42}$$

$$B_n=\sum_{v=4y-3}^{4y}\delta_v+z_y+e_v \quad \text{for y} = 1, 2, \ldots, b \tag{2.43}$$

where

δ_v represents the values in the quarters;

d_v stands for the total effect
c_v represents the sub-annual series
e_v the sub-annual serial-correlated residuals

z_y the yearly heteroscedastic residuals from the yearly data d_y, not related with e_y,

having

$E(e_v)=0,\ \ E(e_v e_{v-b})\neq 0$ expected value of the error is zero and the autocorrelation of the residuals is not equal to zero
$E(z_y)=0,\ \ E(z_y^2)=\delta_y^2$ the variance of the residuals is constant
$E(e_v z_y)=0$ the autocovariance of the error is zero

The Cholette and Dagum technique using Autoregressive error is carried out in order to improve on the quarterly national extrapolations. This joining reveals the assumptions underlying the definition of the general regression based model as stated in (2.42) - (2.43), the implementation of this model is better carried out using the matrix notation

The quarterly series c_v happens to be the determinant of δ_v and mingled with the effect d_v and with high frequency residual e_v. The equation in (2.42) associates the yearly benchmarks d_y to the sum total of the sub-annual observations e_v having some estimates of residuals z_y. The regression-based technique changes with respect to the hypothesis for the effect d_v, sub-annual residual e_v, and the yearly residuals z_v.

One of the roles of the yearly residual or error z_v is to give explanation to cases when the benchmark is also dependent on the residual. This kind of benchmarks are referred to as non-binding. This kind of benchmarks are dependent on the changes in the processes of benchmarking. Considering the quarterly national accounts, the yearly totals are binding constraints in most occasions for the quarterly observations, that is (i.e., $E(z_y^2)=0$).

y

What brings about the combined effects d_v is the combined regressors $t_{v,b}$ when multiplied by their coefficients of regressions β_b, that is

$$d_v = \sum_{m=1}^{p} t_{v,b}\beta_h \tag{2.44}$$

g is the combined effects that was investigated. An unchanging entity was employed to capture the difference in the bias level which may exist between the yearly and high frequency series. A constant bias is also adjusted by reducing the indicator series using the benchmarking-indicator (B-I) ratio. This kind of change is easy since it is not a strict requirement in this case to produce any estimation of parameter in the level bias. A diverging way between the sub-annual and the penalty function can be caught using a combined trend. On the other hand, extrapolations can be caused at the two ends of the series through the combined trend, therefore adequate caution should be taken.

The residual e_v is the sub-annual difference between the expected variable δ_v and the sub-annual indicator c_v. For the reason that an important aim of benchmarking is to preserve the movement pattern in Since a key objective of benchmarking is to keep the developments in δ_v as close as conceivable to the pattern in c_v, the residual e_v has two main features:

i. be similar in ratio to the value of the high frequency series c_v. This acquired quality is needed in distributing the deviations around the level of the high frequency series, likewise to the solution provided through the proportion method by Denton.

ii. introduce freedom from obstruction movements from one quarter to another one. A distribution which is smooth in nature of e_v will sure make movements of δ_v and c_v very close to one another.

In order to have an adjustment that is similar in ratio, the values of the high frequency series c_v are used to standardize the error or deviation e_v

$$e_v' = \frac{e_v}{c_v} \tag{2.45}$$

The result of this is that the indicator series will be assumed to be equal to the standard deviation of e_v. (This condition shows that the coefficient of variation will be same throughout, which can be expressed as $\frac{\sigma_v}{c_v}$ is equal to one for any chosen t quarter.) In order to have a non-distorted distribution the standardized error e' is expected to be an autoregressive stationary model at first difference:

$$e_v' = \phi e_{v-1}' + u_v \tag{2.46}$$

It should be noted that | φ |< 1, and u_v's are expected to be normally distributed, i.e:

$E(u_v) = 0,\ E(u^2) = 1,\ E(u_v,\ u_{v-vb}) = 0$ true for all v & b.

This method having an autoregressive error involves the differentiation of a penalty function having a relationship with the proportional condition which has been minimised by Denton. It is evident that the adjusted high frequency series, otherwise known as the benchmarked series of Cholette similar in ratio model with autoregressive model (2.46) minimizes the penalty function:

$$\min_{\delta_v}\left(\frac{1}{1-\phi^2}\right)\left(\frac{\delta_1}{c_1^d}\right)^2+\sum_{v=2}^{q}\left[\frac{\delta_v}{c_v^d}-\phi\frac{\delta_{v-1}}{c_v^d}\right]^2 \tag{2.47}$$

The function stated in (2.47) explains that apart from extrapolation, the autoregressive parameter φ performs important function in preserving the short term dynamics of the high frequency series. It is assumed that the closer the parameter to 1, the faster the function (2.47) converges to this function:

$$\min_{\delta_t}\sum_{v=2}^{4z}\left[\frac{\delta_v-c_v}{c_v}-\frac{\delta_{v-1}-c_{v-1}}{c_{v-1}}\right]^2=\min_{\delta_q}\sum_{v=2}^{4z}\left[\frac{\delta_v}{c_v}-\frac{\delta_{v-1}-\delta_{v-1}}{c_{v-1}}\right]^2 \tag{2.48}$$

This function is minimised by the Denton's proportional technique. It is noticed that the movement preservation becomes weaker as φ deviates from 1. And this affect the benchmarking to indicator ratio.

2.8.1 Cholette and Dagum Technique using Matrix

The autoregressive error benchmarking solution proposed by Cholette and Dagum is:

$$\delta = c + X_eK^{\lrcorner}(KX_eK^{\lrcorner})^{-1}[d - Kc] \tag{2.49}$$

where

c^* the V x 1 dimension having the bias adjusted high-frequency series
d the M x 1 dimension of the yearly benchmarks,

δ the V x 1 dimension for the modified high frequency series of benchmarked data,

K a U x V dimension,

V the frequency of high frequency periods (months or quarters),

U the frequency of years where yearly totals is obtainable

$X_e = diag(c^*)(\Psi^{-1})diag(c^*)$ is the V x V matrix of variance and covariance of the error of high frequency series e_v

Ψ is the matrix of autocorrelation of the Autoregressive model of order one using parameter φ:

$$\Psi = \begin{bmatrix} 1 & \phi & \phi^2 & . & . & . & \phi^{V-1} \\ \phi & 1 & \phi & . & . & . & \phi^{V-2} \\ \phi^{V-2} & \phi & 1 & . & . & . & \phi^{V-3} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \phi^{V-1} & \phi^{V-2} & \phi^{V-3} & . & . & . & 1 \end{bmatrix}$$

2.9 Method of Pro-rata

A careful examination on an instance when the high frequency and benchmarking series have only elements that are positive in nature. This part deals with bench- marking constraints as addition of flow variables. The distribution of pro-rata forces the high frequency series to be consistent with the benchmark series by auditing the annual sums proportionally to the portions of the sub-period observations in the corresponding aggregates of the indicators.

One of the objectives of this section is to show how pro rata distribution creates step problem in the series and how it performs extrapolation commencing it from last benchmark. The instance given is given is that if ratio of quarterly national accounts estimates of benchmarked to the indicator, this is also known as quarterly BI ratio. This outcome brought about by method of pro rata distribution reveals that this technique brings unwelcomed discontinuities into the series.

Revealing how the quarterly benchmarking to indicator ratio which is produced by the technique of distribution and the quarterly benchmarking to indicator ratio produced by extrapolation from an indicator method help us to know how distribution and extrapolation using high frequency series can fit into a unique framework of benchmarking to indicator ratio. Now having established this, the method of pro rata distribution is unacceptable in solving benchmarking problem.

$$\delta_v = c_v \left[\frac{z_y}{\bar{c}_y} \right] \quad \text{for } y = 1, \ldots, q \text{ and } v = 4y - 3. \ldots 4y \tag{2.50}$$

where

δ_v is the dimension of the quarterly countries estimate for quarter v, c_v is the dimension of the quarterly indicator for quarter t,

z_y is the dimension of the yearly countries estimate for year y,

$\bar{c}_y$ is the yearly aggregate (total or average) of the quarterly estimates of the indicator for year y,

y the years,

q the last accessible year y, q the last available year,

v the temporal index for the quarters.

According to literature, distribution in this regard is the allocation of a yearly total continous series to its every 3-month in the particular period. In summary, a prorata technique distributes the yearly sum based on the corresponding proportions by the four quarterly observations.

The step problem occurs due to lack of continuities in the yearly benchmarking to indicator ratio between the years. Immediately, a difference is noticed in yearly growth rates in higher or lower than the yearly benchmark, then benchmarking to indicator ratio has to move from one year to the next year. Each time the bench- marking to indicator ratio is used in increasing the indicator’s value for all the quarters in the year, the whole difference in the sub annual growth rates will not change. However, the variations size in the yearly benchmark to indicator ratio determines the devastating effect of the step problem in any series.
Extrapolation using an indicator in this context is using the transition in the original series to update the estimates of quarterly national accounts time series. For the sub-annual series, for which no yearly data are available.

Employing mathematical notations, extrapolation using an indicator can be made formal by making use of the same benchmarks to indicator ratio presentation for the distribution instance.

$$\delta_v = c_v \left[\frac{z_q}{\bar{c}_q} \right] \text{ for } v = 1, \ldots\ldots, q \text{ and } v = 4q + 1, 4q + 3, 4(q + 1)$$

where q stands for the year having the last accessible yearly benchmark and extrapolations can be provided for the quarterly observations of the year $q + 1$. It is also assumed that the indicator is made available for every quarter of year $q + 1$.

With the application of equation (2.20) the rates of growth observed in quarters found in the original series in the year $q + 1$ is exactly reproduced by rates of growth in the forward series. Considering the equation (2.20), the BI ratio common for year q located in the right-hand after computation produces:

$$\left(\frac{\delta_{4q+i}}{\delta_{4q+i-1}} \right) = \left(\frac{\delta_{4q+i}}{c_{4q+i-1}} \right) \quad \text{where } i = 1, 2, 3, 4$$

Likewise, it is easily seen that the countries' quarterly accounts series has equal year-on-year indicator rates of growth in fixing quarters outside the range. Notwithstanding, generally these attributes appear like the expected qualities, the series of a function outside the range of the known values need to be at variance with the pattern in the indicator to match unequal annual patterns in the yearly countries' accounts for the following years.

Summarily, the back series using the corresponding benchmark to indicator ratios for each year is calculated. The annual national accounts should be made available to serve as adjustment factors to increase or reduce the indicator and the forward series by bringing the last yearly BI ratio forward. This technique is not acceptable for quarterly national accounts reasons because step problem can be introduced in the year's first quarter, and as a result, the objective of preservation of the original patterns in the indicator is violated. It is well believed by many researchers that socio economic variable does not grow within the year, not knowing that growth is not impossible within and between the years.

2.10 Benchmarking Method using the Entropy

Obtaining the parameter back in a faulty system, we use a generalised cross-entropy approach. It is used whenever the pattern preservation is done through the additive order one difference and proportional order one difference techniques by considering the high frequency series (Temurshoev, 2012).

2.10.1 Generalisation of Cross-entropy method

The linear inverse problem having noise equation can be considered using the following (Golan, Judge, and Miller (1996)):

$$w = \Gamma\delta + e \tag{2.51}$$

Where w is a vector of I-dimension containing observations, δ is a vector of N-dimension having unknown parameters, is also known as the benchmarked series Γ is a linear operator matrix with non-square attribute, and e is the vector of the disturbance. It is in cases that $N > I$ which implies that it is an undermined system. This means in the normal regression, the number of parameters is larger than the number of observations. As a result, the conventional method of regression such as the ordinary least square may not be suitable for estimating δ.

The ill-posed problem can be handled by the entropy-based methods. These technique brought about by Golan et al. (1996) are the core of a generalised cross entropy (GCE). Formalism of entropy is done through theory of information provided by Shannon (1948) and the investigations of Kullback and Leibler (1951), Janes (1957a) and Janes (1957b).

The first step in GCE is to use information having no sample about the unknown parameter parameter δ otherwise known as signal component and disturbance component e. Among the content of the information are background beliefs about the signs, dimensions, and ranges of commendable, value of these components which are unknown. The next step is to obtain discrete random variables having prior information on δ and e.

The GCE beginning stage is that one needs to utilise his/her constrained earlier or non-test data about the obscure parameters δ (additionally called the flag segment) and noise part e.

This data incorporates earlier convictions about the signs, extents, or potentially scopes of conceivable estimations of these obscure parts. Then one develops discrete arbitrary variables with earlier loads (probabilities) and limited backings that is steady with the given non-test data about δ and e. As needs be, the straight reverse issue is re-parametrised as far as discrete irregular variables on limited backings, and the estimation issue winds up recouping "back" likelihood circulations for δ and e predictable with the accessible earlier data and observed sample information.

According to Golan et al. (1996) δ can be represented by the expected value of the random variables having compact supports. The discrete random variable having a compact of K feasible outcomes $t_g = (t_{g1}, \ . \ . \ . \ , t_{gJ})$, note that $2 \leq J < \infty$, and t_{g1} and t_{gJ} which are the lower and upper bounds of δ_g. Thus, δ_g is communicated as:

$$\delta_g = \sum_{j=1}^{J} t_{gj} \tilde{x}_{gj} = \mathrm{t'_g} \, \tilde{\mathrm{q}}_{\mathrm{j}} \tag{2.52}$$

Note that $\tilde{\mathrm{x}}_{\mathrm{g}}$ is a K-vector of positive probabilities having totals equal one (this is not to be taken as $\mathbf{x_r}$). If the matrix arrangement of the combination which is convex is done, then δ may be denoted by:

$$\delta = \mathrm{T}\tilde{\mathrm{q}} = \begin{bmatrix} \mathrm{t'_1} & 0' & . & . & . & 0' \\ 0' & \mathrm{t'_2} & . & . & . & 0' \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0' & 0' & & & & \mathrm{t'_M} \end{bmatrix} \begin{bmatrix} \tilde{\mathrm{q}}_1 \\ \tilde{\mathrm{q}}_2 \\ . \\ . \\ . \\ \tilde{\mathrm{q}}_{\mathrm{M}} \end{bmatrix} \tag{2.53}$$

where $\mathbf{T}$ is a $MXMJ$ matrix, $\tilde{q}$ is a MJ -vector of probabilities, $\mathbf{0}$ is a matrix of ones having the correct structure.

In the same vein, the outcome of error calculation can be shown by taking each e_h like a countable random variable having $2 \leq Y < \infty$ results. Assuming we have lower and upper limits s_{h1} and s_{hY} for each e_h in such a way that $1-(s_{h1} < e_h < s_{hY})$ is made small without following a particular rule. Each of the disturbances can be expressed as:

$$e_h = \sum_{h=1}^{J} u_{hy} z_{hy} = \mathrm{u'_h}\ \mathrm{z_h} \tag{2.54}$$

where $\mathbf{u_h} = (\mathbf{u_{h1}}, \ . \ . \ . \ , \ \mathbf{u_{hY}})'$ is finite support for e_h and $\mathbf{z_h} = (\mathbf{z_{h1}}, \ . \ . \ . \ , \ \mathbf{z_{hY}})'$ is a Y - dimensional vector of positive weight whose totals is equal to one. Therefore, the I-dimensional vector of unknown errors can be expressed as:

$$\mathrm{e} = \mathrm{Xz} = \begin{bmatrix} \mathrm{u'_1} & 0' & . & . & . & 0' \\ 0' & \mathrm{u'_2} & . & . & . & 0' \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0' & 0' & & & & \mathrm{u'_I} \end{bmatrix} \begin{bmatrix} \mathrm{z_1} \\ \mathrm{z_2} \\ . \\ . \\ . \\ \mathrm{z_I} \end{bmatrix} \tag{2.55}$$

It should be noted that **X** has a size of $H \times HY$ matrix and **z** is an HY - vector of non-negative probabilities. Making use of the unknown restructured parameters in models (2.10.13) and (2.10.15), and model (2.10.11) can be expressed as:

$$\tilde{\mathbf{w}} = \mathbf{\Gamma}\delta + \mathbf{e} = \mathbf{\Gamma T\tilde{q}} + \mathbf{Xz} \tag{2.56}$$

The information having no sample on δ and **e** can be indicated as a collection of biased probability distributions on their agreeing supports **T** and **X**. Let **r** has MJ dimensional vector having prior probabilities for the parameter δ, yet not known, therefore **Tr** is the prior mean for δ. In the same way, if **v** has HY dimensional of probabilities on residuals e having average of prior of **Xv**. Hence,

$$min_{\tilde{q},z} H(\tilde{\mathbf{q}},\ \mathbf{r}, \mathbf{x}, \mathbf{v}) = \tilde{\mathbf{q}}'\mathbf{log}(\tilde{\mathbf{q}}/\mathbf{r}) + \mathbf{z}'\mathbf{log}(\mathbf{z}/\mathbf{v})) \tag{2.57}$$

subject to

$$\tilde{\mathbf{w}} = \mathbf{\Gamma T\tilde{q}} + \mathbf{Xz}, \tag{2.58}$$

$$h_{\mathbf{M}} = (\mathbf{H_M} \otimes h'_{\mathbf{J}})\tilde{\mathbf{q}}, \tag{2.59}$$

$$h_{\mathbf{H}} = (\mathbf{H_H} \otimes h'_{\mathbf{Y}})\mathbf{z}, \tag{2.60}$$

where the $/$ in equation (2.57) indicates an element-wise divisions while agreeing dimensions are represented by the subscripts of the identity matrices and vectors of summations.

Kullback (1959) proposed the rule of minimum discrimination information (MDI). This principle serves as background for the GCE penalty function (2.56). It explains that $\tilde{\mathbf{q}}$ and $\mathbf{z}$, which are the new distributions provided with the restrictions of the model in (2.56), are estimated in such a way that they are treated unequally with the original distributions which are $\mathbf{r}$ and $\mathbf{v}$, separately. The MDI can also be referred to as the minimum cross entropy principle. In case, $\mathbf{r}$ and $\mathbf{v}$, which are the prior information becomes consistent with data, therefore the solution of the MDI also becomes $\tilde{\mathbf{q}} = \mathbf{r}$ and $\mathbf{z} = \mathbf{v}$ with $H(\tilde{\mathbf{q}}, \mathbf{r}, \mathbf{z}, \mathbf{v}) = \mathbf{0}$. This also means that the series does not have any information in common to the prior.

Analysts choose prior as a uniform distribution as a result of unavailability of non-sample information information. In such an instance, the cross entropy is now the same as a maximum entropy technique. Therefore, the MDI becomes an extension of maximum entropy rule as deduced by Janes (1957,a,b) which states that be- fore an inference is made on partial information basis, the probability distribution having maximum entropy must be known (Janes 1957a), Cover and Thomas, (2006).

It should be noted that the model (2.58) is the constraint of consistency, while (2.59) and (2.60) are the expected constraint for additivity with respect to $\tilde{\mathbf{q}}$ and $\mathbf{z}$, respectively. The set of additivity constraints strictly determines the GCE objectives $H(\tilde{\mathbf{z}}, \mathbf{r}, \mathbf{z}, \mathbf{v})$ A unique solution will exist if there is interaction between the constraints of the consistency and additivity. With the application of Lagrangean, the solution of the GCE problem can be expressed as:

$$\tilde{q}_{mj} = \frac{r_{mj} \exp\left(t_{mj} \Gamma'_m \hat{\phi}\right)}{\Theta_m\left(\hat{\phi}\right)} \tag{2.61}$$

and

$$\tilde{z}_{hy} = \frac{v_{hy} \exp\left(t_{hy} \hat{\phi}_h\right)}{\Omega_h\left(\hat{\phi}_h\right)} \tag{2.62}$$

where $\mathbf{\Gamma_m}$ is the mth column of $\mathbf{\Gamma}$, φ is an $H \times 1$ vector of Lagrange multipliers, and the normalization factors are:

$$\Theta_m\left(\hat{\phi}\right) = \sum_{j=1}^{J} r_{mj} \exp\left(t_{mj}\Gamma'_m\,\hat{\phi}\right) \tag{2.63}$$

and

$$\Omega_h\left(\hat{\phi}_h\right) = v_{hk} \exp\left(u_{hk}\hat{\phi}_h\right) \tag{2.64}$$

The additivity constraints and non-negativity condition are satisfied, by the GCE solutions $\hat{\mathbf{q}}$ and $\hat{\mathbf{z}}$. But there cannot be a closed-form solution because the solutions largely rely on φ which are not determined by the problem order one conditions. Therefore, it must be computed numerically using a reliable algorithm on the basis of dual GCE technique which does not have constraints (Golan et al. 1996). The point estimates of the unknown parameters and error vectors, i.e., $\hat{\delta} = \mathbf{T}\hat{\mathbf{q}}$ and $\mathbf{e} = \mathbf{X}\hat{\mathbf{z}}$ are computed by the vectors optimal probability.

CHAPTER THREE

METHODOLOGY

3.1 Proposed Method

The proposed technique for benchmarking examined here depends on Denton (1971) and Dagum and Cholette (2006). The extension to their work is in formulating new models through the introduction of various levels of coefficient of autocorrelation. Various derived models are used to extrapolate new series from the existing ones.

The principal contribution of this research is utilising a regression based model for seat stamping and to write a program for unraveling Denton's techniques. The major aim therefore is to provide an alternative solution to the problem of benchmarking in economic data through improving on the order one and order two proportional benchmarking models using restrained minimised quadratic polynomials.

3.2 Alternative solution to the benchmarking problem in Economic data

This proposed benchmarking strategy depends on the summed up least squares regression model. It is grounded on a factual model that takes into consideration (i) the nearness of bias and autocorrelated errors in the pointer, and (ii) the nearness of non-restricting benchmarks. The benchmarked series is determined as the summed up least squares arrangement of a relapse model with deterministic impacts, autocorrelated and heteroschedastic disturbance.

Issue of different pattern and various accuracy level when two same variables are collected from different sources is what leads to problem of benchmarking. Normally, data from one source is expected to have higher frequency than the other series. A good example is a monthly series from other sources. Literature has it proven that series having low frequency are more accurate and reliable than the high frequency series. Therefore, as a result, the series with low frequency is universally regarded as benchmark. Disagreement is always noticed between the benchmarks and the 12 months sums of the high frequency series. What benchmarking does is bringing together the effective power embedded in both the pattern of the high frequency and the benchmarks consistency.

For ease of identification of terms and definitions in the study, the within-year or sub-yearly data are also known as high frequency series, and the yearly or twelve months data are the same as benchmarks. The within-year and the yearly series can be represented by:

$$t = 1, 2, 3, \ldots, T \tag{3.1}$$

$$m = 1, 2, 3, \ldots, M \tag{3.2}$$

note that $1, 2, \ldots, T$ are the neighbouring months, quarter, days, and so on; also note that $1, 2, \ldots, M$ are the subsequent existing benchmarks in the economic series under consideration.

The model contains the following two attributes:

$$\begin{aligned} s_t &= \sum_{h=1}^{H} r_{th} B_h + \theta_t + e_t \\ &= \sum_{h=1}^{H} r_{th} B_h + \theta_t + \sigma_t^{\lambda} e_t^{*} \end{aligned} \tag{3.3}$$

where

$$E(e_t) = 0, \quad E(e_t e_{t-1}) = \sigma_t^{\lambda} \sigma_{t-1}^{\lambda} w_l$$

S_t = the sub-yearly or indicator series

$\sum_{h=1}^{H} r_{th} B_h$ = deterministic time impact

e_t = $\sigma_t^{\lambda} e_t^{*}$ the autocorrelated error

r_{th} = deterministic regressors (h = 1, . . . , H)

e_t^{*} = the standardized error of mean zero and unit variance

w_l = $\phi^{|1|}$, 1 = 0, . . . , T-1 are the autocorrelations

θ_t = genuine qualities that fulfills the yearly constraints, they are the benchmarked values

λ = change model parameter (0, 0.5, or 1)

$$a_m = \sum_{t=t_{1m}}^{t_{Lm}} j_{mt} \theta_t + \varepsilon_{m}, \quad m = 1, \ldots, M \tag{3.4}$$

where

a_m = yearly series

$\sum_{t=t_{1m}}^{t_{Lm}} j_{mt}$ = aggregating coverage fraction with the true values

j_{mt} = coverage fractions

ε_m = error having no autocorrelation but with heteroscedastic differences

t_{1m} = the first sub-yearly period by benchmark a_m

t_{2m} = the first sub-yearly period by benchmark a_m

Equations (3.3) and (3.4) can be written thus:

$$\mathbf{S} = \mathbf{R}\beta + \theta, \ \mathbf{E}(\mathbf{e}) = \mathbf{0}, \ \mathbf{E}(\mathbf{e}\mathbf{e}') = \mathbf{V_e} \tag{3.5}$$

$$\mathbf{a} = \mathbf{J}\theta + \varepsilon, \ \mathbf{E}(\varepsilon) = \mathbf{0}, \ \mathbf{E}(\varepsilon\varepsilon') = \mathbf{V}_\varepsilon, \ \mathbf{E}(\mathbf{e}\varepsilon') = \mathbf{0} \tag{3.6}$$

Where

$\mathbf{S} = [\mathbf{S_1}, \ldots, \mathbf{S_T}]', \ \theta = [\theta_\mathbf{1}, \ldots, \theta_\mathbf{T}]', \ \mathbf{a} = [\mathbf{a_1}, \ldots, \mathbf{a_M}]'$
β = the quantities of the bias, it measures the inconsistencies that exist between **a** and the relating entities **s**

Matrix **R** of dimension T by H contains the H regressors (the X's or the sums of the related series) of eqn. (3.23) and $\beta = [\beta_1, \ldots, \beta_H]$ the regression parameters. The covariance matrix has this:

$$V_e = \Xi^\lambda \Omega \Xi^\lambda \tag{3.7}$$

Where V_e is the covariance matrix of the survey error, Ξ is a diagonal matrix of standard deviations, $\sigma_1, \ldots, \sigma_T$ and $\mathbf{\Omega}$ contains the autocorrelations usually following an ARMA error model. For an AR(1) error model with parameter $|\varphi| < 1$, the elements of Ω are given by $\omega_{ij} = \varphi^{|i-j|}$

$$\omega_{ij} = \varphi^{|i-j|} = \Omega = \begin{bmatrix} 1 & \phi & \phi^2 & . & . & . & \phi^{T-1} \\ \phi & 1 & \phi & . & . & . & \phi^{T-2} \\ \phi^2 & \phi & 1 & . & . & . & \phi^{T-3} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \phi^{T-1} & \phi^{T-2} & \phi^{T-3} & . & . & . & 1 \end{bmatrix}$$

The standard deviations are assumed to be 1. This is essential in light of the unreasonable impact bigger qualities may have on exceptions in the series, and accordingly present issue for benchmarking or compromise. It is therefore important to accept a consistent standard deviation (Hillmer and Trabelsi, 1987)

3.2.1 Temporal summation

Matrix **J** of dimension M x T is used as the matrix used to carry out temporal summation, it contains the proportions of coverage,

$$J = \begin{bmatrix} j_{11} & j_{12} & . & . & . & j_{1T} \\ j_{21} & j_{22} & . & . & . & j_{2T} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ j_{M1} & j_{M2} & . & . & . & j_{MT} \end{bmatrix}$$

Where

$j_{mt} = 0$ for $t < t_{1m}$ or $t > t_{Lm}$, and j_{mt} is as defined for $t_{1m} \le t_{1m} \le t_{Lm}$

Each proportion of coverage j_{mt} is used in *m* and *t* of **J**

Equations (3.5) and (3.6) can be pre-written as

$$y = X\alpha + U \tag{3.8}$$

$$E(U) = 0, \quad E(UU') = V_u = block(V_{e,}\ V_\varepsilon),$$

Where

$$y = \begin{bmatrix} s \\ a \end{bmatrix}, \quad X = \begin{bmatrix} R & I_T \\ 0 & J \end{bmatrix}, \quad \alpha = \begin{bmatrix} \beta \\ \theta \end{bmatrix}, \quad V_u = \begin{bmatrix} V_e & 0 \\ 0 & V_\varepsilon \end{bmatrix}$$

where

y, **X**, a, **u**, $\mathbf{I_T}$, $\mathbf{V_u}$ are indicators, regressors, parameters, errors, identity matrix of dimension T x T, and covariance matrix for sub-annual and annual series respectively.

3.2.2 Theoretical Framework

The models (3.5) and (3.6) can be written as

$$\begin{bmatrix} s \\ a \end{bmatrix} = \begin{bmatrix} R & I_T \\ 0 & J \end{bmatrix} \times \begin{bmatrix} \beta \\ \theta \end{bmatrix} + \begin{bmatrix} e \\ \varepsilon \end{bmatrix} \tag{3.9}$$

Model (3.8) follows a standard regression model, having the solution of generalized least square:

$$\hat{\alpha} = \left(X' V_U^{-1} X\right)^{-1} X' V_U^{-1} y = \begin{bmatrix} \hat{\beta} \\ \hat{\theta} \end{bmatrix} \tag{3.10}$$

V_U is the known or true matrix of the covariance of the disturbances **u**, the estimates $\hat{\alpha}$ gives

$$\operatorname{var}(\hat{\alpha}) = \left(X' V_U^{-1} X\right)^{-1} = \begin{bmatrix} \operatorname{var}(\hat{\beta}) & \operatorname{cov}(\hat{\beta}\,\hat{\theta}) \\ \operatorname{cov}(\hat{\theta}\,\hat{\beta}) & \operatorname{var}(\hat{\theta}) \end{bmatrix} \tag{3.11}$$

3.2.3 Matrix representation of the model

$\hat{a} = [\hat{\beta}\ \hat{\theta}]$ produces the estimates of $\hat{\beta}_1, \ldots, \hat{\beta}_H\ \hat{\theta}_1, \ldots, \hat{\theta}_T$;

$var[\hat{a}]$ produces the covariance matrix of the vector $\hat{\beta}$;

$var[\hat{a}]$, gives the covariance matrix of the vector $\hat{\theta}$.

The solution (3.10) is obtained only if Vu is not a singular matrix, this means both Ve and $V\varepsilon$ matrices having feasible determinants. The results of last condition is that the values on the located on the diagonal $V\varepsilon$ are expected to higher than zero, that is the values of the benchmarks should not be binding.

Given eqn. (3.9), model (3.10) can be written as:

$$\begin{bmatrix}\hat{\beta}\\ \hat{\theta}\end{bmatrix}=\begin{bmatrix}R & I_T\\ 0 & J\end{bmatrix}^{-1}\left(\begin{bmatrix}s\\ a\end{bmatrix}-\begin{bmatrix}e\\ \varepsilon\end{bmatrix}\right) \tag{3.12}$$

$$\begin{bmatrix}\hat{\beta}\\ \hat{\theta}\end{bmatrix}=\begin{bmatrix}R'V_e^{-1}R & R'V_e^{-1}\\ V_e^{-1}R & (V_e^{-1}+J'V_\varepsilon^{-1}J)\end{bmatrix}^{-1}\begin{bmatrix}R'V_e^{-1} & 0\\ V_e^{-1} & J'V_\varepsilon^{-1}\end{bmatrix}\begin{bmatrix}s\\ a\end{bmatrix} \tag{3.13a}$$

$$\begin{bmatrix}\hat{\beta}\\ \hat{\theta}\end{bmatrix}=\begin{bmatrix}R'V_e^{-1}R & R'V_e^{-1}\\ V_e^{-1}R & (V_e^{-1}+J'V_\varepsilon^{-1}J)\end{bmatrix}^{-1}\begin{bmatrix}R'V_e^{-1}s\\ V_e^{-1}s+J'V_\varepsilon^{-1}a\end{bmatrix} \tag{3.13b}$$

Having the following dimensions:

$R = Tx1$, V_e and $V_\varepsilon = TxT$, $J=MxT$

V_ε is the covariance matrix of the benchmarks error

3.2.4 Derivations of the method

From (3.11)

$$\operatorname{var}(\hat{\alpha})=\begin{bmatrix}R'V_e^{-1}R & R'V_e^{-1}\\ V_e^{-1}R & (V_e^{-1}+J'V_\varepsilon^{-1}J)\end{bmatrix}^{-1}$$

$$=\begin{bmatrix}E_{11} & E_{12}\\ E'_{12} & E_{22}\end{bmatrix}^{-1}=\begin{bmatrix}E^{11} & E^{12}\\ E^{12'} & E^{22}\end{bmatrix} \tag{3.14}$$

Combining (3.12) and (3.13)

$$\hat{\beta}=\operatorname{var}\left[\hat{\beta}\right]R'V_e^{-1}s+\operatorname{cov}\left[\hat{\beta},\hat{\theta}\right]\left(V_e^{-1}s+J'V_\varepsilon^{-1}a\right) \tag{3.15}$$

$$\hat{\theta}=\operatorname{cov}\left[\hat{\theta},\hat{\beta}\right]R'V_e^{-1}s+\operatorname{var}\left[\hat{\theta}\right]\left(V_e^{-1}s+J'V_\varepsilon^{-1}a\right) \tag{3.16}$$

From eqn. (3.10) V_ε should be positive, this means that having binding benchmarks is not feasible. The quantity in (3.10) is now expressed as:

$$\hat{\beta}=-\left(R'J'V_d^{-1}JR\right)^{-1}R'J'V_d^{-1}\left[a-Js\right] \tag{3.17}$$

$$\operatorname{var}\left[\hat{\beta}\right]=\left(R'J'V_d^{-1}JR\right)^{-1} \tag{3.18}$$

Where $\mathbf{s}$= indicator series and $\mathbf{V_d} = \mathbf{J}^{\mathbf{J}}\mathbf{V_e J} + \mathbf{V}_\varepsilon$
derivation assumes that both V_e and V_ε are invertible. When V_ε is not non-singular, the expression has to change to V_δ giving $V_\varepsilon + \delta I_M$ and δ which is not far from zero. The quantity δ should equal zero upon derivation.

The bottom-right segment of (3.14) which is the same as E_{22} in (3.15) using the matrix identities be transformed and expansion of the algebra:

$$(D+BCB')^{-1} \equiv D^{-1} - D^{-1}B\left(B'D^{-1}B + C^{-1}\right)^{-1} B'D^{-1}$$

$$\left(V_e^{-1} + J'V_e^{-1}J\right)^{-1} \equiv V_e J'\left(J'V_e^{-1}J + V_e\right)^{-1} JV_e$$

$$\equiv V_e - V_e J'V_d^{-1} JV_e = E_{22}^{-1} \tag{3.19}$$

Given E_{22}^{-1} from (3.19), the matrix inversion in (3.14) yields this after transformation:

Proof from R.H.S.

$$= D^{-1} - D^{-1}B\,B'D^{-1}(B'D^{-1}B + C^{-1})^{-1}$$

$$= D^{-1}\left(1 - D^{-1}B\,B'\left(\frac{1}{B'D^{-1}B + C^{-1}}\right)\right)$$

$$= D^{-1}\left(1 - \frac{D^{-1}B\,B'}{B'D^{-1}B + C^{-1}}\right) = D^{-1}\left(\frac{B'D^{-1}B + C^{-1} - D^{-1}B\,B'}{B'D^{-1}B + C^{-1}}\right)$$

$$= \frac{D^{-1}C^{-1}}{B'D^{-1}B + C^{-1}} = \frac{(DC)^{-1}}{B'D^{-1}B + C^{-1}}$$

$$= \frac{\frac{1}{DC}}{B'D^{-1}B + C^{-1}} = \left(\frac{1}{DC}\right)\left(\frac{1}{B'D^{-1}B + C^{-1}}\right) \tag{3.20}$$

since $= CC^{-1} = I$ and $DD^{-1} = I$

Therefore

$$= \frac{1}{D\,C\,B'D^{-1}B + D\,C\,C^{-1}} = \frac{1}{I\,B\,C\,B' + I\,D}$$

$$= \frac{1}{I\left(B\,C\,B'+D\right)} = \frac{1}{B\,C\,B'+D}$$

$$= \left(B\,C\,B'+D\right)^{-1} \quad \text{or} \quad \left(D + B\,C\,B'\right)^{-1} \tag{3.21}$$

$$\begin{bmatrix} E_{11} & E_{12} \\ E'_{12} & E_{22} \end{bmatrix}^{-1} = \frac{\begin{bmatrix} E_{22} & -E_{12} \\ -E'_{12} & E_{11} \end{bmatrix}}{E_{11}E_{22} - E'_{12}E_{12}} \tag{3.22}$$

$$E^{11} = \frac{E_{22}}{E_{11}E_{22} - E'_{12}E_{12}}$$

$$= \frac{E_{22}}{E_{22}\left(E_{11} - E'_{12}\,E_{12}E_{22}^{-1}\right)}$$

$$= \frac{1}{E_{11} - E'_{12}\,E_{12}E_{22}^{-1}} = \left(E_{11} - E'_{12}\,E_{12}E_{22}^{-1}\right)^{-1} \tag{3.23}$$

Therefore

$$E^{11} = \left(E_{11} - E_{12}'\,E_{12}E_{22}^{-1}\right)^{-1} \tag{3.24}$$

$$E^{12} = \frac{-E_{12}}{E_{11}E_{12} - E_{12}'\,E_{12}} = \frac{-E_{12}}{E_{22}\left(E_{11} - E_{12}'\,E_{12}E_{22}^{-1}\right)}$$

$$= -E_{12}E_{22}^{-1}\left(E_{11} - E_{12}'\,E_{12}E_{22}^{-1}\right)^{-1} \tag{3.25}$$

Therefore

$$E^{12} = -E_{12}E_{22}^{-1}E^{11} \tag{3.26}$$

$$E^{22} = E_{22}^{-1} - E^{12'}E_{12}E_{22}^{-1} = E_{22}^{-1} + E_{22}^{-1}E'_{12}\,E^{11}E_{12}E_{22}^{-1}$$

$$= E_{22}^{-1} + E_{22}^{-1}E^{12'}\left(E_{11} - E_{12}E_{22}^{-1}E'_{12}\right)^{-1}E_{12}E_{22}^{-1} \tag{3.27}$$

Substituting (3.24) in the upper left partition of (3.11) yields:

$$\operatorname{var}\left[\hat{\beta}\right] = E^{11} = \left[\left(R'V_e^{-1}R\right) - \left(R'V_e^{-1}\right)\left(V_e - V_eJ'V_d^{-1}JV_e\right)\left(V_e^{-1}R\right)\right]^{-1}$$

$$= \left(R' J' V_d^{-1} JR\right)^{-1} \tag{3.28}$$

this proves (3.18)

Substituting (3.25) in the upper right partition of (3.11) yields:

$$\operatorname{cov}\left[\hat{\beta}, \hat{\theta}\right] = E^{12} = -\operatorname{var}\left[\hat{\beta}\right]\left(R' V_e^{-1}\right)\left(V_e - V_e J' V_d^{-1} JV_e\right)$$

$$= -\operatorname{var}\left[\hat{\beta}\right] R'\left(I \; - J' V_d^{-1} JV_e\right) \tag{3.29}$$

$$\operatorname{cov}\left[\hat{\theta}, \hat{\beta}\right] = E^{12'} = -\left(I - V_e J' V_d^{-1} J\right) R \operatorname{var}\left[\hat{\beta}\right] \tag{3.30}$$

Substituting (3.27) into the lower right partition of (3.11) yields:

$$\operatorname{var}\left[\hat{\theta}\right] = E^{22} = \left(V_e - V_e J' V_d^{-1} JV_e\right) - \operatorname{cov}\left[\hat{\theta}, \hat{\beta}\right]\left(R' V_e^{-1}\right)\left(V_e - V_e J' V_d^{-1} JV_e\right)$$

$$= \left(V_e - V_e J' V_d^{-1} JV_e\right) + \left(I - V_e J' V_d^{-1} J\right) R \operatorname{var}\left[\hat{\beta}\right] R'\left(I - J' V_d^{-1} JV_e\right) \tag{3.31}$$

Substituting (3.28), (3.29), (3.30) into (3.11) and expanding leads to cancellations and simplifications

$$\hat{\beta} = \operatorname{var}\left[\hat{\beta}\right] R' V_e^{-1} s - \operatorname{var}\left[\hat{\beta}\right] R'\left(I - J' V_d^{-1} JV_e\right)\left(V_e^{-1} s + V_\varepsilon^{-1} a\right)$$

$$= \operatorname{var}\left[\hat{\beta}\right] R' V_d^{-1} Js + \operatorname{var}\left[\hat{\beta}\right] R' J' V_d^{-1}\left(JV_e J' s - V_d\right) V_\varepsilon^{-1} a \tag{3.32}$$

Substituting the content in V_d using the expression in the last $JV_e J' + V_\varepsilon$ gives:

$$\hat{\beta} = \operatorname{var}\left[\hat{\beta}\right] R' J' V_e^{-1} Js + \operatorname{var}\left[\hat{\beta}\right] R' J' V_d^{-1}\left(JV_e J' - JV_e J' - V_\varepsilon\right) V_\varepsilon^{-1} a \tag{3.33}$$

$$\hat{\beta} = -\left(R' J' V_d^{-1} JR\right)^{-1} R' J' V_d^{-1}\left(a - Js\right) \tag{3.34}$$

Substituting (3.28), (3.29), (3.30) into (3.11) and expanding yields:

$$\hat{\theta} = -\left(I - BJ\right) R \operatorname{var}\left[\hat{\beta}\right] R' V_e^{-1} s + \left(\left(V_e - BJV_e\right) + \left(I - BJ\right) R \operatorname{var}\left[\hat{\beta}\right] R'\left(I - J' B'\right)\right)\left(V_e^{-1} s + J' V_\varepsilon a\right) \tag{3.35}$$

Where $B = V_e J' V_d^{-1}$

The following model was obtained after a long mathematical computations have been performed.

Some terms in $V_\varepsilon a$ and s were transformed and replaced:

$$\hat{\theta} = s - V_e J' V_d^{-1} Js + V_e J' V_d^{-1} JR \operatorname{var}\left[\hat{\beta}\right] R' J' V_d^{-1} Js - R \operatorname{var}\left[\hat{\beta}\right] R' J' V_d^{-1} Js$$
$$- V_e J' \left[V_d^{-1} JV_e J' - I\right] V_\varepsilon^{-1} a - \left[R \operatorname{var}\left[\hat{\beta}\right] R' J'\right]\left[V_d^{-1} JV_e J' - I\right] V_\varepsilon^{-1} a$$
$$+ \left[V_e J' V_d^{-1} JR \operatorname{var}\left[\hat{\beta}\right] R' J'\right]\left[V_d^{-1} JV_e J' - I\right] V_\varepsilon^{-1} a \tag{3.36}$$

This model derived is called Autocorrelated Benchmarking model (AIBM)

The above models can be implemented using three mathematical softwares for modeling; MATLAB and MAPLE 18. Computer programs are also developed for Denton's and Cholette-Dagum methods of benchmarking, the former is numerical while the latter is model-based.

3.2.5 Major contributions to the existing solutions to benchmarking problems

i. Deriving an alternative matrix solution to benchmarking problem

ii. Obtain a new benchmarked series that reproduces the movement in the original series

iii. Developing a computer program for solving benchmarking problem using Den- ton and Cholette-Dagum methods

iv. Modification of the Denton's method to remove it's short-coming of: (a) Implicit forecast of the next of the next discrepancy at the end of the series, on the basis of the last two discrepancies only (b) Introduction of a transient movement at the beginning of the series, which defeats the stated principle of movement preservation.

3.3 Growth Rates

These are employed in order to measure how close the benchmarked series obtained from various techniques θ to the high frequency series s: Growth rates can be presented in a bar chart or in an analytical table. They are useful to evaluate the movement preservation. Growth rates of the indicator series to the growth of benchmarked series from various models can be compared.
If the growth rates are similar, it shows that the benchmarking method ensures good movement

preservation from the indicator series to the benchmarked series. It is given as:

$$\text{Growth rate} = \frac{V_{present} - V_{past}}{V_{past}} \text{ x } 100$$

Where

$V_{present}$ = present value

V_{past} = past value

It is noteworthy that values obtained for growth rates in this study were not converted to percentages. This is necessary in order to be able to compute other necessary statistics from them.

3.3.1 Importance of the growth rates to benchmarking

1. Growth rates are used in measuring the rate of discrepancies between the indicator series and benchmarked series
2. They are preferred in percentages for ease of interpretation (Latendresse et al, 2007)
3. They are useful in evaluating the movement preservation
4. Better benchmarking method can be determined through it

3.4 Coefficient of Variation

The coefficient of variation, also known as relative standard deviation, is the ratio of the standard deviation σ to the mean μ. It is given by

$$CV = \frac{\sigma}{\mu}$$

It reveals the degree of variability in relation to the mean of the sample or population. The CV is especially useful when the researcher wants to compare outcomes from two different surveys or tests having different measures or values. Distributions with CV < 1 are considered low-variance, while those with CV > 1 are considered high-variance (Broverman, 2001)

CHAPTER FOUR

DATA ANALYSIS AND RESULTS

4.1 Introduction

In this chapter, we report the results of benchmarking various datasets using the alternative method of benchmarking. Also comparison of other methods of benchmarking with newly developed one is made, performance tests using various statistics were all carried out.

4.2 Application of the alternative method of benchmarking to various datasets

There are various benchmarking methods as described in chapter two of this thesis, the gaps left in many of them is what our alternative method filled, that is the Autocorrelated Integrated Benchmarking Model (AIBM). The model is applied to many data sets in this chapter:

$$\hat{\theta} = s - V_e J' V_d^{-1} Js + V_e J' V_d^{-1} JR \operatorname{var}\left[\hat{\beta}\right] R' J' V_d^{-1} Js - R \operatorname{var}\left[\hat{\beta}\right] R' J' V_d^{-1} Js$$
$$- V_e J' \left[V_d^{-1} JV_e J' - I\right] V_\varepsilon^{-1} a - \left[R \operatorname{var}\left[\hat{\beta}\right] R' J'\right]\left[V_d^{-1} JV_e J' - I\right] V_\varepsilon^{-1} a$$
$$+ \left[V_e J' V_d^{-1} JR \operatorname{var}\left[\hat{\beta}\right] R' J'\right]\left[V_d^{-1} JV_e J' - I\right] V_\varepsilon^{-1} a \qquad (4.1)$$

Where

$\hat{\theta}$ T by 1 matrix for the generated benchmarked series estimates

a M by 1 matrix for annual series

s T by 1 matrix for the sub-annual series

V_e T by T covariance matrix for the sub-annual series

J M by T matrix for the temporal sum operator

V_d M by M covariance matrix for the annual discrepancies

R T by 1 matrix for the regressors

$\hat{\beta}$ Estimate of the bias constant

V_e M by M covariance matrix for the sub-annual discrepancies

I T by T identity matrix of ones

The MATLAB code for the full implementation of this model is in appendix II of this write up.

4.3 Implementation of the model using the Cholette data, Cholette and Dagum (2006)

In order to apply the derived model, the simulated quarterly series by Cholette and Dagum (2006) are used.

$$V_d = \begin{bmatrix}
\frac{6347038342708793}{562949953241312} & \frac{35966473322019}{70368744177664} & \cdot & \cdot & \cdot & \frac{2343356126012245}{72057594073927936} & \frac{5294666344162943}{576460752033423488} \\
\frac{5754635735152305}{1125899908642624} & \frac{6347038342708793}{562949953241312} & \cdot & \cdot & \cdot & \frac{8297134930264777}{72057594073927936} & \frac{2343356126012245}{72057594073927936} \\
\frac{3250558196827209}{2251799816385248} & \frac{5754635735152305}{1125899908642624} & \cdot & \cdot & \cdot & \frac{3672214587129203}{9007199257440992} & \frac{8297134930264777}{72057594073927936} \\
\frac{3672214587129203}{9007199257440992} & \frac{3250558196827209}{2251799816385248} & \cdot & \cdot & \cdot & \frac{6501116392754417}{4503599623770496} & \frac{3672214587129203}{9007199257440992} \\
\frac{4148567460632389}{36028797018963968} & \frac{3672214587129203}{9007199257440992} & \cdot & \cdot & \cdot & \frac{35966473322019}{70368744177664} & \frac{6501116392754417}{4503599623770496} \\
\frac{2343356126012245}{72057594073927936} & \frac{4148567460632389}{36028797018963968} & \cdot & \cdot & \cdot & \frac{6347038342708793}{562949953241312} & \frac{35966473322019}{70368744177664} \\
\frac{2647333177031471}{288230376511711744} & \frac{2343356126012245}{72057594073927936} & \cdot & \cdot & \cdot & \frac{5754635735152305}{1125899908642624} & \frac{371040489447791}{35184372088832}
\end{bmatrix}$$

Where $V_d = J \times Ve \times J'$ is the Covariance matrix for annual discrepancies of 7 by 7 dimension

Matrix V_d reveals the covariance matrix for annual discrepancies. It shows the covariances for the aggregated seven years data. The numerator and denominator of each fraction represents total benchmarks and the sum of the annual series respectively. In a binding series the outcome of each of these fractions should be approximately equals 1 and less or greater than 1 in the case of unbinding series. Irregular benchmarks come up when the economic data is unbinding in many cases.

$$V_d* = \begin{pmatrix} 11.2746 & 5.1111 & 1.4435 & 0.4077 & 0.1151 & 0.0325 & 0.0092 \\ 5.1111 & 11.2746 & 5.1111 & 1.4435 & 0.4077 & 0.1151 & 0.0325 \\ 1.4435 & 5.1111 & 11.2746 & 5.1111 & 1.4435 & 0.4077 & 0.1151 \\ 0.4077 & 1.4435 & 5.1111 & 11.2746 & 5.1111 & 1.4435 & 0.4077 \\ 0.1151 & 0.4077 & 1.4435 & 5.1111 & 11.2746 & 5.1111 & 1.4435 \\ 0.0325 & 0.1151 & 0.4077 & 1.4435 & 5.1111 & 11.2746 & 5.1111 \\ 0.0092 & 0.0325 & 0.1151 & 0.4077 & 1.4435 & 5.1111 & 10.5456 \end{pmatrix}$$

Where

V_d* is the computed covariance matrix for the annual discrepancies of 7 by 7 dimension

The covariance matrix for annual discrepancies presented V_{d*} depicts the computed values earlier presented in V_d. The first value 11.2746 shows that the discrepancies between the first year benchmark (annual) and the sub-annual series for the same year is 11.2746. The discrepancy for the second year is 5.1111, the discrepancy for the third is 1.4435, while the discrepancy for the seventh year is 0.0092. None of the years is binding in terms of discrepancies, the benchmarks are irregular and be evaluated and reconciled for suitability in usage for future research and application.

In a binding benchmarked series the discrepancies are not more than one, although this is not common in most economic data, the annual total and the sub-annual total are always contrary to each other, the commonest series are non-binding (irregular).

The matrix is a confirmation that the series is irregular in nature, therefore there is need for evaluation and reconciliation in order to obtain an hybrid and improved series.

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Where

B is the matrix of the bias

The matrix of the Bias is represented as *B*, the bias is very important in reconciling low frequency economic series with high frequency series. The bias is a regressor or predictor that is deterministic in nature. It is a constant that covers the difference in average level that exist between the low frequency series and the high frequency series. It normally occurs when the high series is not fair with the low series. This happens mostly when there is problem of under coverage or non-coverage in a survey. Bias is an expected phenomenon is benchmarking.

The column matrices of ones indicate the quarters in the years, we have four of such in a year, the one represented on page 62 contains 28 quarters in seven years. This is inserted in the equation 4.1 before a new benchmarked series could be obtained.

$$
V_e = \begin{pmatrix}
1.0000 & 0.7290 & 0.5314 & . & . & . & 0.0003 & 0.0002 \\
0.7290 & 1.0000 & 0.7290 & . & . & . & 0.0004 & 0.0003 \\
0.5314 & 0.7290 & 1.0000 & . & . & . & 0.0005 & 0.0004 \\
0.3874 & 0.5314 & 0.7290 & . & . & . & 0.0007 & 0.0005 \\
0.2824 & 0.3874 & 0.5314 & . & . & . & 0.0010 & 0.0007 \\
0.2059 & 0.2824 & 0.3874 & . & . & . & 0.0013 & 0.0010 \\
0.1501 & 0.2059 & 0.2824 & . & . & . & 0.0018 & 0.0013 \\
0.1094 & 0.1501 & 0.2059 & . & . & . & 0.0025 & 0.0018 \\
0.0798 & 0.1094 & 0.1501 & . & . & . & 0.0034 & 0.0025 \\
0.0581 & 0.0798 & 0.1094 & . & . & . & 0.0046 & 0.0034 \\
0.0424 & 0.0581 & 0.0798 & . & . & . & 0.0064 & 0.0046 \\
0.0309 & 0.0424 & 0.0581 & . & . & . & 0.0087 & 0.0064 \\
0.0225 & 0.0309 & 0.0424 & . & . & . & 0.0120 & 0.0087 \\
0.0164 & 0.0225 & 0.0309 & . & . & . & 0.0164 & 0.0120 \\
0.0120 & 0.0164 & 0.0225 & . & . & . & 0.0225 & 0.0164 \\
0.0087 & 0.0120 & 0.0164 & . & . & . & 0.0309 & 0.0225 \\
. & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . \\
0.0005 & 0.0007 & 0.0010 & . & . & . & 0.5314 & 0.3874 \\
0.0004 & 0.0005 & 0.0007 & . & . & . & 0.7290 & 0.5314 \\
0.0003 & 0.0004 & 0.0005 & . & . & . & 1.0000 & 0.0000 \\
0.0002 & 0.0003 & 0.0004 & . & . & . & 0.7290 & 1.0000
\end{pmatrix}
$$

Where

V_e is the covariance matrix for sub-annual discrepancies of 28 by 28 dimension

The matrix on page 64 shows the covariance matrix for sub-annual discrepancies. Each of the values in the cell represents the discrepancies in the sub-annual (high frequency series). For instance, 0.7200 is the discrepancy between the annual and sub-annual series for the first year in the series. In this case, we have seven annual series, the sub-annual is twenty-eight, and therefore we are expected to have a 28 by 28 matrix. It can be seen from the matrix that only few values are close to 1, which means they are not binding. This shows that there exists irregular benchmarks in the series, as a result, there is need for evaluation.

The evaluation was carried out and the results are presented in tables 4.1a and 4.1b on page 64. The outcome of the evaluation is expected to be an improved series, better than the irregular series. For any economic data to be reliable for usage and forecast, the B-I ratio obtained from it's annual and sub-annual totals must be close to 1.

Table 4. 1a: Indicator series, bias-adj, benchmarked, annual, and growth rates

Quarters	Ind.	bias-adj	S*	Annual	Ind.(%)	AIBM(%)
1998-1	85.00	136.75	134.81			
1998-2	95.00	146.75	144.10		0.11	0.06
1998-3	125.00	176.75	173.12		0.24	0.17
1998-4	95.00	146.75	141.77	494.00	-0.32	-0.22
1999-1	85.00	136.75	129.93		-0.12	-0.09
1999-2	95.00	146.75	137.36		0.11	0.05
1999-3	125.00	176.75	163.80		0.24	0.16
1999-4	95.00	146.75	128.90	560.00	-0.32	-0.27
2000-1	85.00	136.75	112.16		-0.12	-0.15
2000-2	95.00	146.75	120.63		0.11	0.07
2000-3	125.00	176.75	154.14		0.24	0.22
2000-4	95.00	146.75	133.07	520.00	-0.32	-0.16
2001-1	85.00	136.75	138.30		-0.12	0.04
2001-2	95.00	146.75	156.70		0.11	0.12
2001-3	125.00	176.75	189.11		0.24	0.17
2001-4	95.00	146.75	155.78	640.00	-0.32	-0.21
2002-1	85.00	136.75	136.38		-0.12	-0.14
2002-2	95.00	146.75	142.21		0.11	0.04
2002-3	125.00	176.75	172.88		0.24	0.18
2002-4	95.00	146.75	148.43	600.00	-0.32	-0.16
2003-1	85.00	136.75	149.44		-0.12	0.01

Table 4.1b: Indicator series, bias-adj, benchmarked, annual, and growth rates (contd).

Quarters	**Ind.**	**bias-adj**	**S***	**Annual**	**Ind. (%)**	**AIBM(%)**
2003-2	95.00	146.75	165.79		0.11	0.10
2003-3	125.00	176.75	198.12		0.24	0.16
2003-4	95.00	146.75	166.66	680.00	-0.32	-0.19
2004-1	85.00	136.75	151.28		-0.12	-0.10
2004-2	95.00	146.75	157.35		0.11	0.04
2004-3	125.00	176.75	184.48		0.24	0.15
2004-4	95.00	146.75	152.39	661.00	-0.32	-0.21

Table 4.1a and 4.1b present the Cholette simulated series. Column 1 shows the quarters (1981 quarter 1 to 2004 quarter 4). Column 2 the indicator series, column 3 the annual series, column 4 the bias-adjusted series, column 5 the benchmarked series, column 6 and 7 show the growth rates from the indicator series growth rate from benchmarked series using the improved benchmarking method called the autocorrelated indicator benchmarking model (AIBM). Growth rates are used in measuring the rate of discrepancies between the indicator series and the benchmarked series. It can be observed from the table that rate of growth in the column 7 is lower than the ones observed in column 6. Since the percentage of growth rate is reduced in AIBM it shows that the movement in the original series is better preserved in all the twenty-eight quarters, no higher value of growth rate is observed in the new method.

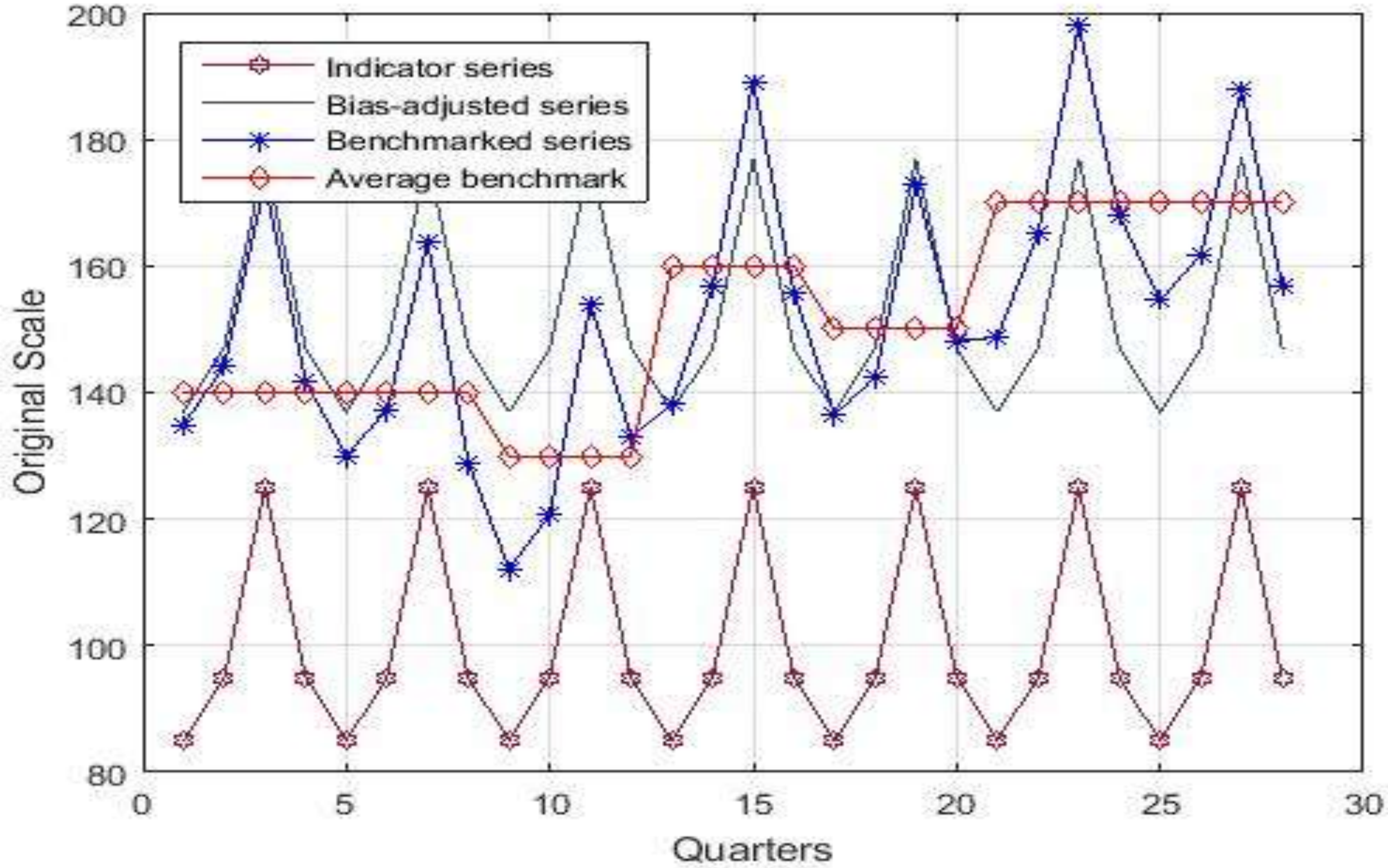

Figure 4. 1: The benchmarked series using Cholette simulated data

According to the chart in fig. 4.1, none of the methods reconciled well with the indicator series. They are far away from the indicator series considering the generated data, movement and pattern. The bias-adjusted series, average benchmarks, and the benchmarked series are at variance with the indicator series. This may be due to the nature of the data, therefore the methods may not be appropriate for Cholette simulated data.

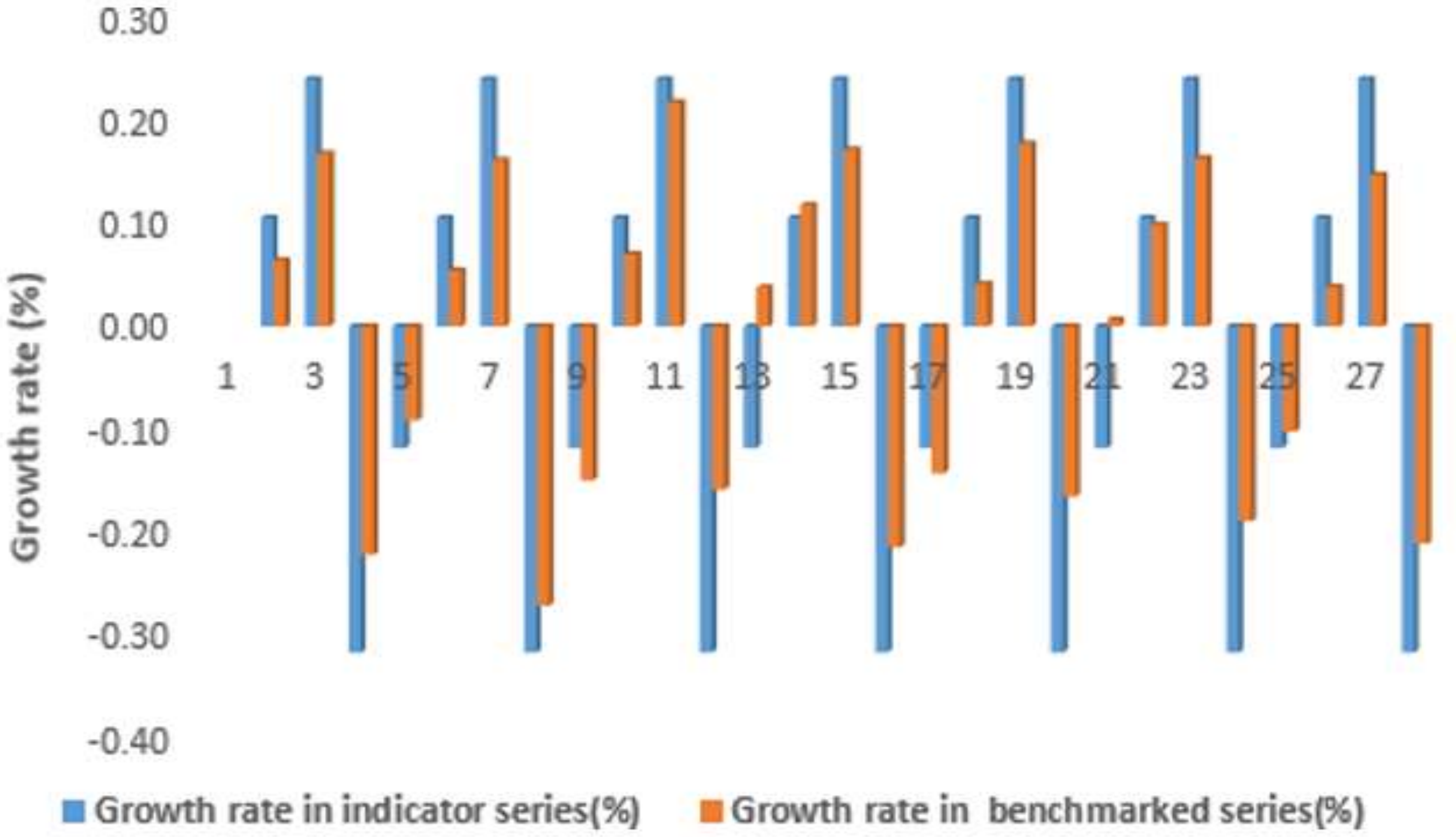

Figure 4. 2: Growth rates graph using Cholette simulated data

As shown in fig. 4.2, the discrepancies between the growth rates from indicator and benchmarked series is wide. The growth are close to one another in few pairs of sample points such as quarters 14, 17, 22, and 25. Other 24 pairs of points are not showing favourable growth rates when compared with the indicator.

The implication of this is that the series is irregular and that the AIBM is not appropriate for evaluation of simulated data. For a regular series, there should be similarities in the growth rates of most of the sample points. The obvious difference in the bars show that the discrepancies between the indicator series and the benchmarked series is too high.

4.4 Using the Denton data

Table 4. 2: Indicator series, bias-adj, benchmarked, annual, and growth rates

Quarters	Ind.	Bias-adj.	S*	Annual	Ind.(%)	AIBM(%)
1998-1	50	59.48	76.13			
1998-2	100	109.48	127.45		0.5000	0.4027
1998-3	150	159.48	175.70		0.3333	0.2746
1998-4	100	109.48	120.72	500	-0.5000	-0.4555
1999-1	50	59.48	61.99		-1.0000	-0.9473
1999-2	100	109.48	104.10		0.5000	0.4045
1999-3	150	159.48	146.25		0.3333	0.2882
1999-4	100	109.48	87.65	400	-0.5000	-0.6686
2000-1	50	59.48	27.43		-1.0000	-2.1952
2000-2	100	109.48	72.57		0.5000	0.6220
2000-3	150	159.48	122.57		0.3333	0.4079
2000-4	100	109.48	77.43	300	-0.5000	-0.5829
2001-1	50	59.48	37.65		-1.0000	-1.0566
2001-2	100	109.48	96.25		0.5000	0.6088
2001-3	150	159.48	154.10		0.3333	0.3754
2001-4	100	109.48	111.99	400	-0.5000	-0.3760
2002-1	50	59.48	70.72		-1.0000	-0.5837
2002-2	100	109.48	125.70		0.5000	0.4374
2002-3	150	159.48	177.45		0.3333	0.2916
2002-4	100	109.48	126.13	500	-0.5000	-0.4069

Table 4.2 presents the Denton data. The seven columns represent the quarters, indicator series, annual series, bias adjusted series, benchmarked and the growth rates respectively. Checking the growth rates critically, most of the quarters between 1998-1 and 2002-4 have favourable values, except in quarters 2001-1 and 2002-1 which account for just 2 out of 28 pairs of points. This shows that the method maintains the pattern in the original data, therefore the method is recommended for use considering Denton data.

For the observations in 1998, the growth rates look similar. This indicates that AIBM model can predict well for the data though it is also a simulated series. Since almost all the observations in the series have similar growth rates with the indicators, it means that model is good.

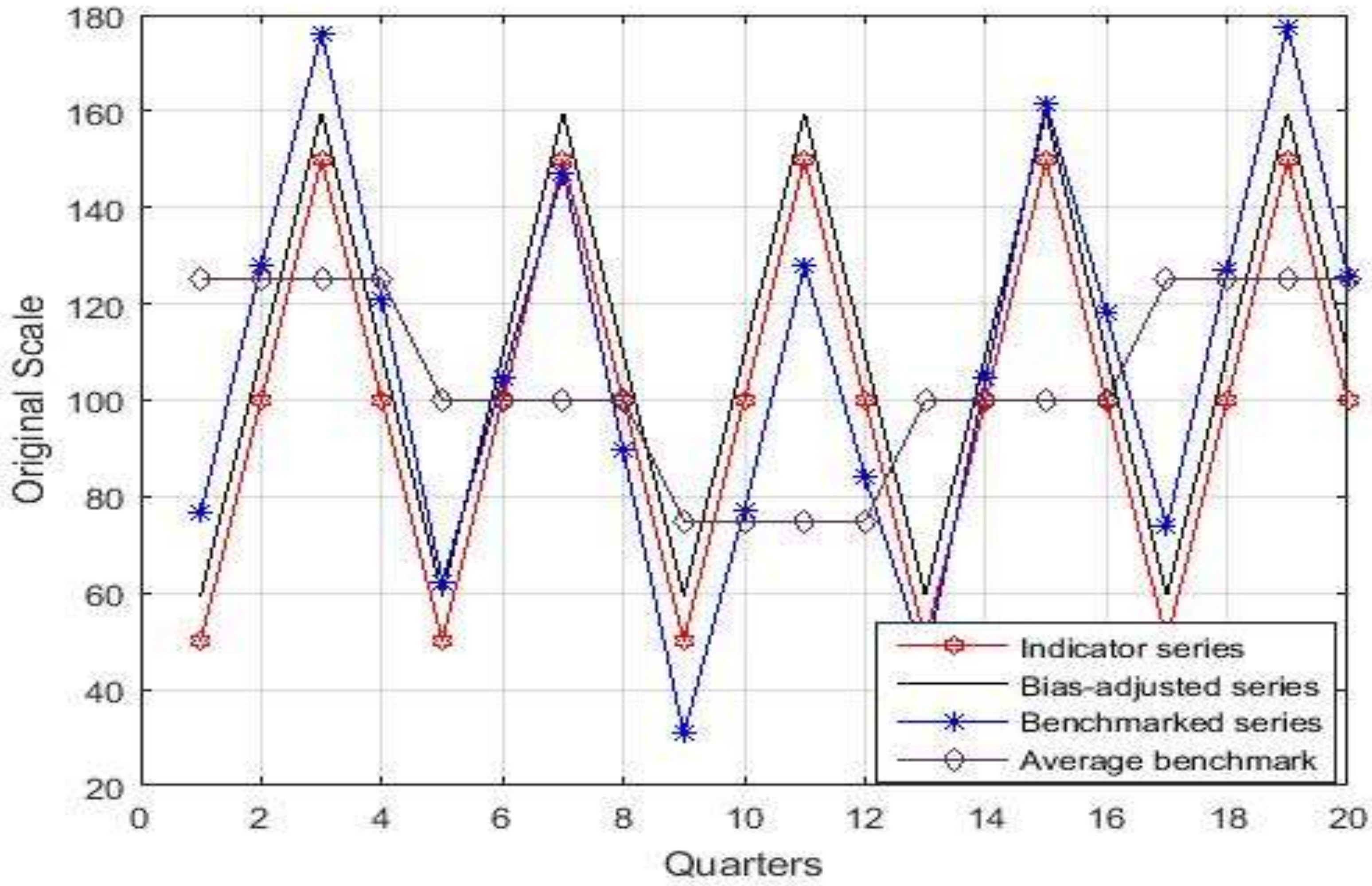

Figure 4. 3: The benchmarked series using Denton simulated data

The chart in fig. 4.3 diagrammatically shows the information in table 4.2, it can be seen that the benchmarked series using the new method (AIBM) compares favourably well when placed side by side with the original series. It is also observed that the patterns and movements in the series generated by the new method and the indicator series are almost the same.

It can be observed that the biased-adjusted and the benchmarked series are similar in growth rates, they compare with the original series (indicator). The only non-conforming series is the average benchmarks series. This an indication that the model is good for reconciliation of the data.

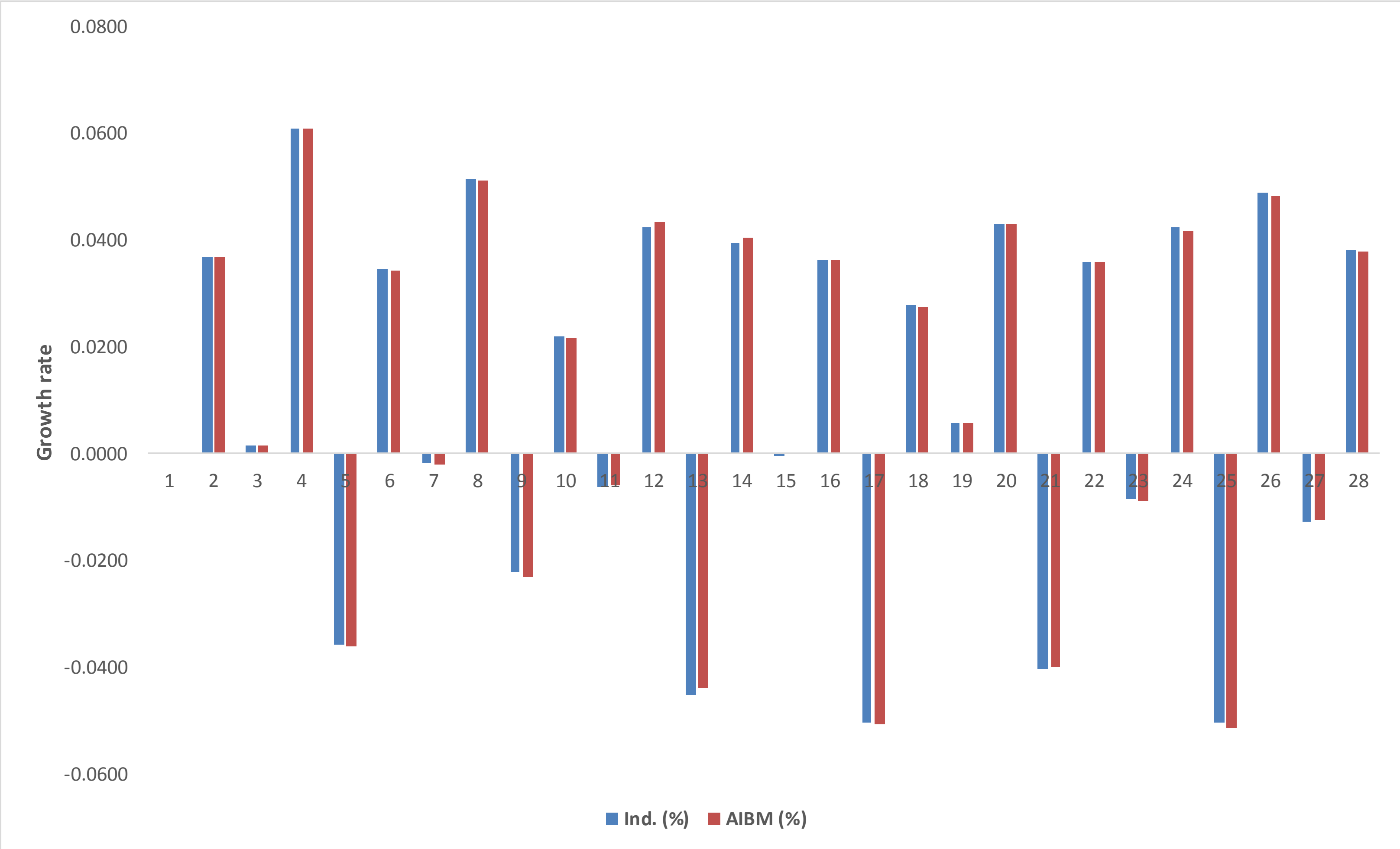

Figure 4. 4: Growth Rates graph using Denton simulated data

Fig. 4.4 represents the growth rates in the indicator and the generated series using the AIBM method. The equality in the lengths of pairs of bars shown in the chart depicts equal or almost equal growth rates percentages The Denton simulated data appear to be good for the method since there is no obvious difference in the growth rate. This means that the pattern and movements in the original series are maintained using the improved model.

4.5 Using the IMF data

Table 4. 3: Indicator series, bias-adj, benchmarked, annual, and growth rates

Quarters	Ind.	Bias.adj.	S*	Annual	Ind. (%)	AIBM (%)
1998-1	613216	613155.06	613408			
1998-2	636852	636791.06	637052		0.0371	0.0371
1998-3	637890	637829.06	638037		0.0016	0.0015
1998-4	679437	679376.06	679467	2567964	0.0611	0.0610
1999-1	656030	655969.06	655866		-0.0357	-0.0360
1999-2	679720	679659.06	679309		0.0349	0.0345
1999-3	678584	678523.06	677848		-0.0017	-0.0022
1999-4	715545	715484.06	714374	2727397	0.0517	0.0511
2000-1	699993	699932.06	698232		-0.0222	-0.0231
2000-2	715684	715623.06	713742		0.0219	0.0217
2000-3	711180	711119.06	709448		-0.0063	-0.0061
2000-4	742734	742673.06	741623	2863045	0.0425	0.0434
2001-1	710551	710490.06	710534		-0.0453	-0.0438
2001-2	739767	739706.06	740451		0.0395	0.0404
2001-3	739391	739330.06	740453		-0.0005	0.0000
2001-4	767218	767157.06	768372	2959810	0.0363	0.0363
2002-1	730367	730306.06	731337		-0.0505	-0.0506
2002-2	751285	751224.06	752172		0.0278	0.0277
2002-3	755601	755540.06	756497		0.0057	0.0057
2002-4	789540	789479.06	790538	3030545	0.0430	0.0431
2003-1	758885	758824.06	760088		-0.0404	-0.0401
2003-2	787249	787188.06	788449		0.0360	0.0360
2003-3	780660	780599.06	781648		-0.0084	-0.0087
2003-4	815231	815170.06	815776	3145961	0.0424	0.0418
2004-1	776189	776128.06	776017		-0.0503	-0.0512
2004-2	816059	815998.06	815445		0.0489	0.0484
2004-3	805903	805842.06	805403		-0.0126	-0.0125
2004-4	837922	837861.06	837095	3233960	0.0382	0.0379

Table 4.3 represents the IMF data. The seven columns represent the quarters, indicator series, annual series, bias adjusted series, benchmarked and the growth rates respectively. It can be seen here that the growth rates in the indicator and the benchmarked series obtained from the improved method are approximately the same. This shows that there is no loss of information in the generated series when compared with the original series. The movement and patterns in the indicator series are preserved

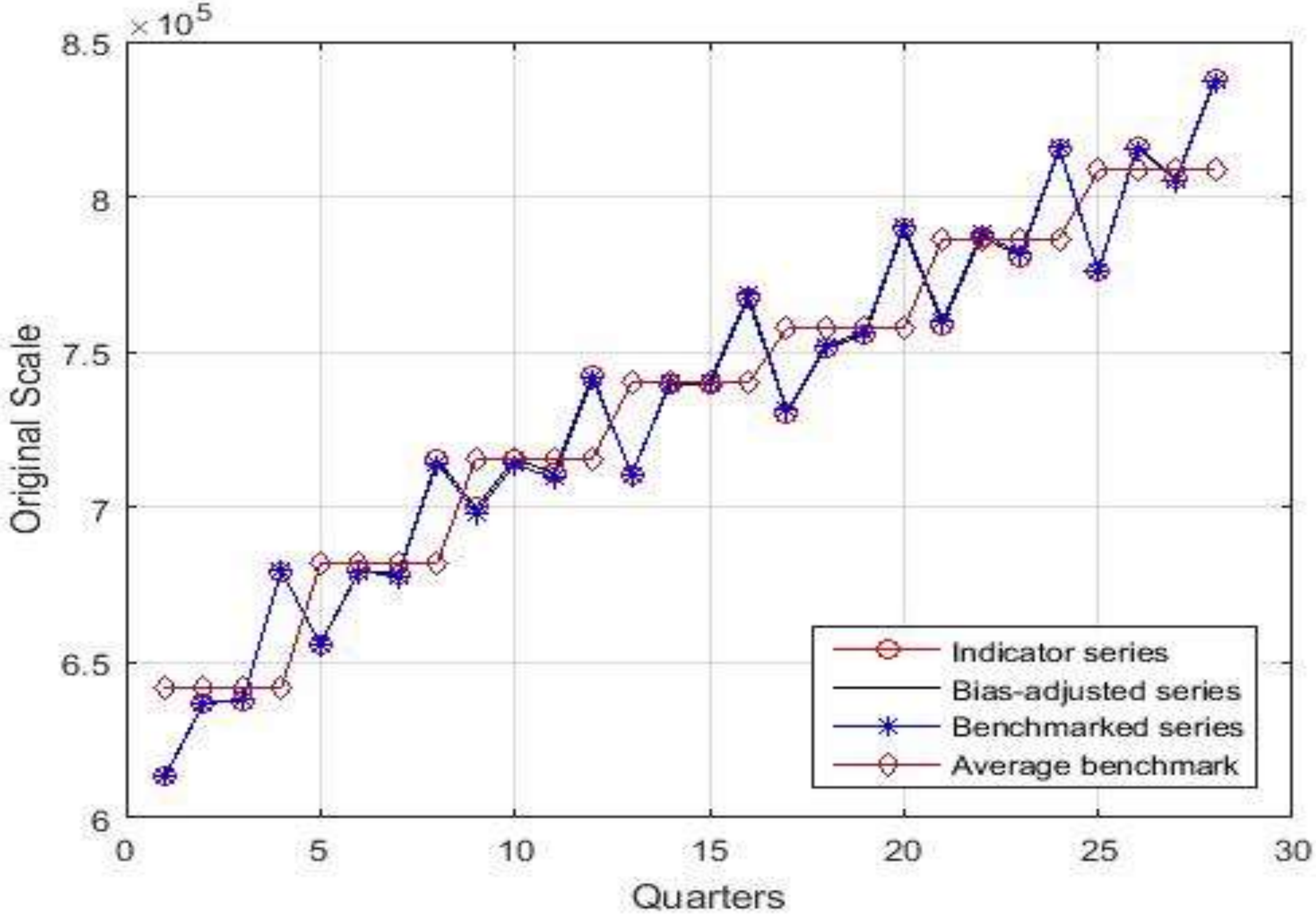

Figure 4. 5: The benchmarked series using the IMF data

The chart in figure 4.5 is a representation of the information in table 4.9. It can be seen that the indicator series, bias adjusted series, the benchmarked series, and the average benchmark (obtained by dividing the benchmarks by 4) follow almost the pattern. It is noteworthy that the indicator and the generated benchmarked series are exactly on the same line, points, and direction, there is no distinction, and one is embedded in the other. This indicates a strong graphic similarity that exists between the series. This shows that the method of benchmarking can be recommended for real life data.

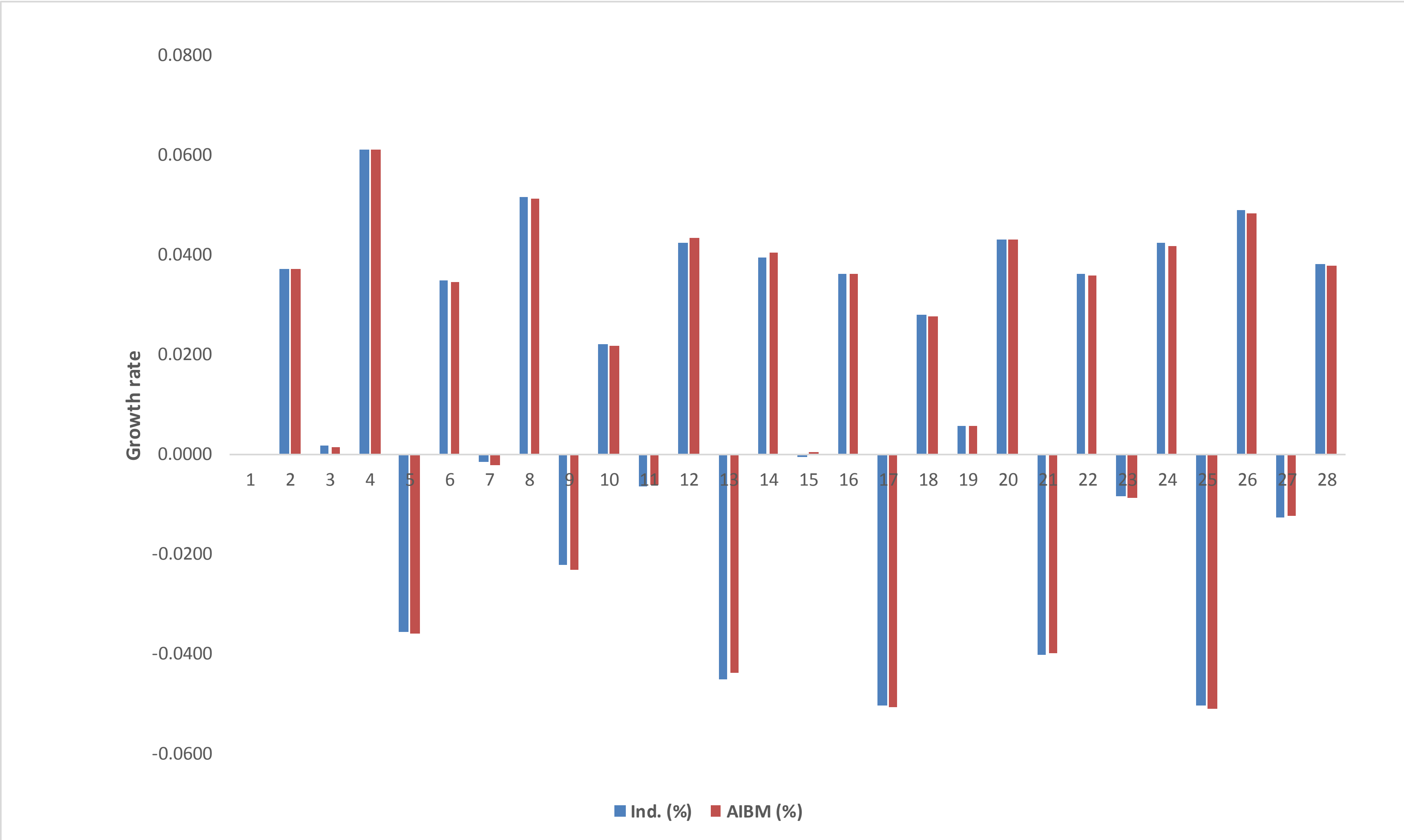

Figure 4. 6: Growth rates graph using the IMF data

The growth rate graph in fig. 4.6 shows the performance of the new method of benchmarking in terms of growth rate percentage. The two bars for each quarter represents the pairs of points from the indicator and the benchmarked series. It can be seen from the chart that lengths of each of the pairs are the same for most of the quarters except in quarters 8 and 9, others are quarters 24 to 28 where the bars are almost the same in length. This also shows that the new method of benchmarking is good enough to handle data reconciliation, especially economic data. It also shows that the new method maintains the original pattern and direction in the original series which is very important in determining a good model in benchmarking.

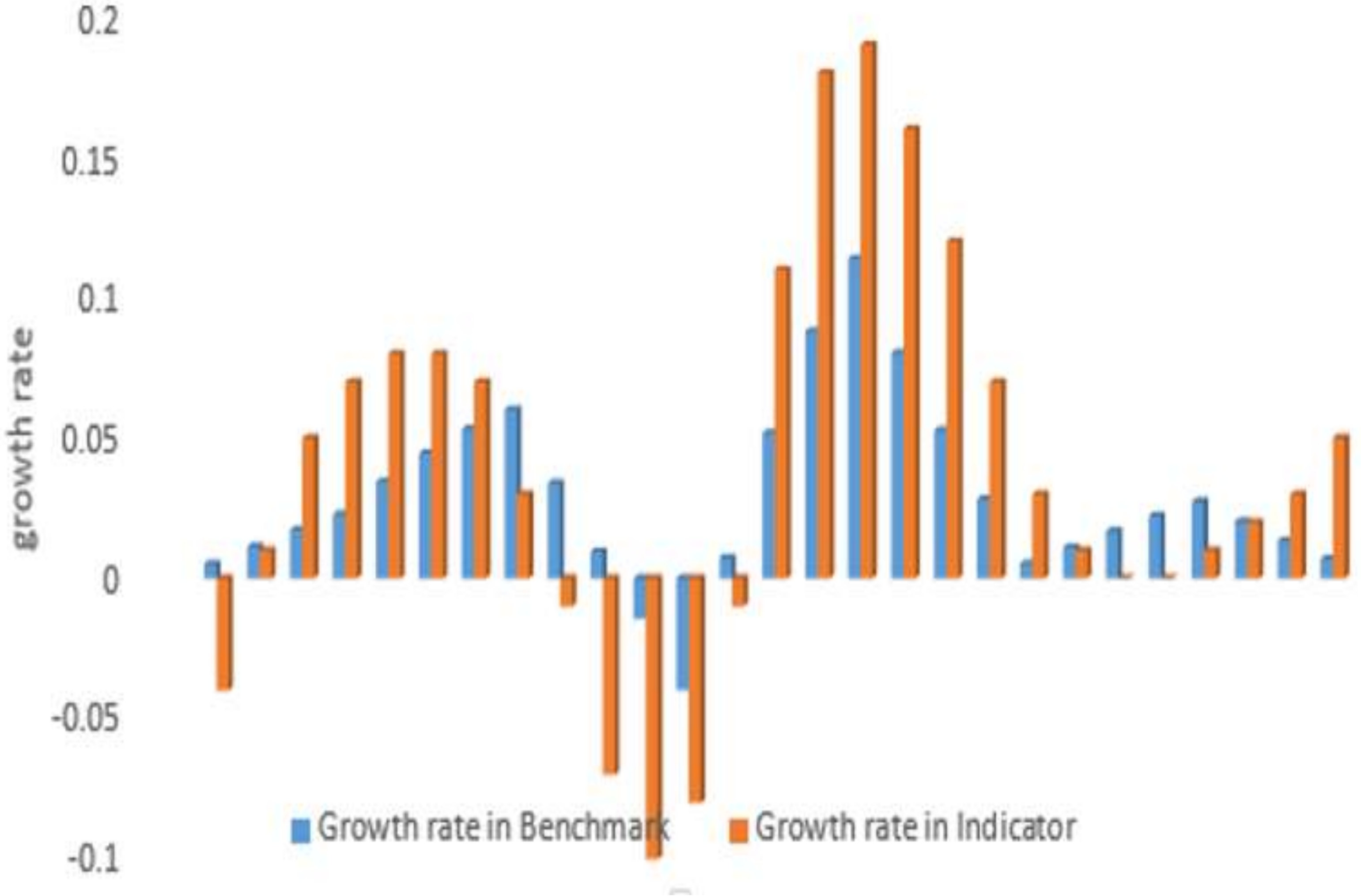

Figure 4. 7: Growth rates graph using the Nigeria GDP

According to fig. 4.7, it can be seen from the chart that differences in the new method and the indicator series is obvious. This may be due to fluctuations in the observations in the data.

Table 4. 4: Benchmarked series from various methods using Cholette data

IND.	ABD	PBD	AOOD	POOD	AOTD	POTD	AIBM
85.00	133.50	119.27	134.50	127.10	132.44	125.58	134.90
95.00	143.50	137.81	144.10	141.70	143.39	140.98	144.15
125.00	173.50	199.11	173.30	185.48	174.09	186.10	173.15
95.00	143.50	137.81	142.10	139.71	144.08	141.33	141.80
85.00	125.00	113.26	130.50	123.54	132.64	125.31	129.94
95.00	135.00	130.31	137.66	135.54	138.82	136.74	137.36
125.00	165.00	186.12	163.58	173.72	162.77	172.93	163.80
95.00	135.00	130.31	128.25	127.20	125.76	125.02	128.90
85.00	115.00	106.20	111.68	108.61	110.17	107.11	112.16
95.00	125.00	121.48	120.38	119.59	119.46	118.56	120.63
125.00	155.00	170.84	154.35	160.90	154.90	161.83	154.14
95.00	125.00	121.48	133.59	130.90	135.47	132.50	133.06
85.00	145.00	127.40	138.10	128.85	137.95	128.91	138.28
95.00	155.00	147.96	156.66	152.61	156.87	152.85	156.70
125.00	185.00	216.69	189.28	205.47	189.46	205.25	189.14
95.00	155.00	147.96	155.95	153.07	155.72	152.98	155.88
85.00	135.00	120.33	136.68	129.69	138.34	130.96	136.56
95.00	145.00	139.13	142.29	140.51	142.76	140.98	142.38
125.00	175.00	201.41	172.81	184.48	172.04	184.02	172.89
95.00	145.00	139.13	148.22	145.32	146.85	144.04	148.17
85.00	155.00	134.46	148.53	138.30	145.51	135.92	148.73
95.00	165.00	156.78	165.24	161.10	163.96	159.83	165.17
125.00	195.00	231.97	198.36	216.55	199.53	217.44	198.15
95.00	165.00	156.78	167.88	164.05	170.99	166.81	167.95
85.00	150.25	131.11	153.81	143.59	158.51	147.39	154.56
95.00	160.25	152.59	160.76	157.67	163.66	160.42	161.76
125.00	190.25	224.71	188.72	204.88	187.63	203.95	187.90
95.00	160.25	152.59	157.71	154.86	151.20	149.24	156.79

Table 4.4 shows the benchmarked series obtained from various benchmarking models using the Cholette simulated data. the first column is the indicator series while the other columns are the ABD (Additive Balanced Difference), PBD (Proportional Balanced Difference), AOOD (Additive Order One Difference), POOD (Proportional Order One Difference), AOTD (Additive Order Two Difference), POTD (Proportional Order Two Difference), and the AIBM (Autocorrelated Indicator Benchmarking Model). Mere checking the series generated from both the existing models (ABD to POTD) and the new model (AIBM), none of the series reconciled well with the indicator series. The disparities are high when the original series is compared with each of the benchmarking methods

Table 4. 5: Benchmarked series from various methods using Denton data

IND.	ABD	PBD	AOOD	POOD	AOTD	POTD	AIBM
50.00	75.00	55.56	79.30	64.33	81.26	66.49	76.13
100.00	125.00	122.22	127.58	127.81	127.26	128.49	127.45
150.00	175.00	200.00	174.14	187.82	173.09	185.91	175.70
100.00	125.00	122.22	118.98	120.04	118.39	119.10	120.72
50.00	50.00	50.00	62.11	56.56	62.64	56.77	61.99
100.00	100.00	100.00	104.51	105.98	105.14	106.70	104.10
150.00	150.00	150.00	146.20	147.50	146.01	147.53	146.25
100.00	100.00	100.00	87.18	89.96	86.21	88.99	87.65
50.00	25.00	44.44	27.44	40.55	27.50	40.09	27.43
100.00	75.00	77.78	72.56	74.45	72.50	74.22	72.57
150.00	125.00	100.00	122.56	108.34	122.50	109.20	122.57
100.00	75.00	77.78	77.44	76.66	77.50	76.49	77.43
50.00	50.00	50.00	37.18	42.76	36.21	42.08	37.65
100.00	100.00	100.00	96.20	94.15	96.01	93.53	96.25
150.00	150.00	150.00	154.51	153.42	155.14	154.01	154.10
100.00	100.00	100.00	112.11	109.67	112.64	110.38	111.99
50.00	75.00	55.56	68.98	58.29	68.39	58.25	70.72
100.00	125.00	122.22	124.14	122.63	123.09	121.63	125.70
150.00	175.00	200.00	177.58	190.41	177.26	189.38	177.45
100.00	125.00	122.22	129.30	128.67	131.26	130.73	126.13

Similar to what is obtained in table 4.1 on Cholette simulated data, Denton data in table 4.2 also show that the benchmarking methods failed to maintain the same pattern and direction in the original series. This is common to simulated data generated by these authors. Further research in this work show that real life data work best for the new model.

Table 4. 6: Benchmarked series from various methods using IMF data

IND.	ABD	PBD	AOOD	POOD	AOTD	POTD	AIBM
613216.00	613358.25	613345.66	613450.56	613442.60	613328.59	613355.36	613407.77
636852.00	636994.25	636991.85	637049.64	637051.41	637001.98	637009.08	637051.56
637890.00	638032.25	638030.31	638013.79	638016.38	638059.76	638045.40	638037.46
679437.00	679579.25	679596.18	679450.01	679453.61	679573.67	679554.16	679467.21
656030.00	655409.50	655457.20	655895.32	655891.09	656027.88	656006.43	655866.00
679720.00	679099.50	679105.09	679331.91	679327.97	679402.91	679396.88	679308.94
678584.00	677963.50	677971.14	677836.80	677854.36	677784.18	677812.39	677848.15
715545.00	714924.50	714863.56	714332.98	714323.58	714182.03	714181.30	714373.91
699993.00	698356.50	698435.69	698210.45	698257.56	698152.69	698200.93	698232.37
715684.00	714047.50	714056.09	713731.66	713739.76	713697.29	713704.07	713742.39
711180.00	709543.50	709572.51	709458.60	709468.09	709480.34	709486.51	709447.74
742734.00	741097.50	740980.71	741644.29	741579.60	741714.69	741653.49	741622.51
710551.00	711271.75	711216.42	710493.71	710455.89	710422.56	710387.33	710534.24
739767.00	740487.75	740488.26	740433.89	740416.36	740414.54	740393.85	740451.09
739391.00	740111.75	740111.53	740473.84	740473.36	740512.13	740510.69	740452.67
767218.00	767938.75	767993.79	768408.56	768464.39	768460.76	768518.13	768372.01
730367.00	731305.00	731240.17	731357.03	731348.54	731467.77	731462.69	731337.42
751285.00	752223.00	752208.90	752174.01	752177.98	752206.43	752215.20	752172.28
755601.00	756539.00	756535.55	756488.49	756479.21	756439.18	756429.02	756497.20
789540.00	790478.00	790560.38	790525.47	790539.26	790431.62	790438.09	790538.10
758885.00	759869.00	759802.83	760067.95	760026.75	759914.51	759874.19	760088.24
787249.00	788233.00	788236.72	788431.36	788434.07	788355.80	788354.92	788448.87
780660.00	781644.00	781631.26	781643.71	781642.99	781692.09	781689.99	781647.66
815231.00	816215.00	816290.19	815817.98	815857.20	815998.60	816041.90	815776.23
776189.00	775660.75	775703.12	776181.19	776205.55	776518.24	776535.95	776017.02
816059.00	815530.75	815521.92	815605.10	815613.15	815845.56	815866.85	815445.07
805903.00	805374.75	805379.21	805151.70	805156.98	805094.72	805108.72	805402.77
837922.00	837393.75	837355.76	837022.01	836984.31	836501.48	836448.48	837095.14

Table 4.6 shows the benchmarked series obtained from various benchmarking models using the IMF real life data. It is observed that all the benchmarking methods performed well as revealed in the table, the indicator series and the generated series from the benchmarking models in the same direction. However, further statistical tests such as the mean, standard variation, variation, and coefficient of variation have to be carried out in order to ascertain the best model.

Table 4. 7: Growth rates analysis on benchmarking methods using the Cholette data $\rho =$ 0.729, 0.900

Data/rho	**Stat.**	**PRM**	**AOOD**	**POOD**	**AOTD**	**POTD**	**B.ADJ**	**AIBM**
Cholette	mean	-0.008	-0.005	-0.016	-0.01	-0.017	-0.008	-0.007
rho=0.729	std	0.222	0.152	0.222	0.155	0.224	0.146	0.153
	var	0.049	0.023	0.049	0.024	0.05	0.021	0.023
	cv	-27.376	-29.62	-14.033	-16.296	-13.16	-19.316	-23.236
Cholette	mean	-0.008	-0.005	-0.016	-0.006	-0.017	-0.008	-0.005
rho=0.900	std	0.222	0.152	0.222	0.154	0.224	0.146	0.153
	var	0.049	0.023	0.049	0.024	0.05	0.021	0.023
	cv	-27.376	-29.62	-14.033	-24.353	-13.16	-19.406	-28.478

As seen from tables 4.7, with Cholette simulated data, using rho = 0.729, the AFD method appears to be the best benchmarking method having the lowest growth rate CV of -29.620. the PRM having -27.376 and the new method AIBM having -23.236. Similarly, when rho = 0.900, the AIBM is better with growth rate CV of -28.478, while AFD has -29.620 and PRM maintains -27.378

Table 4. 8: Comparison using various benchmarking methods rho=0.990 & 0.999

Data/rho	Stat.	PRM	AOOD	POOD	AOTD	POTD	B.ADJ	AIBM
Cholette	mean	-0.008	-0.005	-0.016	-0.006	-0.017	-0.007	-0.005
rho=0.990	std	0.222	0.152	0.222	0.154	0.224	0.145	0.152
	var	0.049	0.023	0.049	0.024	0.05	0.021	0.023
	cv	-27.376	-29.62	-14.033	-24.353	-13.16	-19.591	-29.486
Cholette	mean	-0.008	-0.005	-0.016	-0.006	-0.017	-0.007	-0.005
rho=0.999	std	0.222	0.152	0.222	0.154	0.224	0.145	0.152
	var	0.049	0.023	0.049	0.024	0.05	0.021	0.023
	cv	-27.376	-29.62	-14.033	-24.353	-13.16	-19.632	-29.606

When rho = 0.990, from table 4.12, growth rate CV for AOOD and AIBM methods are -29.620 and -29.486 respectively. When rho = 0.999, the growth rate CV for AOOD and AIBM are almost equal, being -29.620 and -29.606 respectively. All these show that AIBM is a good alternative to the AOOD method. In all the methods of benchmarking considered for Cholette simulated data, the POTD is the worst because it has the highest growth rate CV.

Table 4. 9: Growth rates analysis on benchmarking methods using the Denton data $\rho = 0.729,\ 0.900,\ 0.990,\ \&\ 0.999$

Data/rho	**Stat.**	**PRM**	**AOOD**	**POOD**	**AOTD**	**POTD**	**B.ADJ**	**AIBM**
Denton	mean	-0.137	-0.167	-0.125	-0.168	-0.125	-0.123	-0.166
rho=0.729	std	0.681	0.735	0.621	0.732	0.621	0.613	0.737
	var	0.463	0.541	0.386	0.536	0.386	0.376	0.543
	cv	-4.974	-4.402	-4.987	-4.371	-4.954	-4.993	-4.432
Denton	mean	-0.137	-0.167	-0.125	-0.168	-0.125	-0.08	-0.167
rho=0.900	std	0.681	0.735	0.621	0.732	0.621	0.499	0.736
	var	0.463	0.541	0.386	0.536	0.386	0.249	0.541
	cv	-4.974	-4.402	-4.987	-4.371	-4.954	-6.228	-4.407
Denton	mean	-0.137	-0.167	-0.125	-0.168	-0.125	-0.062	-0.167
rho=0.990	std	0.681	0.735	0.621	0.732	0.621	0.445	0.735
	var	0.463	0.541	0.386	0.536	0.386	0.198	0.541
	cv	-4.974	-4.402	-4.987	-4.371	-4.954	-7.137	-4.402
Denton	mean	-0.137	-0.167	-0.125	-0.168	-0.125	-0.06	-0.167
rho=0.999	std	0.681	0.735	0.621	0.732	0.621	0.436	0.735
	var	0.463	0.541	0.386	0.536	0.386	0.19	0.541
	cv	-4.974	-4.402	-4.987	-4.371	-4.954	-7.309	-4.402

From tables 4.9, using the Denton simulated data, with rho = 0.729, the AOOD and the AIBM are on the leading side, their respective growth rate CV are -4.402 and -4.432. For rho = 0.900, the growth rate CV for AOOD is -4.402 while it is -4.432 for AIBM.

In the two levels of autocorrelation, that is 0.729 and 0.900, it is observed that the POTD has the highest growth rate CV being -4.402 which means it is the worst method as far as the Denton simulated data is concerned.

For rho = 0.990, the B.ADJ is the best benchmarking method having the lowest growth rate CV of -7.137, next to it is PRM having -4.974, the AOOD and AIBM have the same values of -4.402.

For rho = 0.999, the B.ADJ is the best method having the least growth rate CV of -7.309, while the worst method of benchmarking is the AOTD having the highest CV of -4.371 for the two autocorrelation coefficient levels.

Table 4. 10: Growth rates analysis on benchmarking methods using the IMF $\rho = 0.729,\ 0.900,\ 0.990,\ \&\ 0.999$

Data/rho	Stat.	PRM	AOOD	POOD	AOTD	POTD	B.ADJ	AIBM
IMF	mean	0.011	0.011	0.011	0.011	0.011	0.011	0.011
rho=0.729	std	0.035	0.035	0.035	0.035	0.035	0.035	0.035
	var	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	cv	3.194	3.195	3.196	3.198	3.200	3.181	3.192
IMF	mean	0.011	0.011	0.011	0.011	0.011	0.011	0.011
rho=0.900	std	0.035	0.035	0.035	0.035	0.035	0.035	0.035
	var	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	cv	3.194	3.195	3.196	3.198	3.200	3.181	3.194
IMF	mean	0.011	0.011	0.011	0.011	0.011	0.011	0.011
rho=0.990	std	0.035	0.035	0.035	0.035	0.035	0.035	0.035
	var	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	cv	3.194	3.195	3.196	3.198	3.200	3.181	3.195
IMF	mean	0.011	0.011	0.011	0.011	0.011	0.011	0.011
rho=0.999	std	0.035	0.035	0.035	0.035	0.035	0.035	0.035
	var	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	cv	3.194	3.195	3.196	3.198	3.200	3.186	3.195

Tables 4.10 present results on IMF data. When rho = 0.729, the B.ADJ is the best having the least growth rate CV of 3.181. Next to it is the AIBM having growth rate CV = 3.192. The worst method is the POTD having growth rate CV of 3.200. When rho = 0.900, the B.ADJ remains the best having growth rate CV = 3.181, while the AIBM is next having equals 3.194. The POTD remains the worst method.

For rho = 0.990, the B.ABJ is the best benchmarking method with growth CV = 3.181 follows by the AIBM with growth rate CV = 3.195. When rho = 0.999, the growth rate CV for B.ADJ and AIBM are 3.186 and 3.195 respectively. It can also be seen that the worst method is the POTD having growth rate CV = 3.200.

Table 4. 11: Growth rates analysis on benchmarking methods using the Nigeria GDP $\rho = 0.729,\ 0.900,\ 0.990,\ \&\ 0.999$

Data/rho	Stat.	PRM	AOOD	POOD	AOTD	POTD	B.ADJ	AIBM
GDP	mean	0.028	0.011	0.011	0.011	0.011	0.000	0.027
rho=0.729	std	0.073	0.035	0.035	0.035	0.035	0.001	0.035
	var	0.005	0.001	0.001	0.001	0.001	0.000	0.001
	cv	2.665	3.195	3.196	3.198	3.200	1.582	1.318
GDP	mean	0.028	0.011	0.011	0.011	0.011	0.000	0.027
rho=0.900	std	0.073	0.035	0.035	0.035	0.035	0.001	0.035
	var	0.005	0.001	0.001	0.001	0.001	0.000	0.001
	cv	2.665	3.195	3.196	3.198	3.200	1.582	1.318
GDP	mean	0.028	0.011	0.011	0.011	0.011	0.000	0.029
rho=0.990	std	0.073	0.035	0.035	0.035	0.035	0.001	0.032
	var	0.005	0.001	0.001	0.001	0.001	0.000	0.001
	cv	2.665	3.195	3.196	3.198	3.200	1.582	1.121
GDP	mean	0.028	0.011	0.011	0.011	0.011	0.000	0.029
rho=0.999	std	0.073	0.035	0.035	0.035	0.035	0.001	0.032
	var	0.005	0.001	0.001	0.001	0.001	0.000	0.001
	cv	2.665	3.195	3.196	3.198	3.200	1.582	1.105

According to table 4.11, using the Nigeria GDP, for rho = 0.729, the benchmarking methods with the lowest growth rate CV are the B.ADJ and the AIBM, being 1.582 and 1.318 respectively. The best is the AIBM while the POTD has the least performing ability growth rate CV of 3.200. The results obtained for growth rate CVs when rho = 0.729 are the same for when rho = 0.900. This means that the impact of the various levels of autocorrelation coefficients is not obvious.

When rho = 0.990, the growth rate CV for AIBM is the least having 1.121. Next to it is the B.ADJ having 1.582. For rho = 0.999, the AIBM has the lowest growth rate CV of 1.105, which is an improvement on the previous result from rho = 0.990. The growth rate CV for B.ADJ method is 1.582, while the worst benchmarking method using the Nigeria GDP is the POTD having growth rate CV of 3.200

4.6 Discussion of Results

The covariance matrix on page 58 shows the covariances for the aggregated seven years data. The numerator and denominator of each fraction represents total benchmarks and the sum of the annual series respectively. In a binding series the outcome of each of these fractions should be approximately equals 1 and less or greater than 1 in the case of unbinding series. Irregular benchmarks come up when the economic data is unbinding in many cases.

The covariance matrix on page 60 is for annual discrepancies presented as V_{d*} depicts the computed values earlier presented in V_d. The first value 11.2746 shows that the discrepancies between the first year benchmark (annual) and the sub-annual series for the same year is 11.2746. The discrepancy for the second year is 5.1111, the discrepancy for the third is 1.4435, while the discrepancy for the seventh year is 0.0092. None of the years is binding in terms of discrepancies, the benchmarks are irregular and be evaluated and reconciled for suitability in usage for future research and application.

Page 62 presents the matrix of the bias. This is very important in reconciling low frequency (annual data) economic series with high frequency series (sub-annual data). The bias is a regressor or predictor that is deterministic in nature. It is a constant that covers the difference in average level that exist between the low frequency series and the high frequency series. It normally occurs when the high series is not fair with the low series. This happens mostly when there is problem of under coverage or non-coverage in a survey. Bias is an expected phenomenon in benchmarking.

It can be seen in the covariance matrix for sub-annual discrepancies that each of the values in the cell represents the discrepancies in the sub-annual (high frequency series). For instance, 0.7200 is the discrepancy between the annual and sub-annual series for the first year in the series. In this case, we have seven annual series, the sub-annual is twenty-eight, and therefore we are expected to have a 28 by 28 matrix. It can be seen from the matrix that only few values are close to 1, which means they are not binding. This shows that there exists irregular benchmarks in the series, as a result, there is need for evaluation.

From the Cholette simulated series (table 4.1), Column 1 shows the quarters (1981 quarter 1 to 2004 quarter 4). Column 2 the indicator series, column 3 the annual series, column 4 the bias-adjusted series, column 5 the benchmarked series, column 6 and 7 show the growth rates from the indicator series growth rate from benchmarked series using the improved benchmarking method called the Autocorrelated Indicator Benchmarking Model (AIBM). Growth rates are used in measuring the rate of discrepancies between the indicator series and the benchmarked series. It can be observed from the table that rate of growth in the column 7 is lower than the ones observed in column 6. Since the percentage of growth rate is reduced in AIBM it shows that the movement in the original series is better preserved in all the twenty-eight quarters, no higher value of growth rate is observed in the new method.

According to the charts in figure 4.1 used to represent Cholette data, none of the methods reconciled well with the indicator series. They are far away from the indicator series considering the generated data, movement and pattern. This may be due to the nature of the data, therefore the methods may not be appropriate for Cholette simulated data. The discrepancies between the growth rates from indicator and benchmarked series is wide. The growth are close to one another in few pairs of sample points such as quarters 14, 17, 22, and 25. Other 24 pairs of points are not showing favourable growth rates when compared with the indicator.

Using the Denton data in table 4.2, the seven columns represent the quarters, indicator series, annual series, bias adjusted series, benchmarked and the growth rates respectively. Checking the growth rates critically, most of the quarters between 1998-1 and 2002-4 have favourable values, except in quarters 2001-1 and 2002-1 which account for just 2 out of 28 pairs of points. This shows that the method maintains the pattern in the original data, therefore the method is recommended for use considering Denton data.

It can be seen that the benchmarked series using the new method (AIBM) compares favourably well when placed side by side with the original series. It is also observed that the patterns and movements in the series generated by the new method and the indicator series are almost the same. This shows how good the method. This shows that the various methods of benchmarking are also good. But the new method is better when handling some specific data set.

As represented by the charts in figures 4.3 and 4.4, the growth rates in the indicator and the generated series using the AIBM method. The equality in the lengths of pairs of bars shown in the chart depicts equal or almost equal growth rates percentages The Denton simulated data appear to be good for the method since there is no obvious difference in the growth rate. This means that the pattern and movements in the original series are maintained using the improved model.

Using the IMF real life data in table 4.3, the seven columns represent the quarters, indicator series, annual series, bias adjusted series, benchmarked and the growth rates respectively. It can be seen here that the growth rates in the indicator and the bench- marked series obtained from the improved method are approximately the same. This shows that there is no loss of information in the generated series when compared with the original series. The movement and patterns in the indicator series are preserved.

It can be seen from figure 4.5 that the indicator series, bias adjusted series, the benchmarked series, and the average benchmark (obtained by dividing the benchmarks by 4) follow almost the pattern. It is noteworthy that the indicator and the generated benchmarked series are exactly on the same line, points, and direction, there is no distinction, and one is embedded in the other. This indicates a strong graphic similarity that exists between the series. This shows that the method of benchmarking can be recommended for real life data.

The growth rate graph in figure 4.6 show the performance of the new method of benchmarking in terms of growth rate percentage. The two bars for each quarter represents the pairs of points from the indicator and the benchmarked series. It can be seen from the chart that lengths of each of the pairs are the same for most of the quarters except in quarters 8 and 9, others are quarters 24 to 28 where the bars are almost the same in length. This also shows that the new method of benchmarking is good enough to handle data reconciliation, especially economic data. It also shows that the new method maintains the original pattern and direction in the original series which is very important in determining a good model in benchmarking.

The benchmarked series obtained from various benchmarking models using the Cholette simulated data are shown in table 4.4. The first column is the indicator series while the other columns are the ABD (Additive Balanced Difference), PBD (Proportional Balanced Difference), AOOD (Additive Order One Difference), POOD (Proportional Order One Difference), AOTD (Additive Order Two Difference), POTD (Proportional Order Two Difference), and the AIBM (Autocorrelated Indicator Benchmarking Model). Mere checking the series generated from both the existing models (ABD to POTD) and the new model (AIBM), none of the series reconciled well with the indicator series. The disparities are high when the original series is compared with each of the bench- marking methods

Similar to what is obtained on Cholette simulated data, it is also shown that the benchmarking methods failed to maintain the same pattern and direction in the original series when the Denton data is applied. This is common to simulated data generated by these authors. Further research in this work show that real life data work best for the new model.

The benchmarked series obtained from various benchmarking models using the IMF real life data as presented the tables in the preceding chapter. It is observed that all the benchmarking methods performed well as revealed in the table, the indicator series and the generated series from the benchmarking models in the same direction. However, further statistical tests such as the mean, standard variation, variation, and coefficient of variation have to be carried out in order to ascertain the best model. As seen from tables 4.7 and 4.8, with Cholette simulated data, using rho = 0.729, the AOOD method appears to be the best benchmarking method having the lowest growth rate CV of -29.620, the PRM having -27.376 and the new method AIBM having -23.236. Similarly, when rho = 0.900, the AIBM is better with growth rate CV of -28.478, while AOOD has -29.620 and PRM maintains -27.378

When rho = 0.990, growth rate CV for AOOD and AIBM methods are -29.620 and -29.486 respectively. When rho = 0.999, the growth rate CV for AOOD and AIBM are almost equal, being -29.620 and -29.606 respectively. All these show that AIBM is a good alternative to the AOOD method. In all the methods of benchmarking considered for Cholette simulated data, the POTD is the worst because it has the highest growth rate CV. From tables 4.9, using the Denton simulated data, with rho = 0.729, the AOOD and the AIBM are on the leading side, their respective growth rate CV are -29.620 and -29.486. For rho = 0.900, the growth rate CV for AOOD is -29.620 while it is -29.606 for AIBM.

In the two levels of autocorrelation, that is 0.729 and 0.900, it is observed that the POTD has the highest growth rate CV being -13.160 which means it is the worst method as far as the Denton simulated data is concerned. When rho = 0.990, the B.ADJ is the best benchmarking method having the lowest growth rate CV of - 7.137, next to it is PRM having -4.974, the AOOD and AIBM have the same values of -4.402. For rho = 0.999, the B.ADJ is the best method having the least growth rate CV of -7.309, while the worst method of benchmarking is the AOTD having the highest CV of -4.371 for the two autocorrelation coefficient levels.

Considering the IMF data in table 4.10 , when rho = 0.729, the B.ADJ is the best having the least growth rate CV of 3.181. Next to it is the AIBM having growth rate CV = 3.192. The worst method is the POTD having growth rate CV of 3.200. When rho = 0.900, the B.ADJ remains the best having growth rate CV = 3.181, while the AIBM is next having equals 3.194. The POTD remains the worst method.

For rho = 0.990, the B.ABJ is the best benchmarking method with growth CV = 3.181 follows by the AIBM with growth rate CV = 3.195. When rho = 0.999, the growth rate CV for B.ADJ and AIBM are 3.186 and 3.195 respectively. It can also be seen that the worst method is the POTD having growth rate CV = 3.200.

Using the Nigeria GDP as seen the table 4.11, for rho = 0.729, the benchmarking methods with the lowest growth rate CV are the B.ADJ and the AIBM, being 1.582 and 1.318 respectively. The best is the AIBM while the POTD has the least performing ability growth rate CV of 3.200. The results obtained for growth rate CVs when rho = 0.729 are the same for when rho = 0.900. This means that the impact of the various levels of autocorrelation coefficients is not obvious.

When rho = 0.990, the growth rate CV for AIBM is the least having 1.121. Next to it is the B.ADJ having 1.582. For rho = 0.999, the AIBM has the lowest growth rate CV of 1.105, which is an improvement on the previous result from rho = 0.990. The growth rate CV for

B.ADJ method is 1.582, while the worst benchmarking method using the Nigeria GDP is the POTD having growth rate CV of 3.200

However, the autocorrelated indicator benchmarking model (AIBM) performs excellently well with real life data like the GDP, export, import and so on, as proved in the tables of results.

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 Summary of results

This study is an effort to develop an improved approach to modelling and evaluation of economic data with irregular benchmarking. Existing benchmarking models were used and compared with the improved model using various data sets, both simulated and real life economic data.

It is observed that the effect of varying the autocorrelation coefficient is not significant on the growth rates of many of the benchmarking methods for all the data sets used in this study, except on B.ADJ and AIBM methods. The AOTD appears to be good when it comes to using simulated data as recorded in the growth rate CV from Cholette and Denton data.

The introduction of autocorrelation coefficients has much effect when real life economic data sets are used. The benchmarks to indicator improves, which reconciles the benchmarks and the annual data better.

5.2 Conclusion

It can be seen from the tables of results presented in table 4.1 page 65 that the bias-adjusted series and the benchmarked series (θ in the derived model) reconciled with the indicator and the annual series better, as also shown in the charts. The growth rates observed in indicator and benchmarked series as revealed in the charts show the variation in the movement of the growth rate. The large variation may be due to the difference in sub-annual total and the annual obtained from two different sources. On the other hand the close growth rates observed in tables of both the indicator and benchmarked may be due to the B-I ratio recorded in the dataset.

The similar growth rates noticed here shows that the new benchmarking method ensures good movement preservation from the indicator series to the benchmarked series. Tables 4.7 to 4.14 show the mean, standard deviation, variance, and co- efficient of variation for the growth rates from various data sets at different autocorrelation coefficient values: 0.729, 0.900, 0.990, 0.999, which are used in the variance-covariance matrix used in the new autocorrelated indicator benchmarking model (AIBM). It is generally observed from tables 4.7 to 4.14 that the variance and the coefficient of variation for both simulated and real life data are at the mini- mum for our new method of benchmarking, which invariably shows that it is better than the existing ones.

It can be seen from the tables of data both simulated and real that the benchmarked values obtained through our method that it competes favourably with the existing methods of benchmarking. The growth rates charts reveal the variation exhibited in the indicator (sub-annual) series and the benchmarked series. It is also observed that the Bias adjusted series performs better as the increment in the autocorrelated coefficient. The higher the variation the worse the method. Although, it appears that real life data such as the GDP reveal the robustness of the method than the simulated data.

5.3 Contributions to knowledge

The work has contributed to knowledge in the following areas:

i. Derived an alternative matrix solution to benchmarking problem from annual to sub-annuals

ii. Obtained a new benchmarked series that reproduces the movement in the original series to better series

iii. Developed an algorithm for solving benchmarking problem using the new method

iv. Improvement in the B-I ratio

v. No loss of information at the beginning of the benchmarked series when compared with Denton's method

vi. Indicator series are more consistent with the benchmarks produced by the various methods of benchmarking

vii. Movements in the annual B-I ratio help identify the quality of the indicator series in tracking the movements of the annual variables over the period than other benchmarking methods

viii. Better benchmarking method is determined through evaluation of the growth rates

5.4 Recommendation

The autocorrelated indicator benchmarking model captured missing first and last values of the original datasets, while keeping the properties of the data. Since having disparities in economic data collected from two different sources is inevitable, the need to evaluate and reconcile the datasets also arises. The model is therefore recommended for handling irregular data.

REFERENCES

Aadland, D. (2000). "Distribution and interpolation using transformed data", *Journal of Applied Statistics*, 27, 141–156.

Abad A. and Quilis E.M.(2005). Software to perform temporal disaggregation of economic time series. http://epp.eurostat.ec.europa.eu/ cache/ITYPUBLIC/KS- DT-05-004/EN/KS-DT-05-004-EN.PDF

Ajani, H. A. (1978). Quarterly Interpolation of Constant Price Gross Domestic Product of Nigeria. 1963/64 1972/73, *CBN Economic and Financial Review*, 16(2): 24-33.

Ajayi, S. I. (1978). Money in a Developing Economy: A Portfolio Approach To Money Supply Determination, Ibadan University Press, 65.

Akanji, O. O. (1993). Strategic And Complementary Benefit of Quarterly national account Series for Effective Monitoring, Evaluation and Redesign of Development Policies, A paper presented at 1993 User/ Producer Seminar on Statistics held at Otta. November, 1993

Almon, C. (1988). Recent methodological advances in input-output in the United States and Canada, paper presented at the Fourth International Conference on Input-Output Techniques, Geneva.

Asogu, J. O. (1997). Quarterly Interpolation of Annual Statistical Series Using Robust Non-Parametric Methods, *CBN Economic and Financial Review*, 35(2):
154- 170

Barbone, Bodo and Visco (1981). "Temporal disaggregation of Time Series: A comparison by a Monte Carlo experiment, 1970-1980", *Bollettino della Banca d'Italia*, 36, pages 465-510

Barcellan R.(2005). The use of benchmarking techniques in the compilation of the European quarterly national accounts: situation and perspectives. http:// epp.eurostat.ec.europa.eu/cache/ITYPUBLIC /KS-DT-05-

026/EN/KS-DT-05- 026-EN.PDF

Bloem A.M., Dippelsman R.J., and Mæhle O.N. (2001) Quarterly national accounts manual, concepts, data sources, and compilation. Washington, DC: International Monetary Fund (//www.imf.org/external/pubs)

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

Bournay, J., and G. Laroque (1979). Reflexions sur la methode d'elaboration des comptes trimestriels," Annales de l'INSEE, 36, 3–30.

Bozik, J., and M. Otto (1988). "Benchmarking: Evaluating methods that pre serve month-to-month changes,"Bureau of the Census - Statistical Research Division, CENSUS/SRD/RR-88/07.

Bruno G., Di Fonzo T., Golinelli R., and Parigi G.(2005). Short-run GDP forecast- ing in G7 countries: temporal disaggregation techniques and bridge models, http: //epp.eurostat.ec.europa.eu/cache/ITYPUBLIC/KS-DT-05-027/EN/KS- DT-05-027-EN.PDF

Causey, B. and M. L. Trager (1981). "Derivation of solution to the benchmarking problem: trend revision". Unpublished research note, U.S. Census Bureau.

Cholette P.A. (1984). Adjusting sub-annual series to yearly benchmarks, Survey Methodology 10(1): 35-49.

Cholette, P.A. and Chhab, N. (1991). Converting Aggregates of Weekly Data into Monthly Values. *Applied Statistics,* 40, No. 3.41 1422.

Cholette, P.A. and Dagum, E.B. (1994). Benchmarking time series with autocorrelated sampling errors. International Statistical Review 62:365-377

Chow, G., and A. Lin (1971). "Best linear unbiased interpolation, distribution and extrapolation of time series by related series, "*The Review of Economics and*

Statistics, 53, 372–375.

Ciammola A., Di Palma F., and Marini M. (2005). Temporal disaggregation techniques of time series by related series: A comparison by a Monte Carloexperiment.http://epp.eurostat.ec.europa.eu/cache/ITYPUBLIC/ KS-DT-05- 018/EN/KS-DT-05-018-EN.PDF

Cover, T. M. and J. A. Thomas (2006). *Elements of Information Theory*. New Jersey: John Wiley and Sons, 2nd edition.

Dagum, E., and P. Cholette (2006). Cholette-Dagum Regression-Based Benchmarking Method - The Multiplicative Model. In: Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series. Lecture Notes in Statistics, vol 186. Springer, New York

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

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

Di Fonzo, T., and M. Marini (2005). "Benchmarking systems of seasonally adjusted time series according to Denton's movement preservation principle,"Dipartimento di Scienze Statistiche, Universita di Padova, working paper 2003.9.

Durbin J. and B. Quenneville (1997), Benchmarking by state space models, International Statistical Review 65(1): 23-48.

Durbin, J. and Koopman, S. J. (2001). Time Series Analysis by State Space Models. Oxford: Oxford University Press.

Eurostat (1999): Handbook of quarterly national accounts. European Commission, Luxembourg.

Fernandez, R. (1981). "A methodological note on the estimation of time series,"*The*

Review of Economics and Statistics, 63, 471–478.

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

Golan, A., G. Judge, and D. Miller (1996). *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. Chichester, UK: John Wiley and Sons.

Gudmundsson G.,(1999). “Disaggregation of annual flow data with multiplicative trends”, *Journal of Forecasting*, Volume 18, Issue 1, Pages 33 – 37

Guerrero V.M. (1990). Temporal disaggregation of time series: an ARIMA-based approach. International Statistical Review 58:29-46.

Guerrero, V., and D. Pena (2000): “Linear combination of restrictions and forecasts in time series analysis,”*Journal of Forecasting*, 19, 103–122.

Harvey, A., and C. Chung (2000). “Estimating the underlying change in unemployment in the UK,”Journal of the Royal Statistical Society A, 163, 303–328.

Helfand, S., N. Monsour, and M. Trager (1977). “Historical revision of current business survey estimates,”American Statistical Association, *Proceedings of the Business and Economic Statistics Section*: 246-250.

Hillmer, S.C., and Trabelsi, A., (1987). Benchmarking of economic time series, *Journal of the American Statistical Association*, Vol. 82, No. 400

Hood C.C.H. (2005). An empirical comparison of methods for benchmarking sea- sonally adjusted series to annual totals, http:// epp.eurostat.ec.europa.eu/cache/ ITYPUBLIC/KS-DT-05-022/EN/KS-DT-05- 022-EN.PDF

Janes, E. T. (1957a). “Information theory and statistical mechanics”. *Physical Review* 106, 620-630.

Janes, E. T.(1957b). “Information theory and statistical mechanics II”. *Physical Review* 108, 171-190.

Keogh G. and Jennings D.(2005). Investigation into the use of confidence indicators with calibration. http:// epp.eurostat.ec.europa.eu/cache/ ITYPUBLIC/KS- DT-05-023/EN/KS-DT-05- 023-EN.PDF

Kullback, S.(1959). *Information Theory and Statistics*. New York: John Wiley and Sons.

Laniel, N., and K. Fyfe (1989). "Benchmarking of economic time series,"*American Statistical Association*, Proceedings of the Business and Economic Statistics Section: 462-466.

Moauro F. and Proietti T. (2005). Temporal disaggregation and seasonal adjust- ment. http:// epp.eurostat.ec.europa.eu/cache/ ITYPUBLIC/KS-DT-05-011/EN/KS- DT-05-011-EN.PDF

Moauro, F., and G. Savio (2004). "Temporal disaggregation using multivariate structural time series models,"mimeo.

Palate J. (2005). Reusable components for benchmarking using Kalman filters. http:// epp.eurostat.ec.europa.eu/cache/ ITYPUBLIC/KS-DT-05-012/EN/KS- DT-05-012-EN.PDF

Proietti and Moauro (2006). "Dynamic factor analysis with non-linear temporal aggregation constraints". Journal of the Royal Statistical Society

Proietti and Moauro (2008). "Extracting the cyclical component in Hours worked: Bayesian Approach ".

Proietti, T. (1998). "Distribution and interpolation revisited: a structural ap- proach,"*Statistica*, 58, 411–432.

Quenneville, B., P. Cholette, G. Huot, K. Chiu, and T. Di Fonzo (2004). "Adjust- ment of seasonally adjusted series to annual totals,"Proceedings of Statistics Canada Symposium 2003.

Quenneville, B., and S. Fortier (2012), "Restoring Accounting Constraints in Time Series Methods and Software for a Statistical Agency," in *Economic Time Series Modeling and Seasonality*, ed. T.S. McElroy, Chapman and Hall, pp. 231–253

Quilis (2005). Theory and Application of Univariate models for official disaggregation, Luxembourg. Office for official publication of the European communities 2005.

Rodrı́guez Feijoó S., Rodrı́guez Caro A., and Dávila Quintana D. (2003). "Methods for quarterly disaggregation without indicators; a comparative study using simulation", *Computational Statistics & Data Analysis*, Volume 43, Issue 1, pages 63-78

Rojo J.L. and Sanz J.A.,(2005). A Bayesian benchmarking method with applica- tions to the Quarterly National Accounts. http:// epp.eurostat.ec.europa.eu/cache/ ITYPUBLIC/KS-DT-05-013/EN/KS-DT-05-013-EN.PDF

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.

Trager, M.L. (1982). "Derivation of Solution to the Benchmarking Problem: Rel- ative Revision". Unpublished research notes, U.S. Census Bureau

Shannon, C. E. (1948). "A mathematical theory of communication". *Bell System Technical Journal*, 27, 379-423.

Smith P. and Hidiroglou M.,(2005). Benchmarking through calibration of weights for microdota. http:// epp.eurostat.ec.europa.eu/cache/ ITYPUBLIC/KS- DT-05-007/EN/KS-DT-05-007-EN.PDF

Somermeyer, W.H., R. Jansen, and A.S. Louter, (1976). "Estimating Quarterly Values of Annually Known Variables in Quarterly Relationships,"*Journal of the American Statistical Association*, Vol. 71 (September), pp. 588–95.

Temurshoev Umed (2012). "Entropy-based benchmarking methods".European Commission, Research Directorate General as part of the 7thFramework Pro- gramme, Theme 8: Socio-Economic Sciences and Humanities.

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

Wei, W., and D. Stram (1990). "Disaggregation of time series models,"*Journal of the Royal Statistical Society* B, 52, 453 - 467.

APPENDICES

APPENDIX I: MATLAB code for various benchmarking methods using Cholette data

```
s =[85;95;125;95;85;95;125;95;85;95;125;95;85;95;125;95;85;95;125;95;85;95;125;95;85;95;
125;95];   %sub-annual series
R        =        [-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1];
%regressors
a = [560;560;520;640;600;680;680];    %annual series or benchmarks for bias-adjusted series
an = [594;560;520;640;600;680;661];    %annual series or benchmarks for new benchmarked series
av= [140;140;140;140;140;140;140;140;130;130;130;130;160;160;160;160;150;150;150;150;
170;170;170;170;170;170;170;170]; %average benchmarks
T = length(s);       %number of sub-annuals
M = length(a);       %number of years
k = T/M;           %length of each sub-annual data
J = kron(eye(M),ones(k,1));      %quarterly sum operator
J = J';
Id = eye([T T]);     %identity matrix of ones in the diagonal and zero elsewhere
d = a - J*s;        %discrepancies between benchmarks and sum of sub-annual series

Ve = [1          0.729   0.531441        0.387420489   0.282429537   0.205891132   0.150094635
        0.109418989   0.079766443   0.058149737   0.042391158   0.030903154   0.0225284
        0.016423203   0.011972515   0.008727964   0.006362685   0.004638398   0.003381392
        0.002465035   0.00179701     0.001310021   0.000955005   0.000696199   0.000507529
        0.000369989   0.000269722   0.000196627
0.729          1          0.729             0.531441 0.387420489   0.282429537   0.205891132
        0.150094635   0.109418989   0.079766443   0.058149737   0.042391158   0.030903154
        0.0225284       0.016423203   0.011972515   0.008727964   0.006362685   0.004638398
        0.003381392   0.002465035   0.00179701     0.001310021   0.000955005   0.000696199
        0.000507529   0.000369989   0.000269722
0.531441        0.729          1          0.729             0.531441 0.387420489   0.282429537
        0.205891132   0.150094635   0.109418989   0.079766443   0.058149737   0.042391158
```

0.030903154 0.0225284 0.016423203 0.011972515 0.008727964 0.006362685
0.004638398 0.003381392 0.002465035 0.00179701 0.001310021 0.000955005
0.000696199 0.000507529 0.000369989
0.387420489 0.531441 0.729 1 0.729 0.531441 0.387420489
0.282429537 0.205891132 0.150094635 0.109418989 0.079766443 0.058149737
0.042391158 0.030903154 0.0225284 0.016423203 0.011972515 0.008727964
0.006362685 0.004638398 0.003381392 0.002465035 0.00179701 0.001310021
0.000955005 0.000696199 0.000507529
0.282429537 0.387420489 0.531441 0.729 1 0.729 0.531441
0.387420489 0.282429537 0.205891132 0.150094635 0.109418989 0.079766443
0.058149737 0.042391158 0.030903154 0.0225284 0.016423203 0.011972515
0.008727964 0.006362685 0.004638398 0.003381392 0.002465035 0.00179701
0.001310021 0.000955005 0.000696199
0.205891132 0.282429537 0.387420489 0.531441 0.729 1 0.729
0.531441 0.387420489 0.282429537 0.205891132 0.150094635 0.109418989
0.079766443 0.058149737 0.042391158 0.030903154 0.0225284 0.016423203
0.011972515 0.008727964 0.006362685 0.004638398 0.003381392 0.002465035
0.00179701 0.001310021 0.000955005
0.150094635 0.205891132 0.282429537 0.387420489 0.531441 0.729 1
0.729 0.531441 0.387420489 0.282429537 0.205891132 0.150094635 0.109418989
0.079766443 0.058149737 0.042391158 0.030903154 0.0225284 0.016423203
0.011972515 0.008727964 0.006362685 0.004638398 0.003381392 0.002465035
0.00179701 0.001310021
0.109418989 0.150094635 0.205891132 0.282429537 0.387420489 0.531441 0.729
1 0.729 0.531441 0.387420489 0.282429537 0.205891132 0.150094635
0.109418989 0.079766443 0.058149737 0.042391158 0.030903154 0.0225284
0.016423203 0.011972515 0.008727964 0.006362685 0.004638398 0.003381392
0.002465035 0.00179701
0.079766443 0.109418989 0.150094635 0.205891132 0.282429537 0.387420489
0.531441 0.729 1 0.729 0.531441 0.387420489 0.282429537
0.205891132 0.150094635 0.109418989 0.079766443 0.058149737 0.042391158
0.030903154 0.0225284 0.016423203 0.011972515 0.008727964 0.006362685
0.004638398 0.003381392 0.002465035

0.058149737 0.079766443 0.109418989 0.150094635 0.205891132 0.282429537
0.387420489 0.531441 0.729 1 0.729 0.531441 0.387420489
0.282429537 0.205891132 0.150094635 0.109418989 0.079766443 0.058149737
0.042391158 0.030903154 0.0225284 0.016423203 0.011972515 0.008727964
0.006362685 0.004638398 0.003381392
0.042391158 0.058149737 0.079766443 0.109418989 0.150094635 0.205891132
0.282429537 0.387420489 0.531441 0.729 1 0.729 0.531441
0.387420489 0.282429537 0.205891132 0.150094635 0.109418989 0.079766443
0.058149737 0.042391158 0.030903154 0.0225284 0.016423203 0.011972515
0.008727964 0.006362685 0.004638398
0.030903154 0.042391158 0.058149737 0.079766443 0.109418989 0.150094635
0.205891132 0.282429537 0.387420489 0.531441 0.729 1 0.729
0.531441 0.387420489 0.282429537 0.205891132 0.150094635 0.109418989
0.079766443 0.058149737 0.042391158 0.030903154 0.0225284 0.016423203
0.011972515 0.008727964 0.006362685
0.0225284 0.030903154 0.042391158 0.058149737 0.079766443 0.109418989
0.150094635 0.205891132 0.282429537 0.387420489 0.531441 0.729 1
0.729 0.531441 0.387420489 0.282429537 0.205891132 0.150094635 0.109418989
0.079766443 0.058149737 0.042391158 0.030903154 0.0225284 0.016423203
0.011972515 0.008727964
0.016423203 0.0225284 0.030903154 0.042391158 0.058149737 0.079766443
0.109418989 0.150094635 0.205891132 0.282429537 0.387420489 0.531441
0.729 1 0.729 0.531441 0.387420489 0.282429537 0.205891132
0.150094635 0.109418989 0.079766443 0.058149737 0.042391158 0.030903154
0.0225284 0.016423203 0.011972515
0.011972515 0.016423203 0.0225284 0.030903154 0.042391158 0.058149737
0.079766443 0.109418989 0.150094635 0.205891132 0.282429537 0.387420489
0.531441 0.729 1 0.729 0.531441 0.387420489 0.282429537
0.205891132 0.150094635 0.109418989 0.079766443 0.058149737 0.042391158
0.030903154 0.0225284 0.016423203
0.008727964 0.011972515 0.016423203 0.0225284 0.030903154 0.042391158
0.058149737 0.079766443 0.109418989 0.150094635 0.205891132 0.282429537
0.387420489 0.531441 0.729 1 0.729 0.531441 0.387420489

0.282429537 0.205891132 0.150094635 0.109418989 0.079766443 0.058149737
0.042391158 0.030903154 0.0225284
0.006362685 0.008727964 0.011972515 0.016423203 0.0225284 0.030903154
0.042391158 0.058149737 0.079766443 0.109418989 0.150094635 0.205891132
0.282429537 0.387420489 0.531441 0.729 1 0.729 0.531441
0.387420489 0.282429537 0.205891132 0.150094635 0.109418989 0.079766443
0.058149737 0.042391158 0.030903154
0.004638398 0.006362685 0.008727964 0.011972515 0.016423203 0.0225284
0.030903154 0.042391158 0.058149737 0.079766443 0.109418989 0.150094635
0.205891132 0.282429537 0.387420489 0.531441 0.729 1 0.729
0.531441 0.387420489 0.282429537 0.205891132 0.150094635 0.109418989
0.079766443 0.058149737 0.042391158
0.003381392 0.004638398 0.006362685 0.008727964 0.011972515 0.016423203
0.0225284 0.030903154 0.042391158 0.058149737 0.079766443 0.109418989
0.150094635 0.205891132 0.282429537 0.387420489 0.531441 0.729 1
0.729 0.531441 0.387420489 0.282429537 0.205891132 0.150094635 0.109418989
0.079766443 0.058149737
0.002465035 0.003381392 0.004638398 0.006362685 0.008727964 0.011972515
0.016423203 0.0225284 0.030903154 0.042391158 0.058149737 0.079766443
0.109418989 0.150094635 0.205891132 0.282429537 0.387420489 0.531441
0.729 1 0.729 0.531441 0.387420489 0.282429537 0.205891132
0.150094635 0.109418989 0.079766443
0.00179701 0.002465035 0.003381392 0.004638398 0.006362685 0.008727964
0.011972515 0.016423203 0.0225284 0.030903154 0.042391158 0.058149737
0.079766443 0.109418989 0.150094635 0.205891132 0.282429537 0.387420489
0.531441 0.729 1 0.729 0.531441 0.387420489 0.282429537
0.205891132 0.150094635 0.109418989
0.001310021 0.00179701 0.002465035 0.003381392 0.004638398 0.006362685
0.008727964 0.011972515 0.016423203 0.0225284 0.030903154 0.042391158
0.058149737 0.079766443 0.109418989 0.150094635 0.205891132 0.282429537
0.387420489 0.531441 0.729 1 0.729 0.531441 0.387420489
0.282429537 0.205891132 0.150094635
0.000955005 0.001310021 0.00179701 0.002465035 0.003381392 0.004638398

```
        0.006362685   0.008727964   0.011972515   0.016423203   0.0225284     0.030903154
        0.042391158   0.058149737   0.079766443   0.109418989   0.150094635   0.205891132
        0.282429537   0.387420489   0.531441      0.729         1        0.729          0.531441
        0.387420489   0.282429537   0.205891132
0.000696199   0.000955005   0.001310021   0.00179701    0.002465035   0.003381392
        0.004638398   0.006362685   0.008727964   0.011972515   0.016423203   0.0225284
        0.030903154   0.042391158   0.058149737   0.079766443   0.109418989   0.150094635
        0.205891132   0.282429537   0.387420489   0.531441      0.729         1        0.729
0.531441      0.387420489   0.282429537
0.000507529   0.000696199   0.000955005   0.001310021   0.00179701    0.002465035
        0.003381392   0.004638398   0.006362685   0.008727964   0.011972515   0.016423203
        0.0225284     0.030903154   0.042391158   0.058149737   0.079766443   0.109418989
        0.150094635   0.205891132   0.282429537   0.387420489   0.531441      0.729         1
0.729         0.531441 0.387420489
0.000369989   0.000507529   0.000696199   0.000955005   0.001310021   0.00179701
        0.002465035   0.003381392   0.004638398   0.006362685   0.008727964   0.011972515
        0.016423203   0.0225284     0.030903154   0.042391158   0.058149737   0.079766443
        0.109418989   0.150094635   0.205891132   0.282429537   0.387420489   0.531441
        0.729         1        0.729         0.531441
0.000269722   0.000369989   0.000507529   0.000696199   0.000955005   0.001310021
        0.00179701    0.002465035   0.003381392   0.004638398   0.006362685   0.008727964
        0.011972515   0.016423203   0.0225284     0.030903154   0.042391158   0.058149737
        0.079766443   0.109418989   0.150094635   0.205891132   0.282429537   0.387420489
        0.531441      0.729         1        0
0.000196627   0.000269722   0.000369989   0.000507529   0.000696199   0.000955005
        0.001310021   0.00179701    0.002465035   0.003381392   0.004638398   0.006362685
        0.008727964   0.011972515   0.016423203   0.0225284     0.030903154   0.042391158
        0.058149737   0.079766443   0.109418989   0.150094635   0.205891132   0.282429537
        0.387420489   0.531441      0.729         1
]; %covariance matrix for sub-annual series

Vd = J*Ve*J';     %covariance matrix for annual discrepancies
B = -inv(R'*J'*inv(Vd)*J*R)*R'*J'*inv(Vd)*d;    %Bias
```

```
sb = s-R*B;          %Bias-adjusted series
varB = inv(R'*J'*inv(Vd)*J*R);

theta = sb + Ve*J'*inv(Vd)*(an-J*sb);     %Benchmarked series
W = R - Ve*J'*inv(Vd)*J*R;
vartheta = Ve - Ve*J'*inv(Vd)*J*Ve + W*varB*W';

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% DISPLAY OF RESULTS USING TABLES AND GRAPHS
Year=
[1998;1998;1998;1998;1999;1999;1999;1999;2000;2000;2000;2000;2001;2001;2001;2001;
    2002;2002;2002;2002;2003;2003;2003;2003;2004;2004;2004;2004];

Results = [s, sb, theta] % indicator series, bias-adjusted series and benchmarked series

qtrs = [1:28]';

plot(qtrs,s,'r-o',qtrs,sb,'g-',qtrs,theta,'b-*',qtrs,av,'r-*');

xlabel('Quarters');

ylabel('Original Scale');

title('{\bf Indicator series to benchmarked series}');

legend('Indicator series','Bias-adjusted series','Benchmarked series','Average
benchmark','Location','NorthWest');%legend at NW

grid on
```

APPENDIX II: MATLAB code for various benchmarking methods using Denton data

```
s =[50;100;150;100;50;100;150;100;50;100;150;100;50;100;150;100;50;100;150;100];   %sub-
annual series
R = [-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1];      %regressors
a = [500;400;300;400;500];   %annual series or benchmarks for bias-adjusted series
an = [502;404;320;430;504];   %annual series or benchmarks for new benchmarked series
av = [125;125;125;125;100;100;100;100;75;75;75;75;100;100;100;100;125;125;125;125];
%average benchmarks
T = length(s);      %number of sub-annuals
M = length(a);      %number of years
k = T/M;          %length of each sub-annual data
J = kron(eye(M),ones(k,1));     %quarterly sum operator
J = J';
Id = eye([T T]);    %identity matrix of ones in the diagonal and zero elsewhere
d = a - J*s;        %discrepancies between benchmarks and sum of sub-annual series

Ve = [1  0.729 0.531441       0.387420489   0.282429537   0.205891132   0.150094635
        0.109418989   0.079766443   0.058149737   0.042391158   0.030903154   0.0225284
        0.016423203   0.011972515   0.008727964   0.006362685   0.004638398   0.003381392
        0.002465035 0.729     1        0.729   0.531441       0.387420489   0.282429537
        0.205891132   0.150094635   0.109418989   0.079766443   0.058149737   0.042391158
        0.030903154   0.0225284       0.016423203   0.011972515   0.008727964   0.006362685
        0.004638398   0.003381392 0.531441       0.729   1        0.729   0.531441
        0.387420489   0.282429537   0.205891132   0.150094635   0.109418989   0.079766443
        0.058149737   0.042391158   0.030903154   0.0225284       0.016423203   0.011972515
        0.008727964   0.006362685   0.004638398 0.387420489     0.531441       0.729   1
        0.729   0.531441       0.387420489   0.282429537   0.205891132   0.150094635
        0.109418989   0.079766443   0.058149737   0.042391158   0.030903154   0.0225284
        0.016423203   0.011972515   0.008727964   0.006362685 0.282429537     0.387420489
```

0.531441 0.729 1 0.729 0.531441 0.387420489 0.282429537
0.205891132 0.150094635 0.109418989 0.079766443 0.058149737 0.042391158
0.030903154 0.0225284 0.016423203 0.011972515 0.008727964 0.205891132
0.282429537 0.387420489 0.531441 0.729 1 0.729 0.531441
0.387420489 0.282429537 0.205891132 0.150094635 0.109418989 0.079766443
0.058149737 0.042391158 0.030903154 0.0225284 0.0164232 0.0119725
0.150094635 0.205891132 0.282429537 0.387420489 0.531441 0.729 1 0.729 0.531441
0.387420489 0.282429537 0.205891132 0.150094635 0.109418989 0.079766443
0.058149737 0.042391158 0.030903154 0.0225284 0.016423203 0.109418989
0.150094635 0.205891132 0.282429537 0.387420489 0.531441 0.729 1
0.729 0.531441 0.387420489 0.282429537 0.205891132 0.150094635
0.109418989 0.079766443 0.058149737 0.042391158 0.030903154 0.0225284
0.079766443 0.109418989 0.150094635 0.205891132 0.282429537 0.387420489
0.531441 0.729 1 0.729 0.531441 0.387420489 0.282429537
0.205891132 0.150094635 0.109418989 0.079766443 0.058149737 0.042391158
0.030903154
0.058149737 0.079766443 0.109418989 0.150094635 0.205891132 0.282429537
0.387420489 0.531441 0.729 1 0.729 0.531441 0.387420489
0.282429537 0.205891132 0.150094635 0.109418989 0.079766443 0.058149737
0.042391158
0.042391158 0.058149737 0.079766443 0.109418989 0.150094635 0.205891132
0.282429537 0.387420489 0.531441 0.729 1 0.729 0.531441
0.387420489 0.282429537 0.205891132 0.150094635 0.109418989 0.079766443
0.058149737
0.030903154 0.042391158 0.058149737 0.079766443 0.109418989 0.150094635
0.205891132 0.282429537 0.387420489 0.531441 0.729 1 0.729
0.531441 0.387420489 0.282429537 0.205891132 0.150094635 0.109418989
0.079766443 0.0225284 0.030903154 0.042391158 0.058149737 0.079766443
0.109418989 0.150094635 0.205891132 0.282429537 0.387420489 0.531441
0.729 1 0.729 0.531441 0.387420489 0.282429537 0.205891132
0.150094635 0.109418989 0.016423203 0.0225284 0.030903154 0.042391158
0.058149737 0.079766443 0.109418989 0.150094635 0.205891132 0.282429537
0.387420489 0.531441 0.729 1 0.729 0.531441 0.387420489

```
0.282429537  0.205891132  0.150094635 0.011972515  0.016423203  0.0225284
0.030903154  0.042391158  0.058149737  0.079766443  0.109418989  0.150094635
0.205891132  0.282429537  0.387420489  0.531441  0.729  1  0.729
0.531441  0.387420489  0.282429537  0.205891132 0.008727964  0.011972515
0.016423203  0.0225284  0.030903154  0.042391158  0.058149737  0.079766443
0.109418989  0.150094635  0.205891132  0.282429537  0.387420489  0.531441
0.729  1  0.729  0.531441  0.387420489  0.282429537  0.006362685
0.008727964  0.011972515  0.016423203  0.0225284  0.030903154  0.042391158
0.058149737  0.079766443  0.109418989  0.150094635  0.205891132  0.282429537
0.387420489  0.531441  0.729  1  0.729  0.531441 0.387420489 0.004638398
0.006362685  0.008727964  0.011972515  0.016423203  0.0225284  0.030903154
0.042391158  0.058149737  0.079766443  0.109418989  0.150094635  0.205891132
0.282429537  0.387420489  0.531441  0.729  1  0.729  0.531441  0.003381392
0.004638398  0.006362685  0.008727964  0.011972515  0.016423203  0.0225284
0.030903154  0.042391158  0.058149737  0.079766443  0.109418989  0.150094635
0.205891132  0.282429537  0.387420489  0.531441  0.729  1  0.729  0.002465035
0.003381392  0.004638398  0.006362685  0.008727964  0.011972515  0.016423203
0.0225284  0.030903154  0.042391158  0.058149737  0.079766443  0.109418989
0.150094635  0.205891132  0.282429537  0.387420489  0.531441  0.729  1
]; %covariance matrix for sub-annual series

Vd = J*Ve*J';  %covariance matrix for annual discrepancies
B = -inv(R'*J'*inv(Vd)*J*R)*R'*J'*inv(Vd)*d;  %Bias
sb = s-R*B;  %Bias-adjusted series
varB = inv(R'*J'*inv(Vd)*J*R);

theta = sb + Ve*J'*inv(Vd)*(an-J*sb);  %Benchmarked series
W = R - Ve*J'*inv(Vd)*J*R;
vartheta = Ve - Ve*J'*inv(Vd)*J*Ve + W*varB*W';

%% DISPLAY OF RESULTS USING TABLES AND GRAPHS

Year =
```

```
[1998;1998;1998;1998;1999;1999;1999;1999;2000;2000;2000;2000;2001;2001;2001;2001;
    2002;2002;2002;2002;2003;2003;2003;2003;2004;2004;2004;2004];

Results = [s, sb, theta] % indicator series, bias-adjusted series and benchmarked series

qtrs = [1:20]';

plot(qtrs,s,'r-o',qtrs,sb,'g-',qtrs,theta,'b-*',qtrs,av,'r-*');

xlabel('Quarters');

ylabel('Original Scale');

title('{\bf Indicator series to benchmarked series}');

legend('Indicator series','Bias-adjusted series','Benchmarked series','Average
benchmark','Location','NorthWest');%legend at NW

grid on
```

APPENDIX III: MATLAB code for various benchmarking methods using IMF data

```
s = [613216;636852;637890;679437;656030;679720;678584;715545;699993;715684;711180;
742734;710551;739767;739391;767218;730367;751285;755601;789540;758885;787249;78066
0;815231;776189;816059;805903;837922];   %sub-annual series
R = [-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1;-1];
%regressors
a = [2567964;2727397;2863045;2959810;3030545;3145961;3233960];   %annual series or
benchmarks for bias-adjusted series
an = [2567984;2727397;2863045;2959820;3030545;3145961;3233966];   %annual series or
benchmarks for new benchmarked series
av =
[641990.98;641990.98;641990.98;641990.98;681849.36;681849.36;681849.36;681849.36;7157
61.35;715761.35;715761.35;715761.35;739952.40;739952.40;739952.40;739952.40;757636.15;
757636.15;757636.15;757636.15;786490.22;786490.22;786490.22;786490.22;808490.11;80849
0.11;808490.11;808490.11]; %average benchmarks
T = length(s);      %number of sub-annuals
M = length(a);      %number of years
k = T/M;            %length of each sub-annual data
J = kron(eye(M),ones(k,1));     %quarterly sum operator
J = J';
Id = eye([T T]);    %identity matrix of ones in the diagonal and zero elsewhere
d = a - J*s;        %discrepancies between benchmarks and sum of sub-annual series

Ve = [1         0.729   0.531441        0.387420489   0.282429537   0.205891132   0.150094635
        0.109418989   0.079766443   0.058149737   0.042391158   0.030903154   0.0225284
        0.016423203   0.011972515   0.008727964   0.006362685   0.004638398   0.003381392
        0.002465035   0.00179701    0.001310021   0.000955005   0.000696199   0.000507529
        0.000369989   0.000269722   0.000196627
0.729       1       0.729           0.531441 0.387420489   0.282429537   0.205891132
        0.150094635   0.109418989   0.079766443   0.058149737   0.042391158   0.030903154
```

0.0225284 0.016423203 0.011972515 0.008727964 0.006362685 0.004638398
0.003381392 0.002465035 0.00179701 0.001310021 0.000955005 0.000696199
0.000507529 0.000369989 0.000269722

0.531441 0.729 1 0.729 0.531441 0.387420489 0.282429537
0.205891132 0.150094635 0.109418989 0.079766443 0.058149737 0.042391158
0.030903154 0.0225284 0.016423203 0.011972515 0.008727964 0.006362685
0.004638398 0.003381392 0.002465035 0.00179701 0.001310021 0.000955005
0.000696199 0.000507529 0.000369989

0.387420489 0.531441 0.729 1 0.729 0.531441 0.387420489
0.282429537 0.205891132 0.150094635 0.109418989 0.079766443 0.058149737
0.042391158 0.030903154 0.0225284 0.016423203 0.011972515 0.008727964
0.006362685 0.004638398 0.003381392 0.002465035 0.00179701 0.001310021
0.000955005 0.000696199 0.000507529

0.282429537 0.387420489 0.531441 0.729 1 0.729 0.531441
0.387420489 0.282429537 0.205891132 0.150094635 0.109418989 0.079766443
0.058149737 0.042391158 0.030903154 0.0225284 0.016423203 0.011972515
0.008727964 0.006362685 0.004638398 0.003381392 0.002465035 0.00179701
0.001310021 0.000955005 0.000696199

0.205891132 0.282429537 0.387420489 0.531441 0.729 1 0.729
0.531441 0.387420489 0.282429537 0.205891132 0.150094635 0.109418989
0.079766443 0.058149737 0.042391158 0.030903154 0.0225284 0.016423203
0.011972515 0.008727964 0.006362685 0.004638398 0.003381392 0.002465035
0.00179701 0.001310021 0.000955005

0.150094635 0.205891132 0.282429537 0.387420489 0.531441 0.729 1
0.729 0.531441 0.387420489 0.282429537 0.205891132 0.150094635 0.109418989
0.079766443 0.058149737 0.042391158 0.030903154 0.0225284 0.016423203
0.011972515 0.008727964 0.006362685 0.004638398 0.003381392 0.002465035
0.00179701 0.001310021

0.109418989 0.150094635 0.205891132 0.282429537 0.387420489 0.531441 0.729
1 0.729 0.531441 0.387420489 0.282429537 0.205891132 0.150094635
0.109418989 0.079766443 0.058149737 0.042391158 0.030903154 0.0225284
0.016423203 0.011972515 0.008727964 0.006362685 0.004638398 0.003381392
0.002465035 0.00179701

0.079766443 0.109418989 0.150094635 0.205891132 0.282429537 0.387420489
0.531441 0.729 1 0.729 0.531441 0.387420489 0.282429537
0.205891132 0.150094635 0.109418989 0.079766443 0.058149737 0.042391158
0.030903154 0.0225284 0.016423203 0.011972515 0.008727964 0.006362685
0.004638398 0.003381392 0.002465035
0.058149737 0.079766443 0.109418989 0.150094635 0.205891132 0.282429537
0.387420489 0.531441 0.729 1 0.729 0.531441 0.387420489
0.282429537 0.205891132 0.150094635 0.109418989 0.079766443 0.058149737
0.042391158 0.030903154 0.0225284 0.016423203 0.011972515 0.008727964
0.006362685 0.004638398 0.003381392
0.042391158 0.058149737 0.079766443 0.109418989 0.150094635 0.205891132
0.282429537 0.387420489 0.531441 0.729 1 0.729 0.531441
0.387420489 0.282429537 0.205891132 0.150094635 0.109418989 0.079766443
0.058149737 0.042391158 0.030903154 0.0225284 0.016423203 0.011972515
0.008727964 0.006362685 0.004638398
0.030903154 0.042391158 0.058149737 0.079766443 0.109418989 0.150094635
0.205891132 0.282429537 0.387420489 0.531441 0.729 1 0.729
0.531441 0.387420489 0.282429537 0.205891132 0.150094635 0.109418989
0.079766443 0.058149737 0.042391158 0.030903154 0.0225284 0.016423203
0.011972515 0.008727964 0.006362685
0.0225284 0.030903154 0.042391158 0.058149737 0.079766443 0.109418989
0.150094635 0.205891132 0.282429537 0.387420489 0.531441 0.729 1
0.729 0.531441 0.387420489 0.282429537 0.205891132 0.150094635 0.109418989
0.079766443 0.058149737 0.042391158 0.030903154 0.0225284 0.016423203
0.011972515 0.008727964
0.016423203 0.0225284 0.030903154 0.042391158 0.058149737 0.079766443
0.109418989 0.150094635 0.205891132 0.282429537 0.387420489 0.531441
0.729 1 0.729 0.531441 0.387420489 0.282429537 0.205891132
0.150094635 0.109418989 0.079766443 0.058149737 0.042391158 0.030903154
0.0225284 0.016423203 0.011972515
0.011972515 0.016423203 0.0225284 0.030903154 0.042391158 0.058149737
0.079766443 0.109418989 0.150094635 0.205891132 0.282429537 0.387420489
0.531441 0.729 1 0.729 0.531441 0.387420489 0.282429537

0.205891132 0.150094635 0.109418989 0.079766443 0.058149737 0.042391158
0.030903154 0.0225284 0.016423203
0.008727964 0.011972515 0.016423203 0.0225284 0.030903154 0.042391158
0.058149737 0.079766443 0.109418989 0.150094635 0.205891132 0.282429537
0.387420489 0.531441 0.729 1 0.729 0.531441 0.387420489
0.282429537 0.205891132 0.150094635 0.109418989 0.079766443 0.058149737
0.042391158 0.030903154 0.0225284
0.006362685 0.008727964 0.011972515 0.016423203 0.0225284 0.030903154
0.042391158 0.058149737 0.079766443 0.109418989 0.150094635 0.205891132
0.282429537 0.387420489 0.531441 0.729 1 0.729 0.531441
0.387420489 0.282429537 0.205891132 0.150094635 0.109418989 0.079766443
0.058149737 0.042391158 0.030903154
0.004638398 0.006362685 0.008727964 0.011972515 0.016423203 0.0225284
0.030903154 0.042391158 0.058149737 0.079766443 0.109418989 0.150094635
0.205891132 0.282429537 0.387420489 0.531441 0.729 1 0.729
0.531441 0.387420489 0.282429537 0.205891132 0.150094635 0.109418989
0.079766443 0.058149737 0.042391158
0.003381392 0.004638398 0.006362685 0.008727964 0.011972515 0.016423203
0.0225284 0.030903154 0.042391158 0.058149737 0.079766443 0.109418989
0.150094635 0.205891132 0.282429537 0.387420489 0.531441 0.729 1
0.729 0.531441 0.387420489 0.282429537 0.205891132 0.150094635 0.109418989
0.079766443 0.058149737
0.002465035 0.003381392 0.004638398 0.006362685 0.008727964 0.011972515
0.016423203 0.0225284 0.030903154 0.042391158 0.058149737 0.079766443
0.109418989 0.150094635 0.205891132 0.282429537 0.387420489 0.531441
0.729 1 0.729 0.531441 0.387420489 0.282429537 0.205891132
0.150094635 0.109418989 0.079766443
0.00179701 0.002465035 0.003381392 0.004638398 0.006362685 0.008727964
0.011972515 0.016423203 0.0225284 0.030903154 0.042391158 0.058149737
0.079766443 0.109418989 0.150094635 0.205891132 0.282429537 0.387420489
0.531441 0.729 1 0.729 0.531441 0.387420489 0.282429537
0.205891132 0.150094635 0.109418989
0.001310021 0.00179701 0.002465035 0.003381392 0.004638398 0.006362685

0.008727964 0.011972515 0.016423203 0.0225284 0.030903154 0.042391158
0.058149737 0.079766443 0.109418989 0.150094635 0.205891132 0.282429537
0.387420489 0.531441 0.729 1 0.729 0.531441 0.387420489
0.282429537 0.205891132 0.150094635
0.000955005 0.001310021 0.00179701 0.002465035 0.003381392 0.004638398
0.006362685 0.008727964 0.011972515 0.016423203 0.0225284 0.030903154
0.042391158 0.058149737 0.079766443 0.109418989 0.150094635 0.205891132
0.282429537 0.387420489 0.531441 0.729 1 0.729 0.531441
0.387420489 0.282429537 0.205891132
0.000696199 0.000955005 0.001310021 0.00179701 0.002465035 0.003381392
0.004638398 0.006362685 0.008727964 0.011972515 0.016423203 0.0225284
0.030903154 0.042391158 0.058149737 0.079766443 0.109418989 0.150094635
0.205891132 0.282429537 0.387420489 0.531441 0.729 1 0.729
0.531441 0.387420489 0.282429537
0.000507529 0.000696199 0.000955005 0.001310021 0.00179701 0.002465035
0.003381392 0.004638398 0.006362685 0.008727964 0.011972515 0.016423203
0.0225284 0.030903154 0.042391158 0.058149737 0.079766443 0.109418989
0.150094635 0.205891132 0.282429537 0.387420489 0.531441 0.729 1
0.729 0.531441 0.387420489
0.000369989 0.000507529 0.000696199 0.000955005 0.001310021 0.00179701
0.002465035 0.003381392 0.004638398 0.006362685 0.008727964 0.011972515
0.016423203 0.0225284 0.030903154 0.042391158 0.058149737 0.079766443
0.109418989 0.150094635 0.205891132 0.282429537 0.387420489 0.531441
0.729 1 0.729 0.531441
0.000269722 0.000369989 0.000507529 0.000696199 0.000955005 0.001310021
0.00179701 0.002465035 0.003381392 0.004638398 0.006362685 0.008727964
0.011972515 0.016423203 0.0225284 0.030903154 0.042391158 0.058149737
0.079766443 0.109418989 0.150094635 0.205891132 0.282429537 0.387420489
0.531441 0.729 1 0
0.000196627 0.000269722 0.000369989 0.000507529 0.000696199 0.000955005
0.001310021 0.00179701 0.002465035 0.003381392 0.004638398 0.006362685
0.008727964 0.011972515 0.016423203 0.0225284 0.030903154 0.042391158
0.058149737 0.079766443 0.109418989 0.150094635 0.205891132 0.282429537

```
        0.387420489   0.531441        0.729          1
]; %covariance matrix for sub-annual series

Vd = J*Ve*J';      %covariance matrix for annual discrepancies
B = -inv(R'*J'*inv(Vd)*J*R)*R'*J'*inv(Vd)*d;     %Bias
sb = s-R*B;          %Bias-adjusted series
varB = inv(R'*J'*inv(Vd)*J*R);

theta = sb + Ve*J'*inv(Vd)*(an-J*sb);     %Benchmarked series
W = R - Ve*J'*inv(Vd)*J*R;
vartheta = Ve - Ve*J'*inv(Vd)*J*Ve + W*varB*W';

%% DISPLAY OF RESULTS USING TABLES AND GRAPHS

Year =
[1998;1998;1998;1998;1999;1999;1999;1999;2000;2000;2000;2000;2001;2001;2001;2001;
    2002;2002;2002;2002;2003;2003;2003;2003;2004;2004;2004;2004];

Results = [s, sb, theta] % indicator series, bias-adjusted series and benchmarked series

qtrs = [1:28]';

plot(qtrs,s,'r-o',qtrs,sb,'black-',qtrs,theta,'b-*',qtrs,av,'r-*');

xlabel('Quarters');

ylabel('Original Scale');

title('{\bf Indicator series to benchmarked series}');

legend('Indicator series','Bias-adjusted series','Benchmarked series','Average
benchmark','Location','SouthEast');%legend at NW
```

grid on

$$J' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Where

J' is the matrix for the temporal sum operator of 7 by 28 dimension

$$sb = \begin{bmatrix} 136.9538 \\ 146.9538 \\ 176.9538 \\ 146.9538 \\ 136.9538 \\ 146.9538 \\ 176.9538 \\ 146.9538 \\ 136.9538 \\ 146.9538 \\ 176.9538 \\ 146.9538 \\ 136.9538 \\ 146.9538 \\ 176.9538 \\ 146.9538 \\ 136.9538 \\ 146.9538 \\ 176.9538 \\ 146.9538 \\ 136.9538 \\ 146.9538 \\ 176.9538 \\ 146.9538 \\ 136.9538 \\ 146.9538 \\ 176.9538 \\ 146.9538 \end{bmatrix}$$

Where
sb = s - R*B is the bias-adjusted series of 28 by 1 dimension