



Horizontal and vertical Aggregation Bias of time series with empirical study

Prepared by

Samar Ahmed Helmy Abdelghany

Lecturer, Department of Statistics, Mathematics & Insurance

Faculty of Commerce, Port Said University

blue_arkedia@hotmail.com

*Scientific Journal for Financial and Commercial Studies and
Researches (SJFCSR)*

Faculty of Commerce – Damietta University

Vol.3, No.1, Part 1., January 2022

APA Citation:

Abdelghany, S. A. H. (2022). Horizontal and vertical Aggregation Bias of time series with empirical study, *Scientific Journal for Financial and Commercial Studies and Research*, Faculty of Commerce, Damietta University, 3(1)1, pp. 397-428.

Website: <https://cfdj.journals.ekb.eg/>

Horizontal and vertical Aggregation Bias of time series with empirical study

Samar Ahmed Helmy Abdelghany

Abstract

Aggregation bias is defined (Theil 1954) as the difference between the macro parameter and the average of the micro parameters. In this article, a new definition of time series aggregation is introduced. Imperially, a new definition of aggregation bias is introduced. Using the new definition, the time series are aggregated in two different methods, horizontal and vertical. Bias resulting from this aggregation is analyzed in both cases. Horizontal aggregation revealed unbiasedness in the slope and intercept of the trend of the series. Vertical aggregation revealed bias in estimating both parameters of the trend of the series depending on the length of the aggregation period. A new theorem "Samar theorem" is introduced to assess bias in both parameters. Proof of the Samar theorem is annexed to the paper. An Empirical study containing data of a number of time series are analyzed in the frame of the new theorem. The data is the weekly income of some Egyptian farmers working on the same farm collected through an Egyptian agrarian year. Samar theorem introduce rules to compute bias in the slope and the intercept of the trend of the series. It also shows the relation between the two biases. In addition, the effect of the length of aggregation period on the amount of bias is discussed.

Key words: Aggregation bias, Egyptian agrarian year, Trend, Samar theorem

Introduction

Moorman (1979) stated that any discussion of aggregation bias necessarily begins with Robinson's (1950) well-known article; "Ecological Correlation and the Behavior of Individuals"; Since that article, there has been considerable discussion of the ecological fallacy, but little attempt to actually measure the amount of bias introduced by aggregation of individuals into areal units when data existed at both levels.

Albuquerque (2003) showed that the exact aggregate representation derived is relatively simple and intuitive. It can be used thereafter in applied issues and in teaching, easing the solving and understanding of aggregation problems.

Franta (2008) focused in his paper on the dynamics of unemployment in the Czech Republic over the period 1992-2007. Unemployment dynamics are elaborated in terms of unemployment inflows and unemployment duration. The paper contributes to the literature dealing with discrete time models of aggregate unemployment duration data by accounting for time aggregation bias.

Sbrana (2011) derived the analytical relationships between structural and reduced form parameters of the local linear trend model with correlated shocks. The results are also valid in the context of temporal aggregation of local level and local linear trend models with their corresponding ARIMA(0,1,1) and ARIMA(0,2,2) reduced forms.

Many research work focused on different aspects of aggregation. Little attention was given to aggregation bias in time series. This article sheds some light on the issue.

1.1 Definition of aggregation bias

Aggregation bias is defined according to Theil (1954) as the difference between the macro parameter and the average of the micro parameter. For example, the following single equation model can be expressed as:

$$Y_{ij} = \alpha_j + \beta_{1i} X_{1ij} + \beta_{2i} X_{2ij} + \dots + \beta_{ki} X_{kij} + e_{ij} \quad (1.1)$$

Where:

- $i = 1, 2, \dots, I$ number of equations
- $j = 1, 2, \dots, n$ number of observations
- $k =$ number of explanatory variables

The model is aggregated over i so the macro relation becomes

$$\sum_{i=1}^I Y_{ij} = A + B_1 \sum_{i=1}^I X_{1ij} + B_2 \sum_{i=1}^I X_{2ij}, \dots, + B_k \sum_{i=1}^I X_{kij} + E_j \quad (1.2)$$

Notice that all variables, dependent or independent, are aggregated.

1.2 Time series

As known, a time series is a collection of observations of well-defined data items obtained through repeated measurements over time. For example, measuring the amount of weekly income of a person would comprise a time series. This is because personal income is well defined, and consistently measured at equally spaced intervals.

An observed time series may be decomposed into four components: the Secular Trend (Trend), the (Seasonal Variations), the (Cyclical Variations) and the (Irregular Variations)

The study focuses on the trend only

1.2 Types of aggregating time series

The functional form of the time series is $Y = f(T)$ where T is the time. If we have (n) observations on a number (I) of time series, we have:

$$Y_{iT} = \alpha_j + \beta_i T + e_{iT} \quad T = 1, 2, \dots, n, \quad i = 1, \dots, I \quad (1.3)$$

Equation (1.3) is aggregated horizontally as follows:

$$\sum_{i=1}^I Y_{iT} = A + B T + E \quad T = 1, 2, \dots, n \quad (1.4)$$

Notice that **only the dependent** variable is aggregated, where T , the independent variable is not aggregated. This is a crucial difference than that in (1.1) which implies different definition of aggregation bias.

Aggregation bias of time series depends on the method of aggregation. Time series is aggregated in two different methods:

- a - Contemporaneous Aggregation (Cross Sectional Aggregation) i.e aggregating some variables in the same time, e.g data necessary to establish an index number. As for time series it means aggregating a number of time series, equation (1.3), to form one series so the aggregated series will be equation (1.4). This type of aggregation will be called **horizontal aggregation**
- b - Temporal Aggregation means aggregation on the same series, i.e grouping the observations to a new timely series, e.g aggregating weekly data to form monthly data. This type of aggregation will be called **vertical aggregation**. Let:

$$Z_X = \sum_{j=1}^k Y_{kX-k+j}, E_X = \sum_{j=1}^k e_{kX-k+j}, X=1, \dots, \frac{n}{k}, j=1, \dots, k$$

Where k is the number of observations added together, so k = 4 if quarterly data are converted into yearly data or k = 12 if the detailed data were monthly etc.

Then the aggregated equation will

$$Z_X = A + BX + E_X \quad X=1, \dots, \frac{n}{k} \quad (1.5)$$

2 Calculation of aggregation bias of time series

To calculate the aggregation bias, the following steps are followed:

- a. Estimate the parameters of the detailed equation using a suitable estimation method
- b. Estimate the parameters of the aggregated equation using the same estimation method
- c. The aggregation bias in the parameters of the time series can be represented for **horizontal aggregation** as

$$\text{Bias in the slope} = B - \sum_{i=1}^I \beta_i \quad (2.1)$$

$$\text{Bias in the intercept} = A - \sum_{i=1}^I \alpha_i \quad (2.2)$$

That is, bias will be the difference between the aggregated parameter and the sum of individual parameters. For example, if the monthly increase in the income of a person is L.E 20 and that of another person is 30 so the aggregated increase will be L.E 50. The same argument applies to the intercept.

For convenience, if we use deviations from the related mean of the variables of (1.3), it becomes

$$y_{ix} = \beta_i x + e_{ix} \quad , i = 1, \dots, I, \quad x = 1, \dots, n \quad (2.3)$$

Where: $y_{ix} = Y_{ix} - \bar{Y}_i$, $i = 1, \dots, I$

$$x = T - \bar{T} \quad (x \text{ is the same in all series})$$

$$e_x = e_x - \bar{e} = e_x - 0 = e_x$$

n is the number of observations on time x

The OLS estimator of β is

$$\hat{\beta}_i = \frac{\sum_{x=1}^n y_x x}{\sum_{x=1}^n x^2} \quad i = 1, \dots, I, \quad x = 1, \dots, n \quad (2.4)$$

, the estimator of α is

$$\hat{\alpha}_i = \bar{Y}_i - \hat{\beta}_i \bar{T} \quad i = 1, \dots, I \quad \bar{T} \text{ is the same in all series} \quad (2.5)$$

and the estimated trend is

$$\hat{Y}_i = \hat{\alpha}_i + \hat{\beta}_i T \quad (2.6)$$

2.1 Bias in the slope and intercept in the horizontal aggregation

For convenience, consider two time series in the deviated form:

$$y_{1x} = \beta_1 x + e_{1x} \quad , x = 1, \dots, n \quad (2.7)$$

$$y_{2x} = \beta_2 x + e_{2x} \quad , x = 1, \dots, n \quad (2.8)$$

the horizontal aggregation of (2.7) and (2.8) is:

$$z_x = Bx + E_x \quad , x = 1, \dots, n \quad (2.9)$$

Where :

- z_x is the sum of the two series $z_x = y_{1x} + y_{2x}$
- x time deviations
- E_x sum of error in the two series $E_x = e_{1x} + e_{2x}$
- n number of observations

Estimator of the aggregated slope is:

$$\hat{B} = \frac{\sum_{x=1}^n z_x x}{\sum_{x=1}^n x^2} \quad (2.10)$$

Estimator of the aggregated intercept is: (2.11)

$$\hat{A} = \bar{Z} - \hat{B} \bar{T}$$

By definition of the aggregation bias in the slope, of equation (2.9)

$$\text{Bias}(\hat{B}) = \hat{B} - (\hat{\beta}_1 + \hat{\beta}_2) \quad (2.12)$$

and the aggregation bias in the intercept, of equation (2.9)

$$\text{Bias}(\hat{A}) = \hat{A} - (\hat{\alpha}_1 + \hat{\alpha}_2) \quad (2.13)$$

To derive the bias in the slope, substitute in (2.12)

$$\begin{aligned} \text{Bias } (\hat{B}) &= \frac{\sum_{t=1}^n z_t x_t}{\sum_{x=1}^n x^2} - \left[\frac{\sum_{x=1}^n y_{1x} x}{\sum_{x=1}^n x^2} + \frac{\sum_{x=1}^n y_{2x} x}{\sum_{x=1}^n x^2} \right] \\ &= \frac{\sum_{t=1}^n z_t x_t - \left[\sum_{x=1}^n y_{1x} x + \sum_{x=1}^n y_{2x} x \right]}{\sum_{x=1}^n x^2} \end{aligned}$$

substitute $z_x = y_{1x} + y_{2x}$

$$\begin{aligned} \text{Bias } (\hat{B}) &= \frac{\sum_{x=1}^n (y_{1x} + y_{2x}) x - \left[\sum_{x=1}^n y_{1x} x + \sum_{x=1}^n y_{2x} x \right]}{\sum_{t=1}^n x_t^2} \\ &= \frac{\sum_{x=1}^n y_{1x} x + \sum_{x=1}^n y_{2x} x - \left[\sum_{x=1}^n y_{1x} x + \sum_{x=1}^n y_{2x} x \right]}{\sum_{x=1}^n x^2} = \frac{0}{\sum_{t=1}^n x_t^2} = 0 \quad (2.14) \end{aligned}$$

Then horizontal aggregation of time series does not result in aggregation bias in the slope

To derive aggregation bias in the intercept, substitute in (2.13)

$$\begin{aligned} \text{Bias } (\hat{A}) &= \hat{A} - (\hat{\alpha}_1 + \hat{\alpha}_2) \\ &= (\bar{Z} - \hat{B}\bar{T}) - \left[(\bar{Y}_1 - \hat{\beta}_1 \bar{T}) + (\bar{Y}_2 - \hat{\beta}_2 \bar{T}) \right] \\ &= \bar{Z} - \hat{B}\bar{T} - \bar{Y}_1 + \hat{\beta}_1 \bar{T} - \bar{Y}_2 + \hat{\beta}_2 \bar{T} \end{aligned}$$

$$\text{since } Z = Y_1 + Y_2 \Rightarrow \bar{Z} = \bar{Y}_1 + \bar{Y}_2 \Rightarrow \bar{Z} - \bar{Y}_1 - \bar{Y}_2 = 0$$

$$\text{Bias } (\hat{A}) = -\hat{B}\bar{T} + \hat{\beta}_1 \bar{T} + \hat{\beta}_2 \bar{T} = \bar{T} (\hat{\beta}_1 + \hat{\beta}_2 - \hat{B})$$

Substitute for the slopes

$$\text{Bias}(\hat{A}) = \bar{T} \left(\left[\frac{\sum_{x=1}^n y_{1x} x}{\sum_{x=1}^n x^2} + \frac{\sum_{x=1}^n y_{2x} x}{\sum_{x=1}^n x^2} \right] - \frac{\sum_{t=1}^n z_t x_t}{\sum_{x=1}^n x^2} \right) \quad (2.15)$$

using (2.14)

$$\text{Bias}(\hat{A}) = \bar{T}(0) = 0 \quad (2.16)$$

Also, horizontal aggregation of time series does not result in bias in the intercept. Since $\hat{A} = \bar{Z} - \hat{B}\bar{T}$ as \bar{Z} and \bar{T} are constants, so if \hat{B} is bias free, then \hat{A} should be bias free too.

Then we conclude that horizontal aggregation does not result in any bias in the parameters of the series.

2.2 Bias in the slope and intercept in the vertical aggregation

To establish a reference criterion to calculate the aggregation bias, an algebraic series, i.e error free, is vertically aggregated with different periods of aggregation k . For the series

$$Y_T = 13 + 3T, \quad T = 1, \dots, 60 \quad (2.17)$$

the intercept $\alpha = 13$ and the slope $\beta = 3$. Linear regression using SPSS on data with different periods of aggregation k , starting with $k = 2$, can be shown as in table (1)

Table (1)
Coefficients of the algebraic equation $Y_T = 13 + 3T$ aggregated
on period $k = 2$
Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	23.000	.000		.	.
k = 2	12.000	.000	1.000	.	.

a. Dependent Variable: Aggregated 2

of course, there is no error and the result is

$$Z_x = 23 + 12 X \quad ,x = 1, \dots, 30 \left(\frac{n}{k} \right) \quad (2.18)$$

the slope B(k) is $B(2) = 12 = 3 \times (2 \times 2) = \beta k^2$ (2.19)

the intercept A(k) is $A(2) = 23$ (2.20)

when $k = 3$ the result is

$$H_R = 30 + 27 R \quad ,R = 1, \dots, 20 \left(\frac{n}{k} \right) \quad (2.21)$$

the slope B(k) is $B(3) = 27 = 3 \times (3 \times 3) = \beta k^2$ (2.22)

the intercept A(k) is $A(3) = 30$ (2.23)

with $k = 4$ the result is

$$M_Q = 34 + 48 Q \quad ,Q = 1, \dots, 15 \left(\frac{n}{k} \right) \quad (2.24)$$

Samar Ahmed Helmy Abdelghany

the slope B(k) is $B(4) = 48 = 3 \times (4 \times 4) = \beta k^2$ (2.25)

the intercept A(k) is $A(4) = 34$ (2.26)

with $k = 5$ the result is

$$S_V = 35 + 75 V, V = 1, \dots, 12 \left(\frac{n}{k}\right) \quad (2.27)$$

the slope B(k) is $B(5) = 75 = 3 \times (5 \times 5) = \beta k^2$ (2.28)

the intercept A(k) is $A(5) = 35$ (2.29)

Generally, aggregating on a period k , the detailed equation is

$$Y_t = \alpha + \beta T + e_t, T = 1, \dots, n \quad (2.30)$$

taking

$$Z_X(k) = \sum_{j=1}^k Y_{kX-k+j}, E_X(k) = \sum_{j=1}^k e_{kX-k+j}, X = 1, \dots, \frac{n}{k}, j = 1, \dots, k$$

the aggregated equation will be

$$Z_X = A + B X(k) + E_X, X = 1, \dots, \frac{n}{k} \quad (2.31)$$

The slope $B = \frac{\sum_{X=1}^{\frac{n}{k}} (Z_X - \bar{Z}_X)(X(k) - \bar{X}(k))}{\sum_{X=1}^{\frac{n}{k}} (X(k) - \bar{X}(k))^2}, X = 1, \dots, \frac{n}{k}$

and the intercept $A = \bar{Z} - B \bar{X}$

Hint: we used A and B rather than \hat{A} and \hat{B} since the detailed and aggregated equations are perfect, i.e error free

To study the effect of aggregation on the intercept, we need the following relations:

a - The relation between \bar{Y} , \bar{Z} :

The sum of all observations $Y_t = \sum_{t=1}^n Y_t$, also $= \sum_{t=1}^{n/k} Z_t$ since the vertical aggregation has no effect on the sum of all observations, then

$$\bar{Z} = \frac{\sum_{t=1}^{n/k} Z_t}{n/k} = \frac{\sum_{t=1}^n Y_t}{n/k} = k \frac{\sum_{t=1}^n Y_t}{n} = k\bar{Y} \quad (2.32)$$

b - The relation between the mean of detailed time \bar{T} and the mean of aggregated time \bar{X}

The detailed time T is an arithmetic progression with first term 1 and the last term n , its sum $= \frac{n}{2}(1+n)$ and its mean $\bar{T} = \frac{1}{2}(1+n)$

Where the time $X(k)$ aggregated on period k , is an arithmetic progression with first term 1 and the last term n/k , its sum $=$

$$\frac{n/k}{2} \left(1 + \frac{n}{k}\right) \text{ and its mean } \bar{X} = \frac{1}{2} \left(1 + \frac{n}{k}\right)$$

Now, the intercept of the aggregated equation is $A = \bar{Z} - B\bar{X}(k)$

Hint : $\bar{X} = \bar{T}$ when $k = 1$

$$A = k\bar{Y} - \beta k^2 \left[\frac{1}{2} \left(1 + \frac{n}{k}\right) \right] = k \left[\bar{Y} - \frac{1}{2} \beta (n+k) \right] \quad (2.33)$$

The mean of the detailed equation $\bar{Y} = 104.5$

Applying the rule (2.33) on equation (2.18) , k = 2 where A = 23

$$A = k \left[\bar{Y} - \frac{1}{2} \beta (n+k) \right] = 2 [104,5 - 0,5 (3) (62)] = 2 (104.5 - 93) = 23$$

Applying the rule (2.33) on equation (2.21) , k = 3 where A = 30

$$A = k \left[\bar{Y} - \frac{1}{2} \beta (n+k) \right] = 3 [104,5 - 0,5 (3) (63)] = 3 (104.5 - 94,5) = 30$$

Applying the rule (2.33) on equation (2.24) , k = 4 the intercept A = 34

$$A = k \left[\bar{Y} - \frac{1}{2} \beta (n+k) \right] = 4 [104,5 - 0,5 (3) (64)] = 4 (104.5 - 96) = 34$$

Applying the rule (2.33) on equation (2.27) , k = 5 the intercept A = 35

$$A = k \left[\bar{Y} - \frac{1}{2} \beta (n+k) \right] = 5 [104,5 - 0,5 (3) (65)] = 5 (104.5 - 96) = 35$$

Equations (2.19), (2.22), (2.25), and (2.28) and realization of the intercept above, prove the following Samar theorem:

Samar theorem

Parameters of aggregated time series

The parameters of aggregated time series depend on aggregation method as follows:

Part (1) : Horizontal aggregation

Given a number I of time series:

$$Y_{iT} = \alpha_j + \beta_i T + e_{iT} \quad T = 1,2, \dots,n, \quad i = 1, \dots,I$$

it can be aggregated horizontally as follows:

$$\sum_{i=1}^I Y_{iT} = A + BT + \sum_{i=1}^I e_{iT}$$

then the slope $B(I) = \sum_{i=1}^I \beta_i$ and intercept $A(I) = \sum_{i=1}^I \alpha_i$

do not imply any bias for any number of time series

Part (2) : Vertical aggregation:

Given a time series Y

$$Y_T = \alpha + \beta T + e_T \quad T = 1, 2, \dots, n$$

aggregated vertically on a time period k (positive number $< n$) as follows:

$$Z_X(k) = A(k) + B(k) X(k) + E_X(k)$$

where

$$Z_X(k) = \sum_{j=1}^k Y_{kX-k+j}, E_X(k) = \sum_{j=1}^k e_{kX-k+j}, X = 1, \dots, \frac{n}{k}, j = 1, \dots, k$$

Then the aggregated slope $B(k) = \beta k^2$ and

$$\text{Aggregated intercept } A(k) = k \left[\bar{Y} - \frac{1}{2} \beta (n+k) \right]$$

Condition that aggregation does not result in any bias.

Corollary (1)

If the condition of vertical aggregation of Samar is violated, then:

Aggregation bias in the slope is $S_\beta(k) = B(k) - k^2 \beta$ and

Aggregation bias in the intercept is $S_a(k) = A(k) - k \left[\bar{Y} - \frac{1}{2} \beta (n+k) \right]$

where: $B(k)$ is the slope of the aggregated series on period k

$A(k)$ is the intercept of the aggregated series on period k

\bar{Y} is the mean of detailed series

β is the slope of detailed series

n is the number of observations of detailed series

k is the aggregation period

Corollary (2)

Aggregation bias in the slope $S_\beta(\mathbf{k})$ results in aggregation bias in the intercept $S_\alpha(\mathbf{k})$ since

$$A(k) = \bar{Z}(k) - B(k) \bar{X}(k) \quad \text{where}$$

$\bar{Z}(k)$ is the mean of the aggregated observations

$\bar{X}(k)$ is the mean of the time sequence of the aggregated series

are fixed do not imply any bias

Corollary (3)

The bias $S_\beta(\mathbf{k})$ and $S_\alpha(\mathbf{k})$ stated in corollary (1) are functions of k

3 - Empirical study

Samar will be used to investigate, empirically, the bias in horizontal and vertical aggregation of time series. 11 series of 52 weekly income data of 11 farmers in a farm near by. Data were collected through the Egyptian agrarian year.2017/2018, (October 2017 - September 2018).

3.1 Model building

The model consists of 11 individual equations

$$Y_{iT} = \alpha_i + \beta_i T_i + e_{iT} \quad i = 1, \dots, 11 \quad T = 1, \dots, 52 \quad (3.1)$$

Where:

Y_{iT} weekly income of farmer i

α_i	intercept of equation i
β_i	slope of equation i
T_i	serial number of weeks 1,2,...,52
e_{iT}	random error of equation i

3.2 Data collection and investigation

The farmers whose data are collected work together in a farm near by Port Said. Most agrarian ownership in Egypt are small and medium farms, individually owned. Therefore, the Egyptian agriculture is half mechanized. So manual work is a must. Each farmer cares about a certain part of the cultivated area of the farm. He earns a proportion of the harvest against his manual work. This ratio varies from $\frac{1}{4}$ or $\frac{1}{3}$ or $\frac{1}{2}$, each ratio implies different obligations on the farmer.

In our case the farmer earns $\frac{1}{4}$ of the harvest, against all the manual work including removing weeds whether manually or using pesticides. The agrarian area is cultivated twice in the agrarian year summer crop, mainly rice, and winter crop mainly wheat.

Thus, income of farmer (1), for instance, consists of his share of farming 5 faddans (faddan is agrarian area unit = 4200 m²) as follows

Summer crop (rice) harvested by October 2017 calculated as follows:

$$5 \text{ faddans} \times 3 = 15 \text{ ton (average productivity 3 ton/faddan)}$$

$$\text{Share of the farmer} = 15 \div 4 = 3.75 \text{ ton}$$

Subtract 0.75 ton cost of pesticides

Net = 3 ton \times LE 3500 = LE 10,500 making LE 404 weekly average (fixed) income through the first half of the year (26 weeks) from October 2017 till March 2018

Winter crop (wheat) harvested by April 2018 calculated as follows:

$$5 \text{ faddans} \times 1.8 = 9 \text{ ton (average productivity 1.8 ton/faddan)}$$

$$\text{Share of the farmer} = 9 \div 4 = 2.25 \text{ ton}$$

Subtract 0.25 ton cost of pesticides

Samar Ahmed Helmy Abdelghany

Net = 2 ton \times LE 4 \times 50 = LE 4,100 in addition to LE 4000 his share in the value of the wheat's hay making LE 504 weekly average (fixed) income through the second half of the year (26 weeks) April / September 2018

In addition he works , randomly, for the others for LE 50 daily wage. Some of his family members may work daily besides him too, so a farmer may have more than 7 working days a week.

This system applies to all farmers with two differences:

- 1 - The area each farmer farms
- 2 - Number of daily wages earned each week

Data are collected under the direct supervision of the researcher presented in the appendix

Using SPSS-20 to test the data collected, the following table (1) shows that all Y's (except Y₃ & Y₄) are random variables. This means that 9 are non stationary, and the other two are stationary. This fact will be considered through the analysis

Table (2)
Randomness test of the series studied

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of income1 is normal with mean 377.02 and standard deviation 111.53.	One-Sample Kolmogorov-Smirnov Test	.209	Retain the null hypothesis.
2	The distribution of income2 is normal with mean 266.48 and standard deviation 70.50.	One-Sample Kolmogorov-Smirnov Test	.124	Retain the null hypothesis.
3	The distribution of income3 is normal with mean 532.69 and standard deviation 108.19.	One-Sample Kolmogorov-Smirnov Test	.017	Reject the null hypothesis.
4	The distribution of income4 is normal with mean 230.19 and standard deviation 84.60.	One-Sample Kolmogorov-Smirnov Test	.046	Reject the null hypothesis.
5	The distribution of income5 is normal with mean 352.62 and standard deviation 121.00.	One-Sample Kolmogorov-Smirnov Test	.071	Retain the null hypothesis.
6	The distribution of income6 is normal with mean 422.38 and standard deviation 126.07.	One-Sample Kolmogorov-Smirnov Test	.371	Retain the null hypothesis.
7	The distribution of income7 is normal with mean 352.35 and standard deviation 125.08.	One-Sample Kolmogorov-Smirnov Test	.918	Retain the null hypothesis.
8	The distribution of income8 is normal with mean 472.11 and standard deviation 116.72.	One-Sample Kolmogorov-Smirnov Test	.557	Retain the null hypothesis.
9	The distribution of income9 is normal with mean 542.00 and standard deviation 154.62.	One-Sample Kolmogorov-Smirnov Test	.889	Retain the null hypothesis.
10	The distribution of income10 is normal with mean 656.74 and standard deviation 139.53.	One-Sample Kolmogorov-Smirnov Test	.494	Retain the null hypothesis.
11	The distribution of income11 is normal with mean 626.72 and standard deviation 178.09.	One-Sample Kolmogorov-Smirnov Test	.428	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

3.1 Bias in the horizontal aggregation of the series

Trend of the 11 detailed series is calculated, and that of the horizontally aggregated series. Parameters of the aggregated trend are compared to the sum of the corresponding parameters to calculate the bias.

Working the data of farmer (1) as an example. Coefficients of the trend are shown in table (4). They are significant at 1% level, where calculated $t = 11.38$ against tabulated $t_{(51, 0.01)} \approx 2.692$. Significance is clear also from column Sig. of table (4) below. So, trend (1) is:

$$\hat{Y}_1 = 224,3 + 5,76 T \quad (3.2)$$

Table (3)
Coefficients of the weekly trend of series (1) of the empirical data
Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	224.274	19.708		11.380	.000
Time	5.764	.647	.783	8.907	.000

a. Dependent Variable: income1

The same procedure is followed for all series, the result is:

Trend (1)	$\hat{Y}_1 = 224,3 + 5,76 T$	$\bar{Y}_1 = 377.02$
Trend (2)	$\hat{Y}_2 = 184,4 + 3,10 T$	$\bar{Y}_2 = 266.48$
Trend (3)	$\hat{Y}_3 = 442,6 + 3,40 T$	$\bar{Y}_3 = 532.69$
Trend (4)	$\hat{Y}_4 = 141,8 + 3,34 T$	$\bar{Y}_4 = 230.19$
Trend (5)	$Y_5 = 201,6 + 5,70 T$	$\bar{Y}_5 = 352.62$
Trend (6)	$\hat{Y}_6 = 376,1 + 1,75 T$	$\bar{Y}_6 = 422.38$
Trend (7)	$\hat{Y}_7 = 359,3 (-) 0,26 T$	$\bar{Y}_7 = 352.35$

Trend (8)	$\hat{Y}_8 = 324,9 + 5,55 T \quad \bar{Y}_8 = 472.11$
Trend (9)	$\hat{Y}_9 = 339,1 + 7,66 T \quad \bar{Y}_9 = 542.00$
Trend (10)	$\hat{Y}_{10} = 448,9 + 7,84 T \quad \bar{Y}_{10} = 656.74$
Trend (11)	$\hat{Y}_{11} = 358,8 + 10,11 T \quad \bar{Y}_{11} = 626.72$
Sum of all trends	$\sum_{i=1}^{11} Y_i = 3401,8 + 53,95 T$
Aggregated trend	$\hat{Z} = 3401,8 + 53,95 T$

It is obvious that $\sum_{i=1}^{11} \hat{\beta}_i = 53.95 = \hat{B}$ and, $\sum_{i=1}^{11} \hat{\alpha}_i = 3401,8 = \hat{A}$ where \hat{A} assessing unbiasedness of the slope and the intercept. **So horizontal aggregation of time series does not involve any bias coping with the theoretical derivations in § 2.1**

3.2 Bias in the vertical aggregation of the series

The Samar theorem showed that bias in the slope S_β and that in the intercept S_α are $f(k)$. So, the available data, 52 observations, will be studied with $k = 2$ and $k = 4$ the only available values for k .

Data are aggregated bi-weekly, $k = 2$ and monthly, $k=4$. Necessary manipulations are done to compute $S_\beta(k)$ and $S_\alpha(k)$ in both cases of k .

Farmer (1)

Working the series of farmer (1) as an example, weekly trend shown in (3.2) was $\hat{Y}_1 = 224,3 + 5,76 T \quad , \quad \bar{Y}_1 = 377.02$

Bi-weekly trend, i.e $k = 2$

Table (4)
Coefficients of the bi-weekly trend of series (1) of the empirical data
Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	444.083	48.148		9.223	.000
1 زمن نصف	22.960	3.118	.833	7.364	.000

a. Dependent Variable: Bi-weekly (1)

Bi-weekly trend $\hat{H}_1 = 444,1 + 22,96 R$

where H Biweekly income

R Biweekly time 1,2,...,26

Monthly trend, i.e k=4

Table (5)
Coefficients of the monthly trend of series (1) of the empirical data
Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	874.077	120.011		7.283	.000
1 Time	90.571	15.120	.875	5.990	.000

a. Dependent Variable: ادخل شهري (1)

Monthly trend $\hat{Z}_1 = 874,1 + 90,57 X$

where Z Monthly income

R Monthly time 1,2,...,13

Calculating aggregation bias using Samar

According to Samar,

$$\text{aggregation bias in the slope } S_{\beta}(\mathbf{k}) = \hat{B} - k^2 \hat{\beta} \quad \text{and} \quad (3.3)$$

$$\text{aggregation bias in the intercept } S_{\alpha}(\mathbf{k}) = \hat{A} - k \left[\bar{Y} - \frac{1}{2} \beta (n+k) \right] \quad (3.4)$$

are computed with $k = 2$, and $k = 4$ and compare the results

Farmer (1) for $k = 2$

Use weekly and bi-weekly parameters and substitute in (3.3) to get:

$$S_{\beta}(\mathbf{2}) = 22.96 - (2)^2 5.76 = - 0.08 \quad = - 0.35\% \text{ of } \hat{B}$$

Use weekly and bi-weekly parameters and substitute in (3.4) to get:

$$S_{\alpha}(\mathbf{2}) = 444.1 - 2 \left[377.02 - \frac{1}{2} (5.76)(52+2) \right] = 1.1 \quad 0.25\% \text{ of } \hat{A}$$

For $k = 4$

Use weekly and monthly parameters and substitute in (3.4) to get:

$$S_{\beta}(\mathbf{4}) = 90.57 - (4)^2 5.76 = - 1.59 \quad = - 1.67\% \text{ of } \hat{B}$$

Use weekly and monthly parameters and substitute in (3.4) to get:

$$S_{\alpha}(\mathbf{4}) = 874.1 - 4 \left[377.02 - \frac{1}{2} (90.57)(52+2) \right] = 11.14 \quad = 1.27\% \text{ of } \hat{A}$$

The same procedure is applied to the remaining series to compute $S_{\beta}(\mathbf{k})$ and $S_{\alpha}(\mathbf{k})$, the results are shown in table (5) below

Table (6)

Aggregation bias in all series according to k

Series	Bias in the slope $S_{\beta}(k)$				Bias in the intercept $S_{\alpha}(k)$			
	k=2		k=4		k=2		k=4	
	amount	%	amount	%	amount	%	amount	%
(1)	- 0.08	- 0.35	- 1.59	- 1.67	1.1	0.25	11.14	1.27
(2)	- 0.05	- 0.4	0.19	0.38	0.67	0.18	- 1.32	- 0.18
(3)	0.11	0.82	- 1.38	- 2.60	1.53	0.17	9.64	0.55
(4)	- 0.07	- 0.52	0.13	0.24	1.11	0.35	- 0.88	- 0.16
(5)	0.02	0.09	0.54	0.59	- 0.28	- 0.07	-3.78	- 0.59
(6)	0.07	0.99	- 0.57	- 2.08	- 0.89	- 0.12	3.98	0.27
(7)	- 0.05	- 4.59	- 0.67	- 13.87	0.67	0.09	4.68	0.32
(8)	0.05	0.22	- 0.46	- 0.52	- 0.72	- 0.11	3.26	0.26
(9)	0.08	0.26	0.03	0.02	1.09	- 0.16	- 0.08	- 0.01
(10)	- 0.17	- 0.55	- 0.04	- 0.02	2.43	0.27	0.42	0.02
(11)	0.04	0.10	- 1.32	- 0.82	- 0.39	- 0.06	9.32	0.67
M	-0.005	-0.357	-0.467	-1.85	0.523	0.072	3.307	0.22
M	0.0718	0.8082	0.6291	2.0736	0.9891	0.1664	4.4091	0.3909

Results:

Two different means of the bias are calculated: the usual mean "M" and the mean of the absolute amounts of the bias |M|, call it free mean.

- 1 - The mean M of $S_{\beta}(2)$ i.e $\bar{S}_{\beta}(2) = - 0.005$ while $\bar{S}_{\beta}(4) = - 0.467$ showing the same direction of the bias, but concerning the amount of bias $|\bar{S}_{\beta}(4)| = |- 0.467| > |\bar{S}_{\beta}(2)| = |- 0.005|$
- 2 - The same argument applies to the mean bias of the intercept. M of $\bar{S}_{\alpha}(2) = 0.523$ while $\bar{S}_{\alpha}(4) = 3.307$ showing the same direction and amount of bias $\bar{S}_{\alpha}(4) > \bar{S}_{\alpha}(2)$
- 3 - The mean M of the relative bias in B and A leads to the same conclusions (1) and (2) above
- 4 - The free mean of the slope $f\bar{S}_{\beta}(4) = 0.6291 > f\bar{S}_{\beta}(2) = 0.0718$.
Also $f\bar{S}_{\alpha}(4) = 4.4091 > f\bar{S}_{\alpha}(2) = 0.9891$

- 5 - The free mean $|M|$ of the relative bias in B and A leads to the same conclusions (3) and (4) above
- 6 - The conclusions (1),..., (5) prove that as k increases, the bias, increases amount and relatively in the slope and the intercept, i.e $S_{\beta}(k)$ and $S_{\alpha}(k)$ are functions of (k) coping with corollary (3) of Samar
- 7 - At $k = 2$ the average bias $\bar{S}_{\beta}(2) = - 0.005$ in the slope induced average bias in the intercept $\bar{S}_{\alpha}(2)$ up to 0.523, at $k = 4$, the average bias in the slope $\bar{S}_{\alpha}(4)$ (increased) to - 0.467 moving the average bias in the intercept $\bar{S}_{\beta}(4)$ up to 3.307 . The same result is achieved if we work with the free means. This realizes corollary (2) of Samar
- 8 - The correlation between $S_{\beta}(2)$ and $S_{\beta}(4)$ $\text{Corr}(S_{\beta}(2), S_{\beta}(4)) = 1$ ensuring that the bias increases with k .
- 9 - $\text{Corr}(S_{\beta}(2), S_{\alpha}(2)) = - 0.69$ and $\text{Corr}(S_{\beta}(4), S_{\alpha}(4)) = - 0.9999$ showing very strong adverse relation as concluded in (7) above.

The adverse relation between $S_{\beta}(k)$ and $S_{\alpha}(k)$ is true because they are changes in the slope and the intercept respectively. Let:

$Y_T = \alpha + \beta T \quad T = 1, 2, \dots, n$ be aggregated on k and k_1 , then

$Z_X = \{A + S_{\alpha}(k)\} + \{B + S_{\beta}(k)\} X, \quad X = 1, 2, \dots, \frac{n}{k}$

$H_R = \{\delta + S_{\alpha}(k_1)\} + \{\theta + S_{\beta}(k_1)\} R, \quad R = 1, 2, \dots, \frac{n}{k_1}$

Since aggregation does not affect the total sum of the series then

$$\Sigma Y_T = \Sigma Z_X = \Sigma H_R$$

If $S_{\alpha}(k_1) > S_{\alpha}(k)$ then $S_{\beta}(k_1)$ should be less than $S_{\beta}(k)$ (at some value) and vice versa, to maintain the total sum of the series unchanged. If $S_{\beta}(k)$ and $S_{\alpha}(k)$ went the same direction, the total sum of the series changes. This explains the adverse relation between $S_{\beta}(k)$ and $S_{\alpha}(k)$

Theoretical derivations, and analysis of the case study above prove validity of the **Samar theorem**. Horizontal aggregation does not produce any bias in the parameters of the trend of the aggregated series. While

vertical aggregation implies aggregation bias as expressed in the corollaries of Samar theorem

References

- 1 - Albuquerque, Pedro H, 2003. "**A practical log-linear aggregation method with examples: heterogeneous income growth in the USA**,"Journal of Applied Econometrics, John Wiley & Sons, Ltd., vol. 18(6), pages 665-678.
- 2 .- Blundell, R. and Stoker, T. 2005. **Aggregation and heterogeneity**. Journal of Economic Literature 43, 347–91.
- 3 - Franta, Michal (2008) "**Time aggregation bias in discrete time models of aggregate duration data**" Czech National Bank, Prague
- 4 - Helmy, Samar (2021) "**Effect of estimation method on the aggregation bias if one of the regressors is stochastic**" (in Arabic) Scientific journal for financial and commercial studies and research, jssub@ekb.eg, July 2021
- 5 - Moorman, Jeanne E. (1979) **Aggregation Bias: An Empirical Demonstration**" Sociological Methods & Research, Vol. 8 No 1, August 1979 69-94 @ 1979 Sage Publications, Inc
- 6 - Robinson, W.S. (1950) "**Ecological Correlations and the Behavior of Individuals,**" American Sociological Review, Vol. 15, No. 3, 1950, pp. 351-357.
- 7 - Sbrana, Giacomo, (2011) "**Structural Time Series Models and Aggregation: Some Analytical Results**" Neoma Business School May 2011
- 8 - Theil, Henri (1954), **Linear Aggregation of Economic Relations**, North-Holland, Amsterdam.

Appendix (1)

Proof of part (2) of Samar theorem, i.e to show that, $\mathbf{B(k)} = \beta k^2$

Using the algebraic series $Y_T = 13 + 3 T$, $T = 1, \dots, 60$

For convenience, let $Z = Z(k)$ and $X = X(k)$ then the aggregated series is

$$Z_X = M + B X \quad X = 1, \dots, \frac{n}{k}$$

where $Z_X(k) = \sum_{j=1}^k Y_{kX-k+j}$, $X = 1, \dots, \frac{n}{k}$, $j = 1, \dots, k$

$$B = \frac{\sum_{X=1}^{\frac{n}{k}} (Z_X - \bar{Z})(X - \bar{X})}{\sum_{X=1}^{\frac{n}{k}} (X - \bar{X})^2} \quad (1)$$

Mathematically, if $T =$ natural numbers $1, \dots, n$ its sum $= \frac{n}{2} (1 + n)$. If aggregated on period $k = +ve$ number $< n$ yields the following arithmetic progression having the same sum

$$\sum_{j=1}^k T_j + Xk^2 , T = 1, \dots, n \quad X = 0, \dots, (\frac{n}{k} - 1) , j = 1, \dots, k \quad (2)$$

if $k = 2$, then

$$\text{first term} = \sum_{j=1}^2 T_j + Xk^2 = 1 + 2 + 0 (k^2) = 3$$

to get the last term put $X = (\frac{n}{k} - 1) = \frac{n}{2} - 1$

$$\text{last term} = \sum_{j=1}^2 T_j + Xk^2 = 3 + (\frac{n}{2} - 1) 4 = 3 + 2n - 4 = 2n - 1$$

sum of the progression = $\frac{n/2}{2} (3 + 2n - 1) = \frac{n/2}{2} (1 + n)$ which is the sum of original progression

if $k = 3$, then

$$\text{first term} = \sum_{j=1}^3 T_j + Xk^2 = 1 + 2 + 3 + 0(k^2) = 6$$

to get the last term put $X = (\frac{n/3}{k} - 1) = \frac{n/3}{3} - 1$

$$\text{last term} = \sum_{j=1}^3 T_j + Xk^2 = 6 + (\frac{n/3}{3} - 1) 9 = 6 + 3n - 9 = 3n - 3$$

sum of the progression = $\frac{n/3}{2} (6 + 3n - 3) = \frac{n/3}{2} (1 + n)$ which is the sum of original progression

The first term in progression (2) is constant, the term Xk^2 is an arithmetic progression, so it suits as regressor for the aggregated series on period k , for convenience, let $Z = Z(k)$ and $X = X(k)$ then the aggregated series will be $Z = A + \Delta(Xk^2)$

Now, regress it on the time $Xk^2 = k^2, 2k^2, \dots, (\frac{n}{k})k^2 = nk$

$$\text{then } \Delta = \frac{\sum_{x=1}^{\frac{n}{k}} (Z_x - \bar{Z})(Xk^2 - \bar{X}k^2)}{\sum_{x=1}^{\frac{n}{k}} (Xk^2 - \bar{X}k^2)^2} \quad (3)$$

for $k = 2$ the trend is: $Z = 23 + 3Xk^2$

Coefficients of the algebraic equation $Y_t = 13 + 3T$ aggregated
on period $k = 2$ and time Xk^2

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	23.000	.000		.	.
Time X k ²	3.000	.000	1.000	.	.

Dependent Variable: Y aggregated 2

the aggregated slope $\Delta = 3$

the intercept $A = \bar{Z} - \Delta (\bar{X}k^2) = k\bar{Y} - \Delta k^2 \left(\frac{1}{2} \left(1 + \frac{n}{k} \right) \right)$

substitute in the right hand side term $A = 2 (104.5) - 3 (4) (15.5) = 23$

notice that the aggregated slope $\Delta = 3$ the slope of the detailed equation and the intercept = 23 = the intercept of equation (2.15) regressed on time

$$X = 1, 2, \dots, 30 \left(\frac{n}{k} \right)$$

For $k = 3$ the trend is : $Z = 30 + 3 Xk^2$

the aggregated slope $\Delta = 3$ the slope of the detailed equation and the intercept = 30 = the intercept of equation (2.17) regressed on time

$$X = 1, 2, \dots, 20 \left(\frac{n}{k} \right)$$

For $k = 4$ the trend is : $Z = 34 + 3 Xk^2$

the aggregated slope $\Delta = 3$ the slope of the detailed equation and the intercept = 34 = the intercept of equation (2.19) regressed on time

$$X = 1, 2, \dots, 15 \left(\frac{n}{k} \right)$$

Which is the same result $\forall k < n$

Now,
$$\Delta = \frac{\sum_{x=1}^{n/k} (Z_x - \bar{Z})(X k^2 - \bar{X} k^2)}{\sum_{x=1}^{n/k} (X k^2 - \bar{X} k^2)^2} = \text{slope of the detailed equation } \beta$$

thus proving that $B = \Delta k^2$ proves that $B = \beta k^2$

from equation (1) above
$$B = \frac{\sum_{t=1}^{n/k} (Z_t - \bar{Z})(X_t - \bar{X})}{\sum_{t=1}^{n/k} (X_t - \bar{X})^2}$$

multiply the right side by $\frac{k^4}{k^4}$ it follows directly that

$$B = k^2 \frac{\sum_{x=1}^{n/k} (Z_x - \bar{Z})(X k^2 - \bar{X} k^2)}{\sum_{x=1}^{n/k} (X k^2 - \bar{X} k^2)^2} = k^2 \Delta = k^2 \beta \quad \text{Q.E.D}$$

Appendix (2)

Weekly income data of 11 farmers used in the empirical study

T	Y ₁	Y ₂	Y ₃	Y ₄	Y ₅	Y ₆	Y ₇	Y ₈	Y ₉	Y ₁₀	Y ₁₁
1	240	240	525	210	400	216	490	380	302	440	412
2	270	198	450	180	216	550	450	398	205	520	375
3	325	201	500	150	216	425	350	437	240	420	415
4	270	201	550	180	216	450	300	416	517	500	359
5	240	198	475	150	216	450	400	274	682	521	382
6	350	282	600	210	216	216	400	261	206	660	422
7	350	228	550	120	216	560	275	298	349	519	517
8	240	196	450	120	216	216	425	319	599	600	511
9	270	198	450	180	216	450	450	469	450	522	512
10	250	201	525	210	216	375	395	354	400	527	417
11	240	228	525	150	216	540	495	439	367	448	377
12	400	198	450	120	216	750	385	384	418	496	550
13	240	168	350	120	216	420	100	357	333	577	530
14	180	162	275	180	216	520	160	429	472	530	543
15	270	228	425	180	276	276	100	470	593	423	460
16	300	228	275	210	306	525	290	385	699	580	681
17	200	164	250	150	306	250	100	395	564	475	560
18	225	207	425	180	276	276	230	385	450	500	429
19	225	168	300	210	306	306	345	460	430	573	477
20	300	198	625	120	276	276	350	465	370	624	527
21	325	244	650	150	276	376	275	410	505	539	467
22	300	243	600	150	246	246	375	450	470	610	585
23	300	245	625	120	216	216	400	390	416	709	648
24	350	254	650	150	276	276	425	400	375	685	607
25	325	243	600	150	246	246	480	390	415	575	350
26	300	243	625	150	246	246	322	260	450	676	455
27	500	245	650	150	246	450	230	395	590	713	515
28	500	265	625	312	560	430	552	450	530	671	572
29	450	343	500	342	560	420	456	570	625	721	677
30	475	354	525	372	560	550	422	475	520	832	560
31	400	264	450	342	530	530	432	390	685	545	690
32	425	296	575	342	530	550	524	510	537	672	614

Samar Ahmed Helmy Abdelghany

T	Y ₁	Y ₂	Y ₃	Y ₄	Y ₅	Y ₆	Y ₇	Y ₈	Y ₉	Y ₁₀	Y ₁₁
33	450	275	450	372	530	540	658	630	607	627	841
34	400	343	500	342	560	476	540	540	658	712	795
35	520	272	450	372	530	480	500	475	596	829	815
36	520	379	650	348	435	435	260	630	658	779	821
37	375	275	450	318	435	750	100	705	687	816	720
38	576	354	500	378	405	405	368	690	618	825	747
39	450	403	625	288	405	405	486	720	559	542	939
40	576	365	650	318	375	375	290	680	570	846	813
41	476	403	600	288	435	435	290	450	675	745	868
42	450	375	625	288	435	435	325	485	670	758	904
43	450	375	625	288	450	650	268	575	771	803	911
44	576	386	600	318	450	450	298	560	577	800	819
45	475	403	600	288	435	460	268	615	537	912	617
46	450	376	650	288	435	435	268	510	783	855	738
47	475	275	625	288	450	450	320	525	823	768	850
48	576	275	625	318	435	460	150	590	686	855	877
49	450	252	625	210	430	430	375	480	672	849	845
50	450	244	600	210	460	460	300	520	825	789	750
51	425	248	600	180	400	450	425	625	780	815	850
52	450	248	600	210	400	400	450	680	669	825	876