Australian Journal of Basic and Applied Sciences, 3(4): 3836-3840, 2009
ISSN 1991-8178

A New Approach for Using Linear Coefficient Correlation and its Application for Investment in Stock Market

Mehdi Saeedikia

Dept.of Industrial Engineering, University of Tehran.

Abstract: A new averaging system has been utilized for the analysis of longitudinal (panel) data with few data points (short time series) that makes it a suitable method in a wide range of applications including the economic, and stock market data analysis. This process deals with data collection, recognition and their applications (e.g. distribution, qualitative or quantitative, longitudinal or cross-sectional nature of data) that utilizes weighting based on longitudinal nature of data and their correlations which are essential to any further statistical analysis. Utilization of this averaging process has rendered a superior performance compared to geometric and arithmetic averaging methods for a wide range of situations and applications. Simulation results confirm the validity and demonstrate advantages of the proposed averaging system.

Key words: Longitudinal Data, Stock market, Investment, Averaging, Data collection and recognition, Weighted Geometric Mean.

INTRODUCTION

Averaging as a measure of tendency towards the center of data is essential in collection and recognition of the nature of data. Recognition of distribution, qualitative vs. quantitative, longitudinal av. the cross-sectional (temporal) nature of data, invariably utilizes the averaging process as a measure. In economics, business, stock market, agriculture and many other fields we often deal with data of longitudinal nature and small sample size, yet often analyzed using the usual arithmetic or some geometric averaging processes as measures of their orientation towards the center.

The response variable, in longitudinal data, is measured repetitively over time for testing purposes (Diggle, 1996). Unlike the cross-sectional studies in which for each experiment one data set is measured, in longitudinal data analysis, such as in economics, several data sets are measured for each experiment.

The systematic development of panel analysis techniques started with the work of Lazarsfeld as the first large-scale panel study (Lazarsfeld, 1948) which explained the basic potentialities and problems of the panel design (Lazarsfeld, 1948; Lazarsfeld and Fisk, 1938; Rosenberg et al., 1951). Lazarsfeld (1948) compared the possibilities of panel studies with those of cross-sectional surveys and trend studies and concluded that the panel design is preferable in all respects. In the above-mentioned fields of application we are dealing with situations of scarce data for the utilization of time series analysis is not adequate. Thus a longitudinal data analysis is warranted.

The nature of the longitudinal data under consideration also includes the auto-correlation between the data entities, which is common in many fields of study. Some economic texts such as Salvatore (1984), and some economic reports, the "average growth" is commonly used as a measure, and by that, they often imply arithmetic mean, geometric mean, and weighted geometric mean (Yound et al., 1969).

This paper uses the weighted geometric method is mainly used for the comparison purposes and introduce weights that are drawn from the longitudinal concept. A simulated data set from the stock market was used to show the measure of success for firms on the stock exchange utilizing the proposed concept of averaging and the other methods currently in use.

Problem Definition:

For a given number of time periods (for example 5 years) of most recent data for a set of k companies on the exchange, with candidate stocks to purchase, data is collected on a specific economic index of interest to the buyer (e.g. profit to investment ratio). The following matrix summarizes this data:

Corresponding Author: Mehdi Saeedikia, Dept.of Industrial Engineering, University of Tehran.
E-mail: saeedikia@yahoo.com

Table 1: Profit to Investment ratio Matrix

	1	2	...	j	...	n
C_1	a_{11}	a_{12}	...	a_{1j}	...	a_{1n}
C_2	a_{21}	a_{22}	...	a_{2j}	...	a_{2n}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
C_i	a_{i1}	a_{i2}	...	a_{ij}	...	a_{in}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
C_k	a_{k1}	a_{k2}	...	a_{kj}	...	a_{kn}

where a_{ij} ,*(i=1,2,…,k and j=1,2,…,n),* is the profit to investment ratio for company C_i in year *j*.

If a sufficiently large volume of data on each company's performance were available, we could have used the time series models for the analysis of each company's case and choose the best alternative. Since quite often such is not the case, we define a proposed averaging process to take into account the data recentness and the trends for the longitudinal data as well as the defined autocorrelations, and compare its performance in application to scarce data situations, with the existing averaging processes already in use.

Definition:

Linear Correlation Coefficient:

The linear coefficient of correlation is generally used to measure the linear association between two variables. The linear correlation coefficient, ρ , between two random variables *X* and *Y* is defined as:

$$\rho = \frac{E[(X-\mu_X),(Y-\mu_Y)]}{\sigma_X \sigma_Y} \tag{1}$$

where E is the expected value operator, μ_X, σ_X and μ_Y, σ_Y are expected value and standard deviation of random variables *X* and *Y*, respectively. The sample linear correlation coefficient of *N* observations of random variables *X* and *Y* can be obtained by replacing μ_X and μ_Y with the sample mean $\bar{X}$ and $\bar{Y}$ and also σ_X and σ_Y with the sample standard deviations s_X and s_Y , respectively. Pearson's correlation coefficient, ρ, has the advantage of being a real-number easy to compute and to interpret. Let us define a column vector X_j for each j (Table 1) as follows:

$$X_j = (a_{1j}, a_{2j}, a_{3j}, \ldots, a_{ij})^T \tag{2}$$

j=1,2,….,n I s the time index and *i=1,2,….k* is the subject (company) index. If ρ_{n-j} ($0 \le j \le n$ is defined as the linear correlation coefficient between vectors X_j and X_n, then based on panel property of longitudinal data the following relation between the correlation coefficients of the 5 year data must hold; $\rho_4 \le \rho_3 \le \rho_2 \le \rho_1 \le \rho_0 = 1$

Geometric and Weighted Geometric Mean (WGM):

For a given data set $x_1, x_2, x_3, \ldots, x_n$ >0, let $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n > 0$ be the weights of the corresponding data items such that: $\lambda_1 + \lambda_2 + \ldots + \lambda_n = 1$ For each $t \varepsilon \mathbb{R}, t \neq 0$ and any given ordered pair (t, x_t) with t>0 the weighted mean of order t, namely , M_t is defined as

$$M_t = \left(\frac{\lambda_1 x_1^{\,t} + \lambda_2 x_2^{\,t} + \ldots + \lambda_n x_n^{\,t}}{n} \right)^{\frac{1}{t}} \tag{3}$$

For the following specific cases we define:

WAM = The Weighted Arithmetic Mean $= M_1 = \dfrac{\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_n x_n}{n}$ and

WHM = Weighted Harmonic Mean = $M_{-1} = \dfrac{n}{\dfrac{\lambda_1}{x_1} + \dfrac{\lambda_2}{x_2} + \ldots + \dfrac{\lambda_n}{x_n}}$, respectively.

Note that $\lim_{t \to 0} M_t = M_0 = x_1^{\lambda_1} x_2^{\lambda_2} \ldots x_n^{\lambda_n}$ which is called Weighted Geometric Mean (*WGM*). For the special case of $\lambda_1 = \lambda_2 = \lambda_3 = \ldots = \lambda_n = \dfrac{1}{n} > 0$ $M_0 = \sqrt[n]{\prod_{i=1}^{n} x_i}$ will be the simple geometric mean.

It is reminded that geometric mean is used as a suitable average for representing the mean of relative values such as quotients, percentages, indices of performance, profit-investment ratios and etc. Geometric mean is alternatively known as the growth mean.

The Proposed Averaging Method (MWGM):

Here we introduce the modified weighted geometric mean (*MWGM*) and derive its weights using correlation of the data over time:

$$MWGM = \sqrt[\sum_{j=0}^{j=n-1} \rho_j]{\prod_{j=1}^{j=n} a_{ij}^{\rho_{n-j}}} \quad (4)$$

where, n = number of years, ρ_j is linear correlation coefficient between year n and year $n\text{-}j$. In our problem introduced in Table 1 *MWGM* for every row of the matrix representing a company) we define MWGM as follows:

$$MWGM = \sqrt[\sum_{j=0}^{j=4} \rho_j]{\prod_{i=1}^{j=5} a_{kj}^{\rho_{5-j}}} \quad (5)$$

It is possible that some $a_{ij} \leq 1$. Therefore, a transformation that will proportionately make all $a_{ij} > 1$, is used which will additionally render the following:

1- Gives the more recent data larger weight, by assigning it a larger exponent $0 < \rho_j < 1$.
2- Allows using old data and trends in the computation of the mean growth.
3- Allows computing the mean growth in succeeding years.
4-Utilizes the mining of the data for the computation of the correlation between the data and utilizes it as a weight used in the MWGM computations.

RESULTS AND DISCUSSION

Results:

We need to remind the readers that the proposed averaging process is intended for situations that the data is longitudinal and the number of data for each company is “few”. In this case the *MWGM* with its weights obtained from correlation coefficients of data over time, has sufficient justification for use. In this paper we have considered only the positive correlations, arranged in descending order.

We raise each a_{ij} to the ρ_j exponent so that $a_{ij}^{\rho_j}$ takes on a value close to the original value (a_{ij} >1, 0< ρ_j <1). The simulation data generated is presented in Table 2 and takes into account all possible cases that may arise in data over the limited number of time periods (here we only consider 5 years) under consideration.

In the data grouping presented in Table 2, the ascending and descending order of data ,(representing positive and negative longitudinal growth), the ascending and descending jump in the data and their combination of orders and jumps have been considered which are explained accordingly in the resulting Table 2. Table 2 summarizes the result of this simulation in the order described in the analysis of the results. The aims of this simulation is the comparison of the performance of *MWGM* and *GM* for all possible cases conceivable based on the following values obtained from relations: $\rho_1 = 0.83, \rho_2 = 0.63, \rho_3 = 0.44, \rho_4 = 0.26$.

Table 2: The Simulated Data.

group	company	(i-4)th year	(i-3)th year	(i-2)th year	(i-1)th year	ith year	GM	MWGM
A	a1	1.05	1.1	1.15	1.17	1.200	1.13	1.1550
	a2	1.2	1.17	1.15	1.10	1.050	1.13	1.1109
B	b1	1.01	1.1	1.1	1.15	1.200	1.11	1.1360
	b2	1.1	1.01	1.1	1.15	1.200	1.11	1.1305
	b3	1.1	1.1	1.01	1.15	1.200	1.11	1.1247
	b4	1.1	1.1	1.15	1.01	1.200	1.11	1.1155
C	c1	1.3	1.1	1.1	1.10	1.100	1.14	1.1152
	c2	1.1	1.3	1.1	1.10	1.100	1.14	1.1259
	c3	1.1	1.1	1.3	1.10	1.100	1.14	1.1373
	c4	1.1	1.1	1.1	1.30	1.100	1.14	1.1493
D	d1	1.2	1.15	1.1	1.05	1.000	1.1	1.0685
	d2	1	1.05	1.1	1.15	1.200	1.1	1.1277
E	e1	1.1	1.15	1.2	1.10	1.050	1.12	1.1097
	e2	1.2	1.1	1.05	1.10	1.150	1.12	1.1132

Group A Companies:

Two companies a1 and a2 with company a1 following a growth pattern while company a2 following a falling pattern of performance. The system is to choose between the two companies based on "mean' of their performance. Based on geometric mean (*GM*) method the two companies have similar performance while *MWGM* method chooses a1 that has a growth pattern.

Group B Companies:

Often due to sudden changes in conditions a jump in data (upward or downward) may occur. Group B of the companies have data that simulates such conditions through b1, b2, b3, b4 companies. The geometric mean for all four companies are similar and thus no choice is forced. Whereas the *MWGM* method by stressing the recent patterns will give a more realistic result namely that b1>b2>b3>b4 and thus the above companies are preferred for investment in reverse order which is as expected.

Group C Companies:

Indicates similar results as the B group of companies, with the difference that these companies are undergoing increasing jumps in their latter years of performance.

Groups D & E Companies:

For situations of increasing and decreasing data patterns groups D and E are the cases for which the *MWGM* present a better performance measurement system.

Conclusion:

A new averaging process has been utilized for the analysis of longitudinal data with few data points that makes it a suitable method in a wide range of applications including the economic, and stock market data analysis. The new averaging system utilizes weighting based on longitudinal nature of date and their correlations which are essential to any further statistical analysis. The results show that the new averaging system has rendered a superior performance compared to geometric method which is useful for a wide range of situations and applications. In summary, we conclude that:

1-MWGM (modified weighted geometric mean) has the ability to distinguish the increasing and decreasing trend of observations, whereas, the geometric mean only focuses on the numerical value of the observation over time.

2- *MWGM* is sensitive to increasing and decreasing jumps and is a suitable criterion of performance, while geometric mean is not capable of presenting comparable performance.

REFERENCES

Lazarsfeld, P.F., 1948. The use of panel in social research. Proceeding of the American Philosophical Society, 92: 405-410.

Lazarsfeld, P.F. and M. Fiske, 1938. The panel as a new tool for measuring opinion. Public Opinion Quarterly, 2: 596-612.

Diggle, P.J., 1996. Analysis of longitudinal data, Oxford press.

Rosenberg, M., W. Thielens and P.F. Lazarsfeld, 1951. The Panel Study. In: Jahoda,Marie-Deutsch,M.-Cook,S.W. (eds.): Research Methods in Social Relations. New York, Dryden Press, pp: 587-609.

Salvatore, D., 1984. Schaum's outline of Theory and problems of Statistics and Economics. (Mcgraw-Hill Publ.Comp.).

Yound, W.E. and R.H. Trent, 1969. Geometric mean approximation. of individual security and portfolio performance; J. Finan. Quant. Anal., 4: 179–199.