Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
European Journal of Operational Research 162 (2005) 433–441
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/dsw

Stochastics and Statistics

Correlation and the time interval in multiple regression models

Rong Jea [a,d], Jin-Lung Lin [b,c], Chao-Ton Su [d,*]

[a] Department of Management Information Systems, Yuanpei University of Science and Technology, Hsinchu, Taiwan
[b] The Institute of Economics, Academia Sinica, Nankang, Taipei, Taiwan
[c] Department of Economics, National Chengchi University, Mucha, Taipei, Taiwan
[d] Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu, Taiwan

Received 24 March 2003; accepted 30 July 2003
Available online 13 December 2003

Abstract

In this paper we investigate the time interval effect of multiple regression models in which some of the variables are additive and some are multiplicative. The effect on the partial regression and correlation coefficients is influenced by the selected time interval. We find that the partial regression and correlation coefficients between two additive variables approach one-period values as *n* increases. When one of the variables is multiplicative, they will approach zero in the limit. We also show that the decreasing speed of the *n*-period correlation coefficients between both multiplicative variables is faster than others, except that a one-period correlation has a higher positive value. The results of this paper can be widely applied in various fields where regression or correlation analyses are employed.

Keywords: Correlation coefficient; Partial regression coefficient; Time interval

1. Introduction

In time series analysis of a given set of variables, practitioners often have to decide whether to use monthly, quarterly, or annual data. They usually try to use the time series data of the higher frequency in order to increase the number of observations. However, the data for such analyses are sometimes limited and available for different periodicities and different time spans. The standard approach is to change them to a common time interval through temporal aggregation or systematic sampling, depending on whether the variables are flow variables or stock variables respectively (Abeysinghe, 1998). This approach, apart from losing information, may defeat the purpose of using the association between variables so as to make a correct decision or to forecast a key variable of interest. Thus, we are concerned with the question of whether the regression and the correlation coefficients are affected by the selected time interval.

The effect of the differencing interval on several economic indices has been studied by Schneller (1975), Levhari and Levy (1977), Levy (1972, 1984), and Lee (1990). In addition, Bruno and

* Corresponding author. Address: Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu, Taiwan. Tel.: +886-3-5731857; fax: +886-3-5722392.
E-mail address: ctsu@cc.nctu.edu.tw (C.-T. Su).

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2003.07.020

Easterly (1998) explain that the inflation-growth correlation is only present with high frequency data and with extreme inflation observations. There is no cross-sectional correlation between long-run averages of growth and inflation. Souza and Smith (2002) show that decreasing the sampling rate will bias the estimation of the long memory parameter towards zero for all estimation methods. All these studies make it clear that the time interval cannot be selected arbitrarily.

Many studies employ some additive variables and some multiplicative simultaneously (Easton and Harris, 1991; Elton et al., 1995; Tang, 1992, 1996; Chance and Hemler, 2001; McAvinchey, 2003), but these are not our present concern. In general, flow variables and stock variables are additive (e.g., gross domestic product (GDP), industrial production, population, inventories, etc.). Examples multiplicative variables include the growth rates of GDP, industrial production, population, etc. Levy and Schwarz (1997) show that when two random variables are multiplicative over time, the coefficient of determination decreases monotonically as the differencing interval increases, approaching zero in the limit. Levy et al. (2001) write that when one of the variables is additive and the other is multiplicative, the squared multi-period correlation coefficient decreases monotonically as n increases and approaches zero when n goes to infinity. Thus far, we have seen the importance of analyzing the time interval effect on the regression coefficients when some of the variables are additive and some multiplicative.

The purpose of this paper is to complement and extend the results in Levy and Schwarz (1997) and Levy et al. (2001). Both studies consider the time interval effect when two random variables are additive or multiplicative. They use the correlation and the regression coefficient to demonstrate the importance of analyzing the time interval effect and provide us with a very good concept. However, using two random variables, we can only construct a simple regression model; that is, a model with a single regressor that has a relationship with a response. Unfortunately, very often we move to the situation with more than one independent variable such that the inferential possibilities increase more or less exponentially. Therefore, it always behooves the investigator to make the underlying rationale and the goals of the analysis as explicit as possible. For practical reasons we study the time interval effect by using the multiple-regression model that can be widely applied in many fields where regression or correlation analyses are employed.

The paper proceeds as follows. Section 2 briefly describes prior research and presents the numerical results with some discussion. Section 3 shows the time interval effect on the partial correlation and the regression coefficients in the multiple-regression model, and gives a numerical example corresponding to the US stock market. Section 4 offers concluding remarks.

2. The correlation coefficients between two random variables

Let $(Y_{11}, X_{11}, X_{21}), \ldots, (Y_{1n}, X_{1n}, X_{2n})$ and $(Y_{21}, X_{11}, X_{21}), \ldots, (Y_{2n}, X_{1n}, X_{2n})$ be sequences of independent, identically distributed variables. We define four new variables to denote an n-fold increase of the differencing interval two multiplicative and two additive variables.

The additive variables, denoted by $Y_1^{(n)}$ and $X_1^{(n)}$, are given by

$$Y_1^{(n)} = Y_{11} + Y_{12} + \cdots + Y_{1n}$$

and

$$X_1^{(n)} = X_{11} + X_{12} + \cdots + X_{1n}.$$

The multiplicative variables, denoted by $Y_2^{(n)}$ and $X_2^{(n)}$, are given by

$$Y_2^{(n)} = Y_{21} \cdot Y_{22} \cdots Y_{2n}$$

and

$$X_2^{(n)} = X_{21} \cdot X_{22} \cdots X_{2n}.$$

Using the above four variables, denoted by $Y_1^{(n)}$, $Y_2^{(n)}$, $X_1^{(n)}$, and $X_2^{(n)}$, we can study a few different cases depending on the types of variables and the number of independent variables in the regression models.

2.1. The additive–additive case

Using two random variables, we can construct a simple regression model. If the independent variable $X_1^{(n)}$ and the dependent variable $Y_1^{(n)}$ are both additive, then the regression coefficients corresponding to the model and the correlation coefficient between them will be unaffected by the selected time interval.

Proof. Let X_{1j} and Y_{1j} be identically independent distributed variables (i.i.d.), $j = 1, 2, \ldots, n$. We have

$$E(X_{1j}) = \mu_x, \quad \mathrm{Var}(X_{1j}) = \sigma_x^2,$$
$$E(Y_{1j}) = \mu_y \quad \text{and} \quad \mathrm{Var}(Y_{1j}) = \sigma_y^2.$$

The one-period correlation coefficient is

$$\rho_1 = \frac{\mathrm{Cov}(X_{1t}, Y_{1t})}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}.$$

Because $X_{11}, X_{12}, \ldots, X_{1n}$ are i.i.d., we have

$$E(X_1^{(n)}) = E\left(\sum_{j=1}^{n} X_{1j}\right) = \sum_{j=1}^{n} \mu_x = n\mu_x \tag{1}$$

and

$$\mathrm{Var}(X_1^{(n)}) = \mathrm{Var}\left(\sum_{j=1}^{n} X_{1j}\right) = \sum_{j=1}^{n} \sigma_x^2 = n\sigma_x^2. \tag{2}$$

Similarly, we can obtain

$$E(Y_1^{(n)}) = n\mu_y \tag{3}$$

and

$$\mathrm{Var}(Y_1^{(n)}) = n\sigma_y^2. \tag{4}$$

The n-period covariance is

$$\mathrm{Cov}(X_1^{(n)}, Y_1^{(n)}) = \mathrm{Cov}\left(\sum_{j=1}^{n} X_{1j}, \sum_{j=1}^{n} Y_{1j}\right) = n\,\mathrm{Cov}(X_{1t}, Y_{1t}) = n\sigma_{xy}. \tag{5}$$

Using Eq. (5), so the n-period correlation coefficient can be easily written as follows:

$$\rho_n = \frac{\mathrm{Cov}(X_1^{(n)}, Y_1^{(n)})}{\sigma_{x_1^{(n)}} \sigma_{y_1^{(n)}}} = \frac{n\sigma_{xy}}{\sqrt{n}\sigma_x \sqrt{n}\sigma_y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \rho_1. \tag{6}$$

Hence, the correlation coefficient between $X_1^{(n)}$ and $Y_1^{(n)}$ is independent of the differencing interval. Using Eq. (6) and the relationship between the correlation coefficient and the regression coefficient, we can easily obtain the same result. That is, the regression coefficient is also unaffected by the time interval employed.

2.2. The multiplicative–multiplicative case

Levy and Schwarz (1997) explain that when two random variables are multiplicative, their correlation coefficient will not be independent of the differencing interval even when each of the random variables is a product of i.i.d. variables over time. They show that unless $Y = kX, k > 0$, the coefficient of determination (ρ^2) decreases monotonically as the differencing interval increases, approaching zero in the limit.

2.3. The additive–multiplicative case

Levy et al. (2001) study the time interval effect when one of the variables is additive and one is multiplicative. They show that the squared multi-period correlation coefficient (ρ_n^2) monotonically decreases in n, and approaches zero when n goes to infinity.

2.3.1. Numerical example

Fig. 1 indicates the change of the correlation coefficients by the selected time interval in the

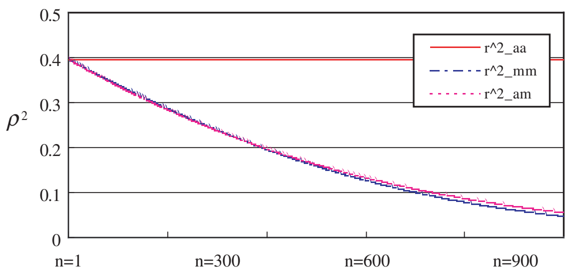

Fig. 1. The squared multi-period correlation coefficient.

above cases. The data used are the monthly rates of returns of IBM stock and the S&P500 index from January 1926 to December 1999. There are 888 observations and the returns include dividends (Tsay, 2002). As the figure indicates, the squared n-period correlation coefficient in the additive–additive case (i.e., r^2_aa) shows a horizontal line. That is, the correlation coefficient between two additive i.i.d. variables is independent of the differencing interval. The squared n-period correlation coefficient in the multiplicative–multiplicative case and additive–multiplicative case are denoted by r^2_mm and r^2_am, respectively. Fig. 1 reveals that they indeed decrease as n increases. Hence, the results of Levy and Schwarz (1997) and Levy et al. (2001) are demonstrated simultaneously.

Table 1 reveals the relationship between ρ_n and ρ_1. The corresponding parameters of the two return series are $E(X) \cong 1.0148$, $\sigma_x^2 \cong 0.0046$, $E(Y) \cong 1.0070$, $\sigma_y^2 \cong 0.0032$ and $\sigma_{xy} \cong 0.0024$, where $\{x_t\}$ and $\{y_t\}$ are the monthly rates of returns of IBM stock and the S&P500 index, respectively. Table 1 illustrates, for various values of a one-period correlation, that the squared correlation coefficients monotonically decrease as n increases. For example, if $\rho_1 = -1$ corresponding

Table 1
The multi-period correlation coefficient between additive or multiplicative variables

n	$\rho_1 = -1$		$\rho_1 = -0.6$		$\rho_1 = -0.1$		$\rho_1 = 0.2$	
	M&M	A&M	M&M	A&M	M&M	A&M	M&M	A&M
2	−0.9962	−0.9992	−0.5982	−0.5995	−0.0998	−0.0999	0.1997	0.1998
3	−0.9925	−0.9984	−0.5964	−0.5991	−0.0996	−0.0998	0.1994	0.1997
4	−0.9887	−0.9976	−0.5946	−0.5986	−0.0994	−0.0998	0.1991	0.1995
5	−0.9850	−0.9969	−0.5928	−0.5981	−0.0992	−0.0997	0.1988	0.1994
6	−0.9812	−0.9961	−0.5910	−0.5976	−0.0990	−0.0996	0.1985	0.1992
7	−0.9775	−0.9953	−0.5892	−0.5972	−0.0987	−0.0995	0.1982	0.1991
8	−0.9738	−0.9945	−0.5874	−0.5967	−0.0985	−0.0995	0.1979	0.1989
9	−0.9702	−0.9937	−0.5856	−0.5962	−0.0983	−0.0994	0.1976	0.1987
10	−0.9665	−0.9929	−0.5838	−0.5958	−0.0981	−0.0993	0.1973	0.1986
50	−0.8306	−0.9618	−0.5166	−0.5771	−0.0901	−0.0962	0.1853	0.1924
100	−0.6871	−0.9233	−0.4426	−0.5540	−0.0808	−0.0923	0.1709	0.1847
500	−0.1500	−0.6426	−0.1197	−0.3856	−0.0303	−0.0643	0.0807	0.1285
1000	−0.0223	−0.3771	−0.0204	−0.2263	−0.0071	−0.0377	0.0255	0.0754
5000	0.0000	−0.0015	0.0000	−0.0009	0.0000	−0.0002	0.0000	0.0003
10000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$\rho_1 = 0.4$		$\rho_1 = 0.5$		$\rho_1 = 0.8$		$\rho_1 = 1$	
2	0.3995	0.3997	0.4995	0.4996	0.7997	0.7994	1.0000	0.9992
3	0.3991	0.3994	0.4990	0.4992	0.7994	0.7987	0.9999	0.9984
4	0.3986	0.3991	0.4985	0.4988	0.7990	0.7981	0.9999	0.9976
5	0.3982	0.3987	0.4981	0.4984	0.7987	0.7975	0.9999	0.9969
6	0.3977	0.3984	0.4976	0.4980	0.7984	0.7969	0.9999	0.9961
7	0.3972	0.3981	0.4971	0.4976	0.7981	0.7962	0.9998	0.9953
8	0.3968	0.3978	0.4966	0.4973	0.7977	0.7956	0.9998	0.9945
9	0.3963	0.3975	0.4961	0.4969	0.7974	0.7950	0.9998	0.9937
10	0.3958	0.3972	0.4956	0.4965	0.7971	0.7944	0.9997	0.9929
50	0.3775	0.3847	0.4763	0.4809	0.7838	0.7694	0.9985	0.9618
100	0.3549	0.3693	0.4522	0.4617	0.7667	0.7386	0.9967	0.9233
500	0.1980	0.2571	0.2754	0.3213	0.6167	0.5141	0.9772	0.6426
1000	0.0795	0.1508	0.1261	0.1885	0.4356	0.3017	0.9465	0.3771
5000	0.0000	0.0006	0.0001	0.0008	0.0176	0.0012	0.7445	0.0015
10000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0000	0.5544	0.0000

to the data, then the correlation coefficient between both multiplicative variables is decreasing in n and approaches zero as $n = 5000$. Similarly, in the additive–multiplicative case, the correlation coefficient decreases to -0.0015 as $n = 5000$ and $\rho_1 = -1$. In addition to these, we also find that the correlation coefficient between two multiplicative time series decreases faster than other cases, except that the one-period correlation has a higher positive value.

As we can see, the correlation coefficient between both multiplicative variables (M&M) is relatively small for all n and $\rho_1 \leqslant 0.5$. On the other hand, the correlation coefficient is relatively large for $\rho_1 = 0.8$ and 1. The reduction in ρ_n is rather minor, particularly for $\rho_1 = 1$. However, this table tells us that the multi-period correlation $|\rho_n|$ indeed decreases as n increases even when $\rho_1 = -1$ or $\rho_1 = 1$. Hence, there is generally good evidence to show that the correlation decreases in n.

3. The partial regression and correlation coefficients in multiple regression models

From what has been mentioned above, we know the effect of the selected time interval when two random variables are additive or multiplicative. Here, we would like to focus on an extension to the multiple regression models. We may consider the subject under the following cases: (1) the dependent variable is additive; (2) the dependent variable is multiplicative.

3.1. The dependent variable is additive

In the multiple regression model, the dependent variable is additive and the regressors are composed of one additive and one multiplicative variable simultaneously. We can then construct the following n-period multiple regression model:

$$Y_1^{(n)} = \alpha_{0n} + \alpha_{1n} X_1^{(n)} + \alpha_{2n} X_2^{(n)} + \varepsilon, \tag{7}$$

where $Y_1^{(n)}$, $X_1^{(n)}$, and $X_2^{(n)}$ are as defined in Section 2. Terms α_{0n}, α_{1n}, and α_{2n} are the regression coefficients corresponding to the n-period multiple regression model. The error term ε is assumed to be normally and independently distributed. We additionally assume that the errors have mean zero and unknown variance σ^2.

Let

$$V_{1n} = \frac{Y_1^{(n)} - \overline{Y}_1^{(n)}}{\sqrt{n-1} S_{y_1^{(n)}}}, \quad U_{1n} = \frac{X_1^{(n)} - \overline{X}_1^{(n)}}{\sqrt{n-1} S_{x_1^{(n)}}} \quad \text{and} \quad U_{2n} = \frac{X_2^{(n)} - \overline{X}_2^{(n)}}{\sqrt{n-1} S_{x_2^{(n)}}}. \tag{8}$$

To apply the above suitable transformation, standardized variables, the regression model becomes

$$V_{1n} = \alpha_{0n}^* + \alpha_{1n}^* U_{1n} + \alpha_{2n}^* U_{2n} + \varepsilon, \tag{9}$$

where

$$\alpha_{0n}^* = 0, \quad \alpha_{1n}^* = \alpha_{1n} \frac{S_{x_1^{(n)}}}{S_{y_1^{(n)}}} \quad \text{and} \quad \alpha_{2n}^* = \alpha_{2n} \frac{S_{x_2^{(n)}}}{S_{y_1^{(n)}}}.$$

We can denote here

$$\boldsymbol{\alpha}_n^* = \begin{pmatrix} \alpha_{1n}^* \\ \alpha_{2n}^* \end{pmatrix} \quad \text{and} \quad \mathbf{U}_n = \begin{pmatrix} U_{1n} & U_{2n} \end{pmatrix}.$$

The least-squares estimator of $\boldsymbol{\alpha}_n^*$ can be expressed as

$$\hat{\boldsymbol{\alpha}}_n^* = (\mathbf{U}_n' \mathbf{U}_n)^{-1} (\mathbf{U}_n' \mathbf{V}_n) = \begin{bmatrix} 1 & r_{12}^{(n)} \\ r_{21}^{(n)} & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_{1y_1}^{(n)} \\ r_{2y_1}^{(n)} \end{bmatrix} = \begin{bmatrix} \frac{r_{1y_1}^{(n)} - r_{12}^{(n)} r_{2y_1}^{(n)}}{1 - (r_{12}^{(n)})^2} \\ \frac{r_{2y_1}^{(n)} - r_{21}^{(n)} r_{1y_1}^{(n)}}{1 - (r_{12}^{(n)})^2} \end{bmatrix}, \tag{10}$$

where $r_{ij}^{(n)}$ is the simple correlation between regressor $x_i^{(n)}$ and $x_j^{(n)}$ (see Neter et al., 1989, p. 290). Similarly, $r_{jy_1}^{(n)}$ is the simple correlation between the regressor $x_j^{(n)}$ and the response $y_1^{(n)}$.

Proposition 1. *Let $\hat{\alpha}_{1n}$ be the n-period partial regression coefficient of the regression as defined in* (7). *We obtain the following results*:

1. *As n approaches infinity*, $\lim_{n\to\infty} \hat{\alpha}_{1n} = \hat{\alpha}_{11}$ (*for the properties of the partial regression coefficient* $\hat{\alpha}_{2n}$, *see* Levy et al., 2001).
2. *If the regressor variables*, $X_1^{(n)}$ *and* $X_2^{(n)}$, *are independent*, *then* $\hat{\alpha}_{1n} = \hat{\alpha}_{11}$.

Proof. 1. Applying the results of Section 2.3 to Eq. (10), we know that $\lim_{n\to\infty} r_{2y_1}^{(n)} = 0$ and $\lim_{n\to\infty} r_{12}^{(n)} = \lim_{n\to\infty} r_{21}^{(n)} = 0$. Hence, as n approaches infinity, the standardized regression coefficients $\hat{\boldsymbol{\alpha}}_n^*$ can be obtained

$$\lim_{n\to\infty} \hat{\boldsymbol{\alpha}}_n^* = \begin{bmatrix} r_{1y_1}^{(n)} \\ 0 \end{bmatrix} = \begin{bmatrix} r_{1y_1}^{(1)} \\ 0 \end{bmatrix}, \tag{11}$$

where $r_{1y_1}^{(n)} = r_{1y_1}^{(1)}$ is shown in Section 2.1. Using the relationship between the original and standardized regression coefficients, we achieve

$$\hat{\alpha}_{jn} = \hat{\alpha}_{jn}^* \frac{S_{y_1^{(n)}}}{S_{x_j^{(n)}}}, \quad j = 1, 2 \tag{12}$$

and

$$\hat{\alpha}_{0n} = \bar{y}_1^{(n)} - \hat{\alpha}_{1n}\bar{x}_1^{(n)} - \hat{\alpha}_{2n}\bar{x}_2^{(n)}.$$

Using Eqs. (2), (4), (11) and (12), the n-period partial regression coefficient $\hat{\alpha}_{1n}$ is as follows:

$$\lim_{n\to\infty} \hat{\alpha}_{1n} = \lim_{n\to\infty} \hat{\alpha}_{1n}^* \frac{S_{y_1^{(n)}}}{S_{x_1^{(n)}}} = r_{1y_1}^{(1)} \frac{\sqrt{n}S_{y_1^{(1)}}}{\sqrt{n}S_{x_1^{(1)}}} = \hat{\alpha}_{11},$$

which completes the proof.

2. Because$X_1^{(n)}$ and $X_2^{(n)}$ are independent, it is obvious that $r_{12}^{(n)} = r_{21}^{(n)} = 0$. Similarly, using Eqs. (2), (4) and (12), we obtain $\hat{\alpha}_{1n} = \hat{\alpha}_{11}$. □

Proposition 2. *Let* $r_{y_1 1.2}^{(n)}$ *and* $r_{y_1 2.1}^{(n)}$ *be the partial correlation coefficients of the regression as defined in* (7). *Therefore,*

1. $\lim_{n\to\infty} r_{y_1 1.2}^{(n)} = r_{y_1 1}^{(1)}$ (*if* $X_1^{(n)}$ *and* $X_2^{(n)}$ *are independent, then* $r_{y_1 1.2}^{(n)} = r_{y_1 1}^{(1)}$).
2. $\lim_{n\to\infty} r_{y_1 2.1}^{(n)} = 0$.

Proof. 1. The partial correlation coefficient $r_{y_1 1.2}^{(n)}$ can be expressed by

$$r_{y_1 1.2}^{(n)} = \frac{r_{y_1 1}^{(n)} - r_{y_1 2}^{(n)} r_{12}^{(n)}}{\sqrt{1 - (r_{y_1 2}^{(n)})^2}\sqrt{1 - (r_{12}^{(n)})^2}}.$$

Because $\lim_{n\to\infty} r_{12}^{(n)} = 0$ and $\lim_{n\to\infty} r_{y_1 2}^{(n)} = 0$ (see Section 2.3), we achieve $\lim_{n\to\infty} r_{y_1 1.2}^{(n)} = r_{y_1 1}^{(n)}$. Using the relationship $r_{y_1 1}^{(n)} = r_{y_1 1}^{(1)}$ (see Section 2.1), we obtain $\lim_{n\to\infty} r_{y_1 1.2}^{(n)} = r_{y_1 1}^{(1)}$. In particular, if $X_1^{(n)}$ and $X_2^{(n)}$ are independent, then $r_{y_1 1.2}^{(n)} = r_{y_1 1}^{(n)} = r_{y_1 1}^{(1)}$.

2. The partial correlation coefficient $r_{y_1 2.1}^{(n)}$ can be expressed by

$$r_{y_1 2.1}^{(n)} = \frac{r_{y_1 2}^{(n)} - r_{y_1 1}^{(n)} r_{12}^{(n)}}{\sqrt{1 - (r_{y_1 1}^{(n)})^2}\sqrt{1 - (r_{12}^{(n)})^2}}.$$

Since $\lim_{n\to\infty} r_{12}^{(n)} = 0$ (see Section 2.3) and $r_{y_1 1}^{(n)} = r_{y_1 1}^{(1)}$ (see Section 2.1), we directly obtain that $\lim_{n\to\infty} r_{y2.1}^{(n)} = 0$, which completes the proof. □

3.2. The dependent variable is multiplicative

When the dependent variable is multiplicative, the regression model is as follows:

$$Y_2^{(n)} = \beta_{0n} + \beta_{1n}X_1^{(n)} + \beta_{2n}X_2^{(n)} + \varepsilon, \tag{13}$$

where $Y_2^{(n)}$, $X_1^{(n)}$, and $X_2^{(n)}$ are as defined in Section 2. Terms β_{0n}, β_{1n}, and β_{2n} are the regression coefficients corresponding to Eq. (13). Here, ε is a random error component.

We similarly let:

$$V_{2n} = \frac{Y_2^{(n)} - \overline{Y}_2^{(n)}}{\sqrt{n-1}S_{y_2^{(n)}}}, \quad U_{1n} = \frac{X_1^{(n)} - \overline{X}_1^{(n)}}{\sqrt{n-1}S_{x_1^{(n)}}} \quad \text{and}$$

$$U_{2n} = \frac{X_2^{(n)} - \overline{X}_2^{(n)}}{\sqrt{n-1}S_{x_2^{(n)}}}. \tag{14}$$

The regression model then becomes

$$V_{2n} = \beta_{0n}^* + \beta_{1n}^* U_{1n} + \beta_{2n}^* U_{2n} + \varepsilon, \tag{15}$$

where

$$\beta_{0n}^* = 0, \quad \beta_{1n}^* = \beta_{1n}\frac{S_{x_1^{(n)}}}{S_{y_2^{(n)}}}, \quad \text{and} \quad \beta_{2n}^* = \beta_{2n}\frac{S_{x_2^{(n)}}}{S_{y_2^{(n)}}}.$$

We can denote

$$\boldsymbol{\beta}_n^* = \begin{pmatrix} \beta_{1n}^* \\ \beta_{2n}^* \end{pmatrix} \quad \text{and} \quad \mathbf{U}_n = (\, U_{1n} \quad U_{2n} \,).$$

The least-squares estimator of $\boldsymbol{\beta}_n^*$ can therefore be expressed as

$$\hat{\boldsymbol{\beta}}_n^* = (U_n'U_n)^{-1}(U_n'V_n) = \begin{bmatrix} 1 & r_{12}^{(n)} \\ r_{21}^{(n)} & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_{1y_2}^{(n)} \\ r_{2y_2}^{(n)} \end{bmatrix} = \begin{bmatrix} \frac{r_{1y_2}^{(n)} - r_{12}^{(n)} r_{2y_2}^{(n)}}{1-(r_{12}^{(n)})^2} \\ \frac{r_{2y_2}^{(n)} - r_{21}^{(n)} r_{1y_2}^{(n)}}{1-(r_{12}^{(n)})^2} \end{bmatrix}.$$

Proposition 3. *Let* $\hat{\beta}_{2n}$ *be the n-period partial regression coefficient of the regression as defined in* (13). *As n approaches infinity*, $\lim_{n\to\infty} \hat{\beta}_{2n} = 0$ (*for the properties of the partial regression coefficient* $\hat{\beta}_{1n}$, *see* Levy et al., 2001).

The proof for Proposition 3 appears in Appendix A.

Proposition 4. *Let* $r_{y_2 1.2}^{(n)}$ *and* $r_{y_2 2.1}^{(n)}$ *be the partial correlation coefficients of the regression as defined in* (13). *Therefore*

1. $\lim_{n\to\infty} r_{y_2 1.2}^{(n)} = 0$.
2. $\lim_{n\to\infty} r_{y_2 2.1}^{(n)} = 0$.

Proof

1. The partial correlation coefficient $r_{y_2 1.2}^{(n)}$ can be expressed by

Table 2
The multi-period partial regression and correlation coefficients

n	Corr (n) A&A (1)	Corr (n) M&M (2)	Corr (n) A&M (3)	$\hat{\alpha}_{1n}$ (4)	$r_{y_1 1.2}^{(n)}$ (5)	$r_{y_1 2.1}^{(n)}$ (6)	$\hat{\beta}_{2n}$ (7)	$r_{y_2 1.2}^{(n)}$ (8)	$r_{y_2 2.1}^{(n)}$ (9)
1	0.62969	0.62969	0.62969	0.32102	0.38639	0.38639	0.52720	0.38639	0.38639
2	0.62969	0.62923	0.62920	0.32154	0.38702	0.38588	0.52664	0.38618	0.38626
3	0.62969	0.62878	0.62870	0.32206	0.38765	0.38538	0.52608	0.38596	0.38614
4	0.62969	0.62832	0.62821	0.32259	0.38828	0.38488	0.52553	0.38575	0.38602
5	0.62969	0.62787	0.62771	0.32311	0.38890	0.38438	0.52497	0.38554	0.38589
6	0.62969	0.62741	0.62722	0.32363	0.38953	0.38388	0.52441	0.38533	0.38577
7	0.62969	0.62695	0.62672	0.32414	0.39015	0.38338	0.52385	0.38512	0.38564
8	0.62969	0.62650	0.62623	0.32466	0.39077	0.38289	0.52329	0.38491	0.38552
9	0.62969	0.62604	0.62574	0.32517	0.39139	0.38239	0.52273	0.38470	0.38539
10	0.62969	0.62558	0.62524	0.32569	0.39201	0.38190	0.52218	0.38449	0.38526
11	0.62969	0.62512	0.62475	0.32620	0.39262	0.38140	0.52162	0.38428	0.38513
12	0.62969	0.62467	0.62426	0.32671	0.39324	0.38091	0.52106	0.38408	0.38501
13	0.62969	0.62421	0.62376	0.32721	0.39385	0.38042	0.52050	0.38387	0.38488
14	0.62969	0.62375	0.62327	0.32772	0.39446	0.37992	0.51994	0.38366	0.38475
15	0.62969	0.62329	0.62278	0.32823	0.39506	0.37943	0.51938	0.38345	0.38462
20	0.62969	0.62100	0.62031	0.33073	0.39808	0.37699	0.51659	0.38241	0.38395
25	0.62969	0.61871	0.61786	0.33319	0.40104	0.37457	0.51380	0.38137	0.38327
50	0.62969	0.60718	0.60561	0.34496	0.41521	0.36278	0.49983	0.37625	0.37966
75	0.62969	0.59558	0.59346	0.35589	0.42836	0.35148	0.48588	0.37122	0.37570
100	0.62969	0.58391	0.58140	0.36605	0.44060	0.34063	0.47196	0.36625	0.37144
500	0.62969	0.39989	0.40466	0.46291	0.55718	0.21094	0.26955	0.28973	0.28238
1000	0.62969	0.21934	0.23745	0.50477	0.60756	0.11652	0.11059	0.19559	0.17269
5000	0.62969	0.00072	0.00097	0.52315	0.62969	0.00046	0.00003	0.00097	0.00072
10000	0.62969	0.00000	0.00000	0.52315	0.62969	0.00000	0.00000	0.00000	0.00000

(1) The correlation coefficient in the additive–additive case.
(2) The correlation coefficient in the multiplicative–multiplicative case.
(3) The correlation coefficient in the additive–multiplicative case.
(4) The partial regression coefficient as defined in Proposition 1.
(5) The partial correlation coefficient as defined in Proposition 2.
(6) The partial correlation coefficient as defined in Proposition 2.
(7) The partial regression coefficient as defined in Proposition 3.
(8) The partial correlation coefficient as defined in Proposition 4.
(9) The partial correlation coefficient as defined in Proposition 4.

$$r^{(n)}_{y_2 1.2} = \frac{r^{(n)}_{y_2 1} - r^{(n)}_{y_2 2} r^{(n)}_{12}}{\sqrt{1-(r^{(n)}_{y_2 2})^2}\sqrt{1-(r^{(n)}_{12})^2}}.$$

Since $\lim_{n\to\infty} r^{(n)}_{12} = 0$ and $\lim_{n\to\infty} r^{(n)}_{y_2 1} = 0$ (see Section 2.3), and $\lim_{n\to\infty} r^{(n)}_{y_2 2} = 0$ (see Section 2.2), we obtain that $\lim_{n\to\infty} r^{(n)}_{y_2 1.2} = 0$.

2. The partial correlation coefficient $r^{(n)}_{y_2 2.1}$ can be expressed by

$$r^{(n)}_{y_2 2.1} = \frac{r^{(n)}_{y_2 2} - r^{(n)}_{y_2 1} r^{(n)}_{12}}{\sqrt{1-(r^{(n)}_{y_2 1})^2}\sqrt{1-(r^{(n)}_{12})^2}}.$$

Similarly, because $\lim_{n\to\infty} r^{(n)}_{12} = 0$, $\lim_{n\to\infty} r^{(n)}_{y_2 1} = 0$ (see Section 2.3) and $\lim_{n\to\infty} r^{(n)}_{y_2 2} = 0$ (see Section 2.2), we obtain that $\lim_{n\to\infty} r^{(n)}_{y_2 1.2} = 0$, which completes the proof. □

3.2.1. Numerical example

Table 2 illustrates the effect of the selected time interval on the partial regression and correlation coefficients in the multiple regression models corresponding to the US stock market. We use the monthly rates of returns of IBM stock and the S&P500 index shown in Table 2 as a numerical example. The sample period is from January 1926 to December 1999. In Table 2, three of the correlation coefficients (Columns (1)–(3)), depending on the additive or multiplicative variables, seem to be helpful in attempting to sketch out the association between variables in the multiple regression models. Using three distinct kinds of correlation coefficients corresponding to the two return series (corresponding to $E(X) \cong 1.0148$, $\sigma_x^2 \cong 0.0046$, $E(Y) \cong 1.0070$, $\sigma_y^2 \cong 0.0032$ and $\sigma_{xy} \cong 0.0024$), the other parameters (Table 2, Columns (4)–(9)) can be easily obtained.

To begin with, we claim that $\lim_{n\to\infty} \hat{\alpha}_{1n} = \hat{\alpha}_{11}$ in Proposition 1 where the dependent variable is additive. Column (4) of Table 2 reveals that $\hat{\alpha}_{1n}$ becomes closer to $\hat{\alpha}_{11}$ (=0.52315) as n increases and $\hat{\alpha}_{1n} = 0.52315$ (i.e., $\hat{\alpha}_{1n} = \hat{\alpha}_{11}$) as $n = 5000$. Therefore, $r^{(n)}_{y_1 1.2}$ approaches $r^{(1)}_{y_1 1}$ and $r^{(n)}_{y_1 2.1}$ approaches zero as n increases (see Columns (5) and (6)). The results also conform with the claim of Proposition 2. Finally, we turn to the case where the dependent variable is multiplicative. Column (7) indicates that $\hat{\beta}_{2n}$ approaches zero and decreases monotonically as n increases. This seems reasonable to support the claim of Proposition 3. The claim of Proposition 4 is shown in Columns (8) and (9).

4. Concluding remarks

We usually use a regression model to express the relationship between a variable of interest (the dependent variable) and a set of related independent variables. The association between variables is often measured by regression and correlation coefficients. The time interval of the data for such analyses cannot be selected arbitrarily. When two random variables are additive or multiplicative, the effect of the time interval employed is well documented in the literature.

In this paper we study the multiple linear regression models with two independent variables, where one of the variables is additive and the other variable is multiplicative. The dependent variable corresponding to the models is either additive or multiplicative. We show that the partial regression and correlation coefficients are affected by the selected time interval. When two variables are both additive, the partial regression and correlation coefficients between them approach one-period values as n goes to infinity. When one of the variables is multiplicative, they approach zero as n increases. The longer time intervals will decrease the relevant association between variables, particularly for the multiplicative dependent variable. We should not overlook these phenomena in such empirical analyses or it might lead to making incorrect decisions and misguided actions. The power of the test for the correlation is also influenced by the differencing interval. We also find that the decreasing speed of the n-period correlation coefficients between both multiplicative variables is faster than others, except that the one-period correlation has a higher positive value. This subject in the case deserves more than a passing notice.

The results of this paper relate to a multiple regression analysis, which is one of the most widely used techniques for analyzing multifactor data. Its broad appeal and usefulness are applied to studies conducted in various fields where variables are additive or multiplicative over time.

Appendix A. Proof of Proposition 3

Here we demonstrate the results of Proposition 3 from Levy and Schwarz (1997). Substituting the variable B (see Levy and Schwarz (1997) Eq. (1), p. 343) with the variable A, we get

$$\rho_n = \frac{C^n - 1}{\sqrt{(A^n - 1)}\sqrt{(A^n - 1)}}.$$

Using the above substitution, ρ_n can be regarded as the regression coefficient between two multiplicative variables. Hence, the results are obtained directly from Levy and Schwarz (1997).

References

Abeysinghe, T., 1998. Forecasting Singapore's quarterly GDP with monthly external trade. International Journal of Forecasting 14, 505–513.

Bruno, M., Easterly, W., 1998. Inflation crises and long-run growth. Journal of Monetary Economics 41, 3–26.

Chance, D.M., Hemler, M.L., 2001. The performance of professional market timers: Daily evidence from executed strategies. Journal of Financial Economics 62, 377–411.

Easton, P.D., Harris, T.S., 1991. Earnings as an explanatory variable for returns. Journal of Accounting Research 29 (1), 19–36.

Elton, E.J., Gruber, M.J., Blake, C.R., 1995. Fundamental economic variables, expected returns, bond fund performance. Journal Finance L (4), 1229–1256.

Lee, Y.W., 1990. Diversification and time: Do investment horizons matter? Journal of Portfolio Management 16 (3), 21–26.

Levhari, D., Levy, H., 1977. The capital asset pricing model and the investment horizon. Review of Economics and Statistics 59, 92–104.

Levy, H., 1972. Portfolio performance and the investment horizon. Management Science 18, 645–653.

Levy, H., 1984. Measuring risk and performance over alternative investment horizons. Financial Analysts Journal 40 (2), 61–62.

Levy, H., Schwarz, G., 1997. Correlation and the time interval over which the variables are measured. Journal of Econometrics 76, 341–350.

Levy, H., Guttman, I., Tkatch, I., 2001. Regression, correlation, and the time interval: Additive–multiplicative framework. Management Science 47 (8), 1150–1159.

McAvinchey, I.D., 2003. Modelling and forecasting in an energy demand system with high and low frequency information. Economic Modelling 20, 207–226.

Neter, J., Wasserman, W., Kutner, M.H., 1989. Applied Linear Regression Models, second ed. Irwin Inc.

Schneller, I.M., 1975. Regression analysis for multiplicative phenomena and its implication for the measurement of investment risk. Management Science 22, 422–426.

Souza, L.R., Smith, J., 2002. Bias in the memory parameter for different sampling rates. International Journal of Forecasting 18, 299–313.

Tang, G.Y.N., 1992. Diversification and intervalling effects on stock returns. Asia Pacific Journal of Management 9 (2), 231–241.

Tang, G.Y.N., 1996. Impact of investment horizon on currency portfolio diversification. International Business Review 5 (1), 99–116.

Tsay, R.S., 2002. Analysis of Financial Time Series. John Wiley & Sons, Inc.