Journal of the Operational Research Society (2005) 56, 1303–1309 0160-5682/05 $30.00

www.palgrave-journals.com/jors

Time aggregation effect on the correlation coefficient: added-systematically sampled framework

R Jea[1], C-T Su[2]* and J-L Lin[3,4]

[1]*National Chiao Tung University, Hsinchu, Taiwan;* [2]*National Tsing Hua University, Hsinchu, Taiwan;* [3]*The Institute of Economics, Academia Sinica, Nankang, Taipei, Taiwan; and* [4]*National Chengchi University, Mucha, Taipei, Taiwan*

The aggregation of financial and economic time series occurs in a number of ways. Temporal aggregation or systematic sampling is the commonly used approach. In this paper, we investigate the time interval effect of multiple regression models in which the variables are additive or systematically sampled. The correlation coefficient changes with the selected time interval when one is additive and the other is systematically sampled. It is shown that the squared correlation coefficient decreases monotonically as the differencing interval increases, approaching zero in the limit. When two random variables are both added or systematically sampled, the correlation coefficient is invariant with time and equal to the one-period values. We find that the partial regression and correlation coefficients between two additive or systematically sampled variables approach one-period values as n increases. When one of the variables is systematically sampled, they will approach zero in the limit. The time interval for the association analyses between variables is not selected arbitrarily or the statistical results are likely affected.

Journal of the Operational Research Society (2005) **56,** 1303–1309. doi:10.1057/palgrave.jors.2601947
Published online 2 March 2005

Keywords: correlation coefficient; time interval; systematic sampling

Introduction

Time-series data of different frequencies and different time spans are often available in empirical studies. They are usually changed to a common time interval through temporal aggregation or systematic sampling, depending on whether the variables are flow variables or stock variables, respectively. Several papers have documented the fact that time aggregation potentially distorts the relationship between variables.[1–5]

According to modern portfolio theory, the point about the gains from international portfolio diversification is inversely related to the correlations of security returns. Diversification benefits depend upon the correlations among different stock markets. Nevertheless, the employed time interval may affect the results if the data are autocorrelated. Even if all random variables are independent over time, the effect is seldom invariant with time. The effect of the differencing interval on several economic indices and finance has been studied by Schneller,[6] Levhari and Levy,[7] Levy,[8,9] Lee,[10] Bruno and Easterly,[11] and Souza and Smith.[12] Levy and Schwarz[13] show that the correlation coefficients are affected by the frequency of data employed when two independent, identically distributed (i.i.d.) random variables are multiplicative over time. Levy *et al*[14] show a similar theoretical effect when one of the i.i.d. variables is additive and the other is multiplicative. Jea *et al*[15] complement and extend the results in Levy and Schwarz[13] and Levy *et al*[14] to the multiple regression model. The recent literature demonstrates the importance of analysing the time interval effect on the association between variables. Therefore, if we select arbitrarily the time interval and neglect its impact, it is likely to lead us to misguided actions.

In this paper, we investigate the influence of the selected time interval on the association between i.i.d. variables over time when one of the variables is additive (eg industrial production, gross domestic product (GDP), etc.) and one is from systematic sampling (eg stock price, money supply, interest rate, etc.). To the best of our knowledge, the implication of the time horizon on the correlation between variables in such cases has not been investigated. Levy and Schwarz[13] and Levy *et al*[14] consider the time interval effect when two random variables are additive or multiplicative. However, the multiplicative variables can be converted to additive variables by taking logarithms. The simplification is not merely in reducing multiplication to addition, but more in modelling the statistical behaviour of some variables (eg, asset returns) over time. It is far easier to derive the time series

*Correspondence: C-T Su, Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu, Taiwan, ROC.
E-mail: ctsu@mx.nthu.edu.tw

properties of additive processes than of multiplicative processes.[16] Systematic sampling is widely applied in many fields, because it creates the best sample coverage over the population region, thus ensuring that sampling units are well spread over the population region. This paper complements the results in Levy and Schwarz[13] and Levy *et al*,[14] and uses the results of Jea *et al*[15] to extend to multiple regression model.

The paper proceeds as follows. The following section briefly describes the impact of the employed time interval on the various correlation coefficients and presents the numerical example corresponding to the US stock market. The subsequent section shows the time interval effect on the partial correlation and the regression coefficients in the multiple-regression model. The last section offers concluding remarks.

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 i.i.d. distributed variables. We define four new variables to denote an n-fold increase in the differencing interval's two additive and two systematically sampled 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 systematically sampled variables, denoted by $Y_2^{(n)}$ and $X_2^{(n)}$, are given by

$$Y_2^{(n)} = Y_{2k}, \quad 1 \leqslant k \leqslant n$$

and

$$X_2^{(n)} = X_{2h}, \quad 1 \leqslant h \leqslant n.$$

Using the above four variables, denoted by $Y_1^{(n)}$, $Y_2^{(n)}$, $X_1^{(n)}$, and $X_2^{(n)}$, we can study the properties of the correlation coefficient between them.

Both are additive

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 are unaffected by the selected time interval (the proof appears in Appendix).

Both are systematically sampled

Let $X_2^{(n)}$ and $Y_2^{(n)}$ be the systematically sampled variables. Because $(X_{21}, Y_{21}), (X_{22}, Y_{22}),\ldots,(X_{2n}, Y_{2n})$ is a sequence of distributed i.i.d. pairs of variables, the n-period expected values of $X_2^{(n)}$ and $Y_2^{(n)}$, respectively, are

$$E(X_2^{(n)}) = \mu_{x_2} \quad \text{and} \quad E(Y_2^{(n)}) = \mu_{y_2} \tag{1}$$

The n-period variances are denoted by

$$Var(X_2^{(n)}) = \sigma_{x_2}^2 \quad \text{and} \quad Var(Y_2^{(n)}) = \sigma_{y_2}^2 \tag{2}$$

The n-period covariance and correlation coefficient are given, respectively, by

$$Cov(X_2^{(n)}, Y_2^{(n)}) = \sigma_{x_2y_2} \tag{3}$$

and

$$\rho_{x_2y_2}^{(n)} = \frac{\sigma_{x_2y_2}}{\sigma_{x_2}\sigma_{y_2}} \tag{4}$$

Equations (1)–(4) provide the fundamental statistics of the systematically sampled variables $X_2^{(n)}$ and $Y_2^{(n)}$, respectively. These results do not depend on the number of periods. It is also easy to show that the correlation and regression coefficients of $X_2^{(n)}$ and $Y_2^{(n)}$ are unaffected by the selected time interval.

One is additive, the other is systematically sampled

Term $Y_1^{(n)}$ is an additive random variable and $X_2^{(n)}$ is a random variable from systematic sampling. Let the n-period expected value and variance of $Y_1^{(n)}$, respectively, be

$$\begin{aligned} E(Y_1^{(n)}) &= E\left(\sum_{j=1}^{n} Y_{1j}\right) = n\mu_{y_1} \quad \text{and} \\ Var(Y_1^{(n)}) &= Var\left(\sum_{j=1}^{n} Y_{1j}\right) = n\sigma_{y_1}^2 \end{aligned} \tag{5}$$

Because $(X_{21}, Y_{11}), (X_{22}, Y_{12}),\ldots,(X_{2n}, Y_{1n})$ is a sequence of i.i.d. variables, we obtain

$$Cov(X_2^{(n)}, Y_1^{(n)}) = Cov\left(X_{2h}, \sum_{j=1}^{n} Y_{1j}\right) = \sigma_{x_2y_1} \tag{6}$$

The n-period correlation coefficient is as follows:

$$\begin{aligned} \rho_{x_2y_1}^{(n)} &= \frac{Cov(X_2^{(n)}, Y_1^{(n)})}{\sqrt{Var(X_2^{(n)})}\sqrt{Var(Y_1^{(n)})}} = \frac{\sigma_{x_2y_1}}{\sigma_{x_2}\sqrt{n}\sigma_{y_1}} \\ &= \frac{1}{\sqrt{n}}\rho_{x_2y_1}^{(1)} \end{aligned} \tag{7}$$

Proposition 1 *Let $\rho_{x_2y_1}^{(n)}$ be the n-period correlation coefficient as defined in Equation (7). We obtain the following results*:

1. $(\rho_{x_2y_1}^{(n)})^2$ *is monotonically decreasing in n.*
2. *As* n *approaches infinity*, $\lim_{n\to\infty} \rho_{x_2y_1}^{(n)} = 0$.

Proof Using Equation (7), we can directly obtain these results. □

Numerical example The data used in this numerical example are obtained from the Center for Research in Security Prices (CRSP) of the University of Chicago. Dividend payments, if any, are included in the returns. Two financial time series are the daily simple returns of the S&P 500 index and American Express stock from January 1990 to December 1999 for 2528 observations.[17] Herein, assume the one-period correlation coefficients of the above three cases are equal to 0.5828. Figure 1 shows the behaviour of the squared correlation coefficients corresponding to the selected time interval. When two random variables are both additive or systematically sampled, as we can see in Figure 1, the squared n-period correlation coefficients in the two cases (ie, ρ^2_{aa} and ρ^2_{ss}) are invariant with time interval. The squared n-period correlation coefficient between one additive and the other systematically sampled is denoted by ρ^2_{as}. Figure 1 reveals that ρ^2_{as} decreases monotonically as n increases, in the situation when one is additive and the other is systematically sampled.

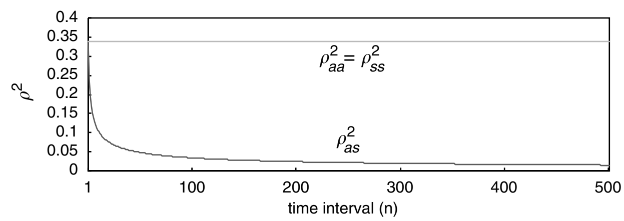

Figure 1 The squared multi-period correlation coefficient.

Table 1 shows that for various values of one-period correlation, the absolute value of the multi-period correlation coefficient decreases monotonically as n increases when one is additive and the other is systematically sampled (denoted by the A&S case). On the other hand, when two random variables are both additive or systematically sampled (denoted by A&A or S&S), the multi-period correlation coefficients are equal to the one-period correlation. These phenomena should be noted when the variables are changed by the aggregation.

The partial regression and correlation coefficients in multiple regression models

In the preceding section, the effect was observed of the selected time interval when two random variables are additive or systematically sampled. Here, we would like to focus on an extension to the multiple regression models. We may consider the subject the following cases: (1) the dependent variable is additive; (2) the dependent variable is systematically sampled.

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

	$\rho^{(1)}=-0.8$			$\rho^{(1)}=-0.5$			$\rho^{(1)}=-0.2$		
Time interval	*A&A*	*S&S*	*A&S*	*A&A*	*S&S*	*A&S*	*A&A*	*S&S*	*A&S*
2	−0.800	−0.800	−0.566	−0.500	−0.500	−0.354	−0.200	−0.200	−0.141
3			−0.462			−0.289			−0.115
4			−0.400			−0.250			−0.100
5			−0.358			−0.224			−0.089
6			−0.327			−0.204			−0.082
7	⋮	⋮	−0.302	⋮	⋮	−0.189	⋮	⋮	−0.076
8			−0.283			−0.177			−0.071
9			−0.267			−0.167			−0.067
10			−0.253			−0.158			−0.063
50			−0.113			−0.071			−0.028
100	−0.800	−0.800	−0.080	−0.500	−0.500	−0.050	−0.200	−0.200	−0.020

	$\rho^{(1)}=0.3$			$\rho^{(1)}=0.6$			$\rho^{(1)}=0.9$		
Time interval	*A&A*	*S&S*	*A&S*	*A&A*	*S&S*	*A&S*	*A&A*	*S&S*	A&S
2	0.300	0.300	0.212	0.600	0.600	0.424	0.900	0.900	0.636
3			0.173			0.346			0.520
4			0.150			0.300			0.450
5			0.134			0.268			0.402
6			0.122			0.245			0.367
7	⋮	⋮	0.113	⋮	⋮	0.227	⋮	⋮	0.340
8			0.106			0.212			0.318
9			0.100			0.200			0.300
10			0.095			0.190			0.285
50			−0.113			−0.071			−0.028
100	0.300	0.300	−0.080	0.600	0.600	−0.050	0.900	0.900	0.090

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 systematically sampled variable simultaneously. We can then construct the following n-period multiple regression model:

$$Y_1^{(n)} = \alpha_{0n} + \alpha_{1n}X_1^{(n)} + \alpha_{2n}X_2^{(n)} + \varepsilon \tag{8}$$

where $Y_1^{(n)}$, $X_1^{(n)}$, and $X_2^{(n)}$ are as defined in the previous section on correlation coefficients. 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

$$V_{1n} = \frac{Y_1^{(n)} - \bar{Y}_1^{(n)}}{\sqrt{n-1}S_{y_1^{(n)}}}, \quad U_{1n} = \frac{X_1^{(n)} - \bar{X}_1^{(n)}}{\sqrt{n-1}S_{x_1^{(n)}}} \quad \text{and} \quad U_{2n} = \frac{X_2^{(n)} - \bar{X}_2^{(n)}}{\sqrt{n-1}S_{x_2^{(n)}}} \tag{9}$$

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

$$V_{1n} = \alpha^*_{0n} + \alpha^*_{1n}U_{1n} + \alpha^*_{2n}U_{2n} + \varepsilon \tag{10}$$

where

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

We can denote here

$$\boldsymbol{\alpha}^*_n = \begin{pmatrix} \alpha^*_{1n} \\ \alpha^*_{2n} \end{pmatrix} \quad \text{and} \quad \mathbf{U}_n = (U_{1n} \quad U_{2n})$$

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

$$\hat{\boldsymbol{\alpha}}^*_n = (\mathbf{U}_n'\mathbf{U}_n)^{-1}(\mathbf{U}_n'\mathbf{V}_{1n}) = \begin{bmatrix} 1 & r_{12}^{(n)} \\ r_{21}^{(n)} & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_{1y_1}^{(n)} \\ r_{2y_1}^{(n)} \end{bmatrix} = \begin{bmatrix} \frac{r_{1y_1}^{(n)} - r_{12}^{(n)}r_{2y_1}^{(n)}}{1-(r_{12}^{(n)})^2} \\ \frac{r_{2y_1}^{(n)} - r_{21}^{(n)}r_{1y_1}^{(n)}}{1-(r_{12}^{(n)})^2} \end{bmatrix} \tag{11}$$

where $r_{ij}^{(n)}$ is the simple correlation between regressor $x_i^{(n)}$ and $x_j^{(n)}$ (see Neter *et al*[18], p 290). Similarly, $r_{jy_1}^{(n)}$ is the simple correlation between the regressor $x_j^{(n)}$ and the response $y_1^{(n)}$.

Proposition 2 *Let $\hat{\alpha}_{1n}$ be the n-period partial regression coefficient of the regression as defined in Equation (8). We obtain the following results*:

1. *As n approaches infinity,* $\lim_{n\to\infty}\hat{\alpha}_{1n} = \hat{\alpha}_{11}$ *and* $\lim_{n\to\infty}\hat{\alpha}_{2n} = \hat{\alpha}_{21} - r_{21}^{(1)}r_{1y_1}^{(1)}S_{y_1^{(1)}}/S_{x_1^{(1)}}$.
2. *If the regressor variables, $X_1^{(n)}$ and $X_2^{(n)}$, are independent, then $\hat{\alpha}_{1n} = \hat{\alpha}_{11}$ and* $\lim_{n\to\infty}\hat{\alpha}_{2n} = \hat{\alpha}_{21}$.

Proof 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 Equation (12) and applying the results of the section 'one is additive, the other is systematically sampled' to Equation (11), the n-period partial regression coefficient $\hat{\alpha}_{1n}$ is as follows:

$$\lim_{n\to\infty}\hat{\alpha}_{1n} = \lim_{n\to\infty}\frac{r_{1y_1}^{(1)} - r_{12}^{(1)}r_{2y_1}^{(1)}/n}{1-(r_{12}^{(1)}/\sqrt{n})^2}\cdot\frac{\sqrt{n}S_{y_1^{(1)}}}{\sqrt{n}S_{x_1^{(1)}}} = \hat{\alpha}_{11} \tag{13}$$

Similarly,

$$\begin{aligned}\lim_{n\to\infty}\hat{\alpha}_{2n} &= \lim_{n\to\infty}\frac{r_{2y_1}^{(1)}/\sqrt{n} - r_{21}^{(1)}r_{1y_1}^{(1)}/\sqrt{n}}{1-(r_{12}^{(1)}/\sqrt{n})^2}\cdot\frac{\sqrt{n}S_{y_1^{(1)}}}{S_{x_2^{(1)}}} \\ &= \hat{\alpha}_{21} - r_{21}^{(1)}r_{1y_1}^{(1)}S_{y_1^{(1)}}/S_{x_1^{(1)}}\end{aligned} \tag{14}$$

which completes the proof.

2. Because $X_1^{(n)}$ and $X_2^{(n)}$ are independent, it is obvious that $r_{12}^{(n)} = r_{21}^{(n)} = 0$ for all n. Similarly, using Equations (11) and (12), then $\hat{\alpha}_{1n} = \hat{\alpha}_{11}$. The result $\lim_{n\to\infty}\hat{\alpha}_{2n} = \hat{\alpha}_{21}$ is obtained by Equations (14). □

Proposition 3 *Let $r_{y_11.2}^{(n)}$ and $r_{y_12.1}^{(n)}$ be the partial correlation coefficients of the regression as defined in Equation (7). Therefore,*

1. $\lim_{n\to\infty} r_{y_11.2}^{(n)} = r_{y_11}^{(1)}$ *(if $X_1^{(n)}$ and $X_2^{(n)}$ are independent, then $r_{y_11.2}^{(n)} = r_{y_11}^{(1)}$).*
2. $\lim_{n\to\infty} r_{y_12.1}^{(n)} = 0$.

Proof 1. The partial correlation coefficient $r_{y_11.2}^{(n)}$ can be expressed by

$$r_{y_11.2}^{(n)} = \frac{r_{y_11}^{(n)} - r_{y_12}^{(n)}r_{12}^{(n)}}{\sqrt{1-(r_{y_12}^{(n)})^2}\sqrt{1-(r_{12}^{(n)})^2}}$$

Because $\lim_{n\to\infty} r_{12}^{(n)} = 0$ and $\lim_{n\to\infty} r_{y_12}^{(n)} = 0$ (see the section 'one is additive, the other is systematically sampled'), we achieve $\lim_{n\to\infty} r_{y_11.2}^{(n)} = r_{y_11}^{(n)}$. Using the relationship $r_{y_11}^{(n)} = r_{y_11}^{(1)}$ (see the section 'both are additive'), we obtain $\lim_{n\to\infty} r_{y_11.2}^{(n)} = r_{y_11}^{(1)}$. In particular, if $X_1^{(n)}$ and $X_2^{(n)}$ are independent, then $r_{y_11.2}^{(n)} = r_{y_11}^{(n)} = r_{y_11}^{(1)}$.

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

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

Since $\lim_{n\to\infty} r^{(n)}_{12} = 0$ (see the section 'one is additive, the other is systematically sampled') and $r^{(n)}_{y_1 1} = r^{(1)}_{y_1 1}$ (see the section 'both are additive'), we directly obtain that $\lim_{n\to\infty} r^{(n)}_{y_1 2.1} = 0$, which completes the proof.

The dependent variable is systematically sampled

When the dependent variable is systematically sampled, the regression model is as follows:

$$Y^{(n)}_2 = \beta_{0n} + \beta_{1n} X^{(n)}_1 + \beta_{2n} X^{(n)}_2 + \varepsilon \qquad (15)$$

where $Y^{(n)}_2$, $X^{(n)}_1$, and $X^{(n)}_2$ are as defined in the previous section on correlation coefficients. Terms β_{0n}, β_{1n}, and β_{2n} are the regression coefficients corresponding to Equation (14). Here, ε is a random error component.

We similarly let:

$$V_{2n} = \frac{Y^{(n)}_2 - \overline{Y}^{(n)}_2}{\sqrt{n-1} S_{y^{(n)}_2}}, \quad U_{1n} = \frac{X^{(n)}_1 - \bar{X}^{(n)}_1}{\sqrt{n-1} S_{x^{(n)}_1}} \quad \text{and}$$
$$U_{2n} = \frac{X^{(n)}_2 - \overline{X}^{(n)}_2}{\sqrt{n-1} S_{x^{(n)}_2}} \qquad (16)$$

The regression model then becomes

$$V_{2n} = \beta^*_{0n} + \beta^*_{1n} U_{1n} + \beta^*_{2n} U_{2n} + \varepsilon \qquad (17)$$

where

$$\beta^*_{0n} = 0 \quad \beta^*_{1n} = \beta_{1n}(S_{x^{(n)}_1}/S_{y^{(n)}_2}), \quad \text{and}$$
$$\beta^*_{2n} = \beta_{2n}(S_{x^{(n)}_2}/S_{y^{(n)}_2})$$

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

$$\hat{\boldsymbol{\beta}}^*_n = (U'_n U_n)^{-1}(U'_n V_{2n}) =$$
$$\begin{bmatrix} 1 & r^{(n)}_{12} \\ r^{(n)}_{21} & 1 \end{bmatrix}^{-1} \begin{bmatrix} r^{(n)}_{1y_2} \\ r^{(n)}_{2y_2} \end{bmatrix} = \begin{bmatrix} \frac{r^{(n)}_{1y_2} - r^{(n)}_{12} r^{(n)}_{2y_2}}{1-(r^{(n)}_{12})^2} \\ \frac{r^{(n)}_{2y_2} - r^{(n)}_{21} r^{(n)}_{1y_2}}{1-(r^{(n)}_{12})^2} \end{bmatrix} \qquad (18)$$

Proposition 4 *Let $\hat{\beta}_{2n}$ be the n-period partial regression coefficient of the regression as defined in (14).*

1. *As n approaches infinity,* $\lim_{n\to\infty} \hat{\beta}_{1n} = 0$ and $\lim_{n\to\infty} \hat{\beta}_{2n} = \hat{\beta}_{21}$.
2. *If the regressor variables, $X^{(n)}_1$ and $X^{(n)}_2$, are independent, then* $\lim_{n\to\infty} \hat{\beta}_{1n} = 0$ *and* $\hat{\beta}_{2n} = \hat{\beta}_{21}$.

Proof The proof for Proposition 4 appears in Appendix.

Proposition 5 *Let $r^{(n)}_{y_2 1.2}$ and $r^{(n)}_{y_2 2.1}$ be the partial correlation coefficients of the regression as defined in (13). Therefore*:

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

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

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

Since $\lim_{n\to\infty} r^{(n)}_{12} = 0$ and $\lim_{n\to\infty} r^{(n)}_{y_2 1} = 0$ (see the section 'one is additive, the other is systematically sampled'), and $\lim_{n\to\infty} r^{(n)}_{y_2 2} = 0$ (see the section 'Both are systematically sampled), 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 presented 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 the section 'one is additive, the other is systematically sampled') and $\lim_{n\to\infty} r^{(n)}_{y_2 2} = r^{(1)}_{y_2 2}$ (see the section 'both are systematically sampled'), we obtain that $\lim_{n\to\infty} r^{(n)}_{y_2 2.1} = r^{(1)}_{y_2 2}$, which completes the proof.

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 daily simple returns of the S&P 500 index and American Express stock shown in Table 2 as a numerical example. The sample period is from January 1990 to December 1999. For the reason of convenient comparison, we use the two variables to simulate the results in order to keep the corresponding parameters the same. Three distinct kinds of the correlation coefficients discussed in the section 'The correlation coefficents between two random variables' seem to be helpful in attempting to sketch out the association between variables in the multiple regression models. Using the various correlation coefficients corresponding to the two return series (corresponding to $E(X) \cong 1.0006$, $\sigma_x \cong 0.0089$, $E(Y) \cong$, 1.0010, $\sigma_y \cong 0.0206$ and $\rho^{(1)}_{xy} \cong 0.5828$), the other parameters (Table 2, Columns (1)–(8)) can be easily obtained.

Table 2 The multi-period partial regression and correlation coefficients

Time interval (n)	$\hat{\alpha}_{1n}$*	$\hat{\alpha}_{2n}$†	$r^{(n)}_{y_1 1.2}$‡	$r^{(n)}_{y_1 2.1}$§	$\hat{\beta}_{1n}$¶	$\hat{\beta}_{2n}$‖	$r^{(n)}_{y_2 1.2}$**	$r^{(n)}_{y_2 2.1}$††
1	0.855	0.855	0.368	0.368	0.368	0.368	0.855	0.855
2	1.155	0.680	0.498	0.232	0.232	0.498	0.340	1.155
3	1.229	0.636	0.530	0.184	0.184	0.530	0.212	1.229
4	1.263	0.617	0.544	0.156	0.156	0.5441	0.154	1.263
5	1.282	0.606	0.552	0.139	0.139	0.552	0.121	1.282
6	1.295	0.598	0.558	0.126	0.126	0.558	0.100	1.295
7	1.303	0.593	0.562	0.116	0.116	0.562	0.085	1.303
8	1.310	0.589	0.564	0.108	0.108	0.564	0.074	1.310
9	1.315	0.587	0.567	0.102	0.102	0.567	0.065	1.315
10	1.319	0.584	0.568	0.096	0.096	0.568	0.058	1.319
11	1.322	0.582	0.570	0.092	0.092	0.570	0.053	1.322
12	1.325	0.581	0.571	0.088	0.088	0.571	0.048	1.325
13	1.327	0.580	0.572	0.084	0.084	0.572	0.045	1.327
14	1.329	0.578	0.573	0.081	0.081	0.573	0.041	1.329
15	1.330	0.577	0.573	0.078	0.078	0.573	0.039	1.330
20	1.336	0.574	0.576	0.068	0.068	0.576	0.029	1.336
25	1.339	0.572	0.577	0.060	0.060	0.577	0.023	1.339
50	1.346	0.568	0.580	0.043	0.043	0.580	0.011	1.346
75	1.348	0.567	0.581	0.035	0.035	0.581	0.008	1.348
100	1.350	0.566	0.581	0.030	0.030	0.581	0.006	1.350
5000	1.353	0.564	0.583	0.004	0.004	0.583	0.000	1.353

*The partial regression coefficient as defined in Proposition 2.
†The partial regression coefficient as defined in Proposition 2.
‡The partial correlation coefficient as defined in Proposition 3.
§The partial correlation coefficient as defined in Proposition 3.
¶The partial regression coefficient as defined in Proposition 4.
‖The partial regression coefficient as defined in Proposition 4.
**The partial correlation coefficient as defined in Proposition 5.
††The partial correlation coefficient as defined in Proposition 5.

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 (1) of Table 2 reveals that $\hat{\alpha}_{1n}$ becomes closer to $\hat{\alpha}_{11}(=1.3527)$ as n increases and $\hat{\alpha}_{1n}=1.3527$ (ie, $\hat{\alpha}_{1n=}\hat{\alpha}_{11}$) as $n=5000$. Also, $r^{(n)}_{y_1 1.2}$ approaches $r^{(1)}_{y_1 1}$ and $r^{(n)}_{y_1 2.1}$ approaches zero as n increases (see Columns (3) and (4)). The results also confirm the result of Proposition 3. Finally, we turn to the case where the dependent variable is systematically sampled. Column (5) indicates that $\hat{\beta}_{1n}$ approaches zero as n increases. The limits claimed in Propositions 4 and 5 are illustrated in the Columns (6)–(8).

Concluding remarks

The relationship between variables is described through regression models and correlation coefficients. If each of the variables is a time series with autocorrelation, then a variety of papers have documented the fact that correlations change over time. When random variables are additive or multiplicative, such effects have been evident even if they are i.i.d. variables over time. However, we should not overlook that some of the variables are from systematic sampling (eg stock prices and interest rates). This paper considers the effect of the time interval when one of the variables is additive and one is from systematic sampling.

Additive and systematically sampled random variables are usually analysed in empirical studies. When the original variables are the stock variables or computed through taking a logarithm for multiplicative variables, they change their frequency by additive operations to become additive variables. Systematic sampling represents the choice of a particular observation value at fixed intervals. Systematically sampled variables are widely applied in many fields.

In this paper, we find that the correlation coefficient is changed with the selected time interval when one is additive and the other is systematically sampled. It is shown that the squared correlation coefficient decreases monotonically as the differencing interval increases, approaching zero in the limit. In sampling for empirical studies, the results should not be ignored, particularly for decisions depending on the correlation between variables. When two random variables are both added or systematically sampled, the correlation coefficient is invariant with time and is equal to the one-period values. Moreover, we also find that the partial regression and correlation coefficients between two additive or systematically sampled variables approach one-period values as n increases. When one of the variables is systematically sampled, they will approach zero in the limit.

These results are similar to the properties of the correlation coefficients. It will be useful to keep these points in mind as we examine the empirical studies.

References

1 Sims CA (1971). Discrete approximations to continuous time distributed lags in econometrics. *Econometrica* **39**: 545–563.
2 Tiao GC and Wei WS (1976). Effects of temporal aggregation on the dynamic relationship between two time series variables. *Biometrika* **63**: 513–523.
3 Wei WS (1982). The effects of systematic sampling and temporal aggregation on causality—a cautionary note. *J Am Statist Assoc* **77**: 316–319.
4 Cunningham SR and Vilasuso JR (1995). Time aggregation and causality tests: results from a monte carlo experiment. *Appl Econ Lett* **2**: 403–405.
5 Cunningham SR and Vilasuso JR (1997). Time aggregation and the money–real GDP relationship. *J Macroecon* **19**(4): 675–695.
6 Schneller IM (1975). Regression analysis for multiplicative phenomena and its implication for the measurement of investment risk. *Mngt Sci* **22**: 422–426.
7 Levhari D and Levy H (1977). The capital asset pricing model and the investment horizon. *Rev. Econ Statist* **59**: 92–104.
8 Levy H (1972). Portfolio performance and the investment horizon. *Mngt Sci* **18**: 645–653.
9 Levy H (1984). Measuring risk and performance over alternative investment horizons. *Finan Anal J* **40**(2): 61–62.
10 Lee YW (1990). Diversification and time: do investment horizons matter? *J Portfol Mngt* **16**(3): 21–26.
11 Bruno M and Easterly W (1998). Inflation crises and long-run growth. *J Monet Econ* **41**: 3–26.
12 Souza LR and Smith J (2002). Bias in the memory parameter for different sampling rates. *Int J Forecast* **18**: 299–313.
13 Levy H and Schwarz G (1997). Correlation and the time interval over which the variables are measured. *J Econom* **76**: 341–350.
14 Levy H, Guttman I and Tkatch I (2001). Regression, correlation, and the time interval: additive-multiplicative framework. *Mngt Sci* **47**(8): 1150–1159.
15 Jea R, Lin JL and Su CT (2004). Correlation and the time interval in multiple regression models. *Eur J Opl Res* **162**: 433–441.
16 Campbell JY, Lo AW and MacKinlay AC (1997). *The Econometrics of Financial Markets*. Princeton University Press: Princeton, NJ.
17 Tsay RS (2002). *Analysis of Financial Time Series*. John Wiley & Sons, Inc: New York.
18 Neter J, Wasserman W and Kutner MH (1989). *Applied Linear Regression models*. 2nd edn. Irwin, Inc: Homewood.

Appendix

Proof of the property 'Both are additive'. Because X_j and Y_j are i.i.d. variables, then for $j=1,2,\ldots,n$, we have

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

The one-period correlation coefficient is

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

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

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

and

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

Similarly, we can obtain

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

and

$$Var(Y_1^{(n)}) = n\sigma_y^2$$

The n-period covariance is

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

The n-period correlation coefficient is as follows:

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

Hence, the correlation coefficient between $X_1^{(n)}$ and $Y_1^{(n)}$ is independent of the differencing interval.

Proof of Proposition 4 The approach here is similar to that of Proposition 2. Substituting the variable $Y_1^{(n)}$ with the variable $Y_2^{(n)}$, we obtain

$$\lim_{n\to\infty} \hat{\beta}_{1n} = \lim_{n\to\infty} \frac{r_{1y_2}^{(1)}/\sqrt{n} - r_{12}^{(1)} r_{2y_2}^{(1)}/\sqrt{n}}{1 - (r_{12}^{(1)}/\sqrt{n})^2} \cdot \frac{S_{y_2^{(1)}}}{\sqrt{n} S_{x_1^{(1)}}} = 0$$

and

$$\lim_{n\to\infty} \hat{\beta}_{2n} = \lim_{n\to\infty} \frac{r_{2y_2}^{(1)} - r_{21}^{(1)} r_{1y_2}^{(1)}/n}{1 - (r_{12}^{(1)}/\sqrt{n})^2} \frac{S_{y_2^{(1)}}}{S_{x_2^{(1)}}} = \hat{\beta}_{21}$$

2. If $X_1^{(n)}$ and $X_2^{(n)}$ are independent, then we have $r_{12}^{(n)} = r_{21}^{(n)} = 0$, and $\lim_{n\to\infty} \hat{\beta}_{1n} = 0$ and $\hat{\beta}_{2n} = \hat{\beta}_{21}$ can be obviously obtained by the above two equations.

Received March 2004;
accepted November 2004 after two revisions