Chapter 1

Introduction

1.1 Overview

Time series data of different frequencies and different time spans are often available to 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 (Sims, 1971; Tiao and Wei, 1976; Wei, 1982; Cunningham and Vilasuso, 1995, 1997). 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.

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 contain autocorrelation. Even if all random variables are independent over time, the effect is documented in recent literature. The effect of the differencing interval on several economic indices and finance has been studied by Schneller (1975), Levhari and Levy (1977), Levy (1972,1984), Lee (1990), Bruno and Easterly (1998), and Souza and Smith (2002). Levy and Schwarz (1997) 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. (2001) write a similar theoretical effect when one of the i.i.d. variables is additive and the other is multiplicative. If we select arbitrarily the time interval and neglect that the corresponding results may be incorrect, then it is likely to lead us to misguided actions. That is, the correlation coefficients are not invariant under changes of the differencing interval, particularly with serial correlation.

1.2 Research Motivations

Generally speaking, a time series variable is either a flow variable or a stock variable. Examples flow variables include gross domestic product (GDP), industrial production, etc. The values of a flow variable are usually obtained through aggregation over equal time intervals. A stock variable such as money supply or inventory levels, can be recorded at each point of time. The data of stock variables are employed by aggregation to transfer various time intervals. Time aggregation involves either temporal aggregation or systematic sampling. Systematic sampling represents the choice of a particular observation value at fixed intervals. Alternatively, temporal aggregation is formed by averaging observations over non-overlapping intervals. This is used to construct the aggregated time series (Cunningham and Vilasuso, 1997).

On the other hand, the multiplicative variables are such as the rate of return on assets, economic growth rate, population growth, which is simply equal to the product of the *n* one-period variables and its net variable over the most recent *n* periods is simply equal to its *n*-period corresponding variable minus one (Campbell et al., 1997).

When random variables are additive or multiplicative, such effects have been evident even if they are independent, identically distributed (i.i.d.) variables over time (Levy and Schwarz, 1997; Levy et al., 2001). This study considers the impact of such analyses containing systematically sampled and temporal aggregated.

1.3 Research Objectives

The objective of this work is to investigate the problem of the time interval effect of the association between two variables that are additive, multiplicative, systematically sampled or temporal aggregated. For flow variables, such as production or imports, they can only be measured over a period of time, so that their values are accumulated. They are denoted by additive variables. 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. However, in many cases the variables are stock variables, then systematically sampled or temporal aggregated variables are often involved in empirical models. For fundamental and practical reasons we study the correlation between multiplicative and aggregate variables that can be widely applied in other fields where correlation analyses are employed.

On the other hand, this study discusses the time interval impact on the partial regression and correlation coefficients in multiple-regression models. 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.

The framework of this research is shown in Figure 1.1.

1.4 Organization

The organization of the remaining chapters for this research is as follows. Chapter 2 reviews the literatures of related work and background. Chapter 3 describes the impact of the employed time interval on the correlation coefficients. Chapter 4 presents the time interval effect of the association between one which is multiplicative and the other which is systematically sampled or temporal aggregated on a simple regression model. Chapter 5 considers the time interval effect when two random variables are additive or multiplicative. 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. Chapter 6 shows the influence of the selected time interval on the association between independent, identically distributed variables over time when one of the variables is additive and one is from systematic sampling. Finally, the conclusions and the directions of future research are given in Chapter 7.

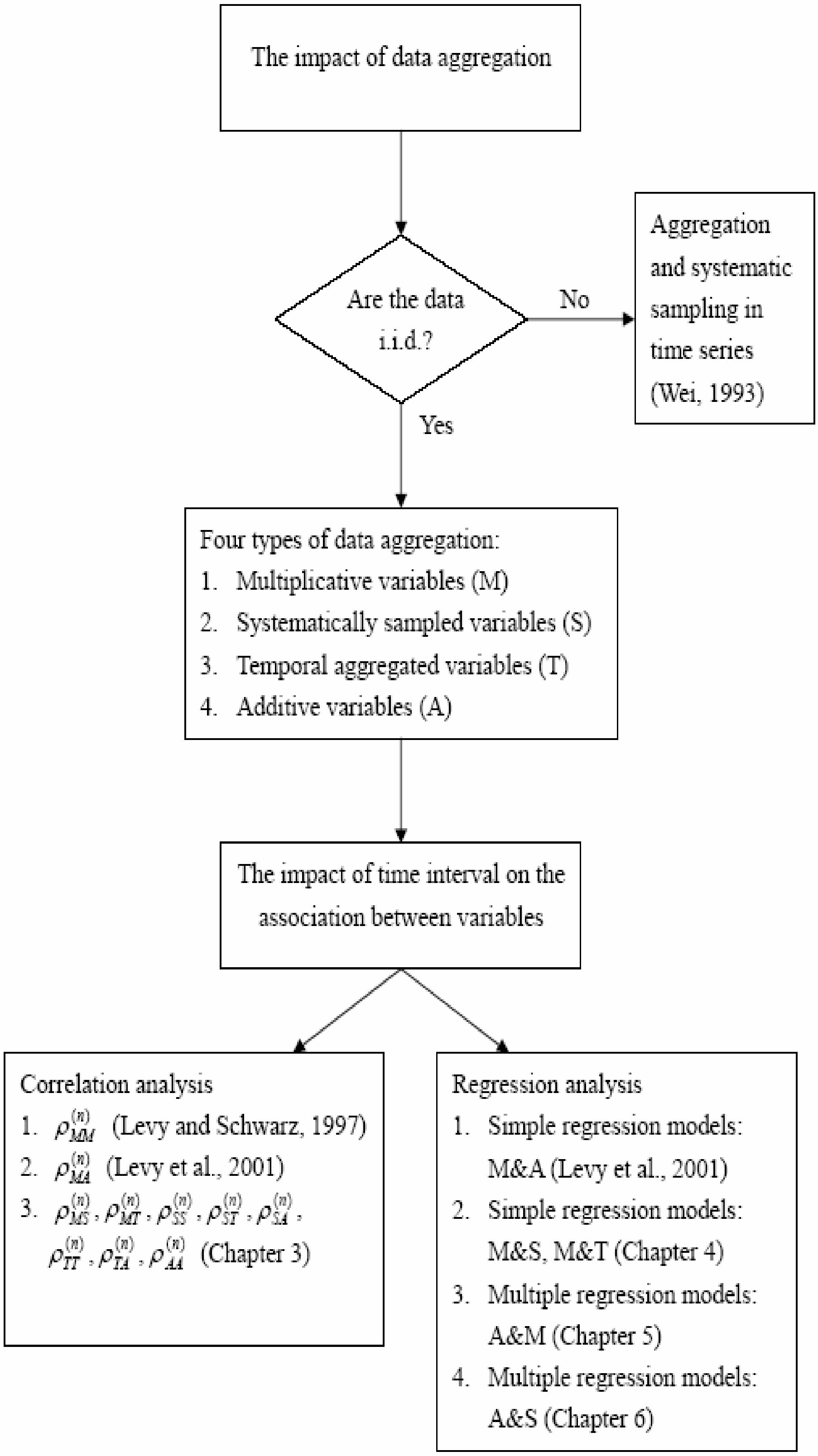

Figure 1.1. The Research Framework

Chapter 2

Related Work

2.1 Literature Review

In time series analysis of a given set of variables, practitioners often have to decide whether to use monthly, quarterly, or annual data. However, in some cases, the investigator often cannot choose the time interval. Thus, the assumed time unit in the interesting investigation may not be the same as the time interval for the employed data. In many studies, data often are available only through aggregation, multiplication, or systematic sampling.

A variety of papers have documented the fact that time aggregation potentially distorts the relationship between variables. Sims (1971) warned that temporal aggregation can result in a spurious causal relationship. Tiao and Wei (1976) and Wei (1982) demonstrate that forming temporally aggregated data can change a true one-way Granger causal relationship into a two-sided causal system. Cunningham and Vilasuso (1995) demonstrate that temporal aggregation is between two and ten time more likely to fail to detect a true causal relationship than is systematic sampling.

In empirical studies, the effect of the selected time interval on variables has been studied extensively. Levy (1972) shows that the performance of mutual funds changes systematically and in a predictable way with changes in the time interval. Fund A may outperform Fund B with monthly data, whereas the opposite holds true with annual data. Lehari and Levy (1977) show that the systematic risk (beta) of securities changes with the length of the time interval. Stocks with high risk become even more risky as the time interval increases, whereas the opposite holds regarding stock characterized by relatively low risk.

If the data contain serial correlation, the effect of the selected time interval on variables has been studied extensively. Levy and Schwarz (1997) and Levy et al. (2001) present that such effects occur even if all random variables are independent over time. 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. The purpose of this dissertation is to complement and extend the results in Levy and Schwarz (1997) and Levy et al. (2001). We consider the various types of variables to study the time interval effect. For practical reasons we also investigates the impact on simple regression models and multiple regression models that can be widely applied in many fields where regression or correlation analyses are employed.

2.2. Definition of Variables

2.2.1 Multiplicative Variables

The asset's gross return over the most recent k periods from date $t-k$ to date t is equal to the product of the k single-period returns from $t-k+1$ to t, i.e.,

$$\begin{aligned} 1+R_t^{(k)} &\equiv (1+R_t)\cdot(1+R_{t-1})\cdots(1+R_{t-k+1}) \\ &= \frac{P_t}{P_{t-1}}\cdot\frac{P_{t-1}}{P_{t-2}}\cdots\frac{P_{t-k+1}}{P_{t-k}} = \frac{P_t}{P_{t-k}}. \end{aligned}$$

Its net return over the most recent k periods is equal to its k-period gross return minus one. These multiperiod returns are called compound returns (Campbell et al., 1997).

2.2.2 Additive Variables

The log return r_t of an asset is defined to be the natural logarithm of its gross return $(1+R_t)$. The multiperiod returns can be written by

$$
\begin{aligned}
r_t^{(k)} &= \log(1+R_t^{(k)}) = \log((1+R_t)\cdot(1+R_{t-1})\cdots(1+R_{t-k+1}) \\
&= \log(1+R_t)+\log(1+R_{t-1})+\cdots+\log(1+R_{t-k+1}) \\
&= r_t + r_{t-1} + \cdots + r_{t-k+1}.
\end{aligned}
$$

The multiperiod log return is the sum of single-period log returns (Campbell et al., 1997).

2.2.3 Temporal Aggregated Variables

Temporal aggregation is an aggregate formed by averaging observations over non-overlapping intervals, i.e., it is the additive variable divided by the number of observations.

2.2.4 Systematically Sampled Variables

The observed series is only a subseries obtained by a scheme of systematic sampling where a single observation from the sampling interval.

Chapter 3

The Relationship Between the One-period and the *n*-period Correlation Coefficients

In time series studies in which the association between variables is an issue, the data collection interval is critical. If each of the variables exists serial correlation, a variety of papers have documented the fact that correlations change over time. This chapter shows the impact of such analyses even if they are independent, identically distributed (i.i.d.) variables over time.

Let $(Y_{11}, Y_{21}, Y_{31}, Y_{41}, X_{11}, X_{21}, X_{31}, X_{41}), \ldots, (Y_{1n}, Y_{2n}, Y_{3n}, Y_{4n}, X_{1n}, X_{2n}, X_{3n}, X_{4n})$ be a sequence of independent, identically distributed variables. These variables are not autocorrelated over time. We define the following variables to denote an *n*-fold increase of the differencing interval involving multiplicative, systematically sampled, temporal aggregated and additive variables.

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

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

and

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

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

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

and

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

The temporal aggregated variables, denoted by $Y_3^{(n)}$ and $X_3^{(n)}$, are given by

$$Y_3^{(n)} = \sum_{j=1}^{n} Y_{3j} / n \quad \text{and} \quad X_3^{(n)} = \sum_{j=1}^{n} X_{3j} / n .$$

The additive variables, denoted by $Y_4^{(n)}$ and $X_4^{(n)}$, are given by

$$Y_4^{(n)} = Y_{41} + Y_{42} + \cdots + Y_{4n}$$

and

$$X_4^{(n)} = X_{41} + X_{42} + \cdots + X_{4n} .$$

Using the above variables, denoted by $Y_1^{(n)}$, $Y_2^{(n)}$, $Y_3^{(n)}$, $Y_4^{(n)}$, $X_1^{(n)}$, $X_2^{(n)}$, $X_3^{(n)}$ and $X_4^{(n)}$, the impact of the selected time interval on the correlation coefficient depending on the various types of variables, is as follows:

3.1 Both Are Multiplicative

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.

3.2 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 independent, identically distributed (i.i.d.) pairs of variables. The *n*-period expected value 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

$$Cov(X_2^{(n)}, Y_2^{(n)}) = \sigma_{x_2 y_2} \tag{3}$$

and

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

Equations (1)-(4) provide the fundamental statistics of the systematically sampled variables $X_2{}^{(n)}$ and $Y_2{}^{(n)}$, respectively. All these results do not depend on the number of periods. Hence, we directly obtain the fact that the correlation and regression coefficient of $X_2{}^{(n)}$ and $Y_2{}^{(n)}$ are unaffected by the selected time interval.

3.3 Both Are Additive

Using two random variables, we can construct a simple regression model. If the independent variable $X_4{}^{(n)}$ and the dependent variable $Y_4{}^{(n)}$ 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_{4j} and Y_{4j} be identically independent distributed variables (i.i.d.), $j = 1,2,\ldots,n$. We have

$$E(X_{4j}) = \mu_{x_4},\ Var(X_{4j}) = \sigma_{x_4}{}^2,\ E(Y_{4j}) = \mu_{y_4} \text{ and } Var(Y_{4j}) = \sigma_{y_4}{}^2.$$

The one-period correlation coefficient is

$$\rho_{x_4y_4}^{(1)} = \frac{Cov(X_{4t}, Y_{4t})}{\sigma_{x_4}\sigma_{y_4}} = \frac{\sigma_{x_4y_4}}{\sigma_{x_4}\sigma_{y_4}}.$$

Because $X_{41}, X_{42}, \ldots, X_{4n}$ are i.i.d., we have

$$E(X_4^{(n)}) = E\left(\sum_{j=1}^{n} X_{4j}\right) = \sum_{j=1}^{n} \mu_{x_4} = n\mu_{x_4} \tag{5}$$

and

$$Var(X_4^{(n)}) = Var\left(\sum_{j=1}^{n} X_{4j}\right) = \sum_{j=1}^{n} \sigma_{x_4}^{\ 2} = n\sigma_{x_4}^{\ 2}. \tag{6}$$

Similarly, we can obtain

$$E(Y_4^{(n)}) = n\mu_{y_4} \tag{7}$$

and

$$Var(Y_4^{(n)}) = n\sigma_{y_4}^{\ 2}. \tag{8}$$

The *n*-period covariance is

$$Cov(X_4^{(n)}, Y_4^{(n)}) = Cov\left(\sum_{j=1}^{n} X_{4j}, \sum_{j=1}^{n} Y_{4j}\right) = nCov(X_{4t}, Y_{4t}) = n\sigma_{x_4y_4}. \tag{9}$$

Using Equation (5), the *n*-period correlation coefficient is as follows:

$$\rho_{x_4y_4}^{(n)} = \frac{Cov(X_4^{(n)}, Y_4^{(n)})}{\sigma_{x_4^{(n)}}\sigma_{y_4^{(n)}}} = \frac{n\sigma_{x_4y_4}}{\sqrt{n}\sigma_{x_4}\sqrt{n}\sigma_{y_4}} = \frac{\sigma_{x_4y_4}}{\sigma_{x_4}\sigma_{y_4}} = \rho_{x_4y_4}^{(1)}. \tag{10}$$

Hence, the correlation coefficient between $X_4^{(n)}$ and $Y_4^{(n)}$ is independent of the differencing interval. Using Equation (10) 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.

3.4 Both Are Temporal Aggregated

Let $X_3^{(n)}$ and $Y_3^{(n)}$ be the temporal aggregated variables. Because $X_3^{(n)}$ and $Y_3^{(n)}$ are i.i.d. random variables over time, it is similar to Section 3.3 that the correlation coefficient between them is unaffected by the selected time interval (e.g. $\rho_{x_3y_3}^{(n)} = \rho_{x_3y_3}^{(1)}$).

3.5 One Is Multiplicative, the Other Is Systematically Sampled

$Y_1^{(n)}$ is an multiplicative random variable and $X_2^{(n)}$ is a random variable from systematic sampling. Because $Y_{11}, Y_{12}, \ldots, Y_{1n}$ are i.i.d., we have

$$E(Y_1^{(n)}) = E(\prod_{j=1}^{n} Y_{1j}) = (\mu_{y_1})^n$$

and

$$Var(Y_1^{(n)}) = Var(\prod_{j=1}^{n} Y_{1j}) = (\sigma_{y_1}^2 + \mu_{y_1}^2)^n - (\mu_{y_1})^{2n} \tag{11}$$

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

$$\begin{aligned} Cov(X_2^{(n)}, Y_1^{(n)}) &= Cov(X_{2h}, \prod_{j=1}^{n} Y_{1j}) \\ &= E(X_{2h} \prod_{j=1}^{n} Y_{1j}) - E(X_{2h}) E(\prod_{j=1}^{n} Y_{1j}) \\ &= (\sigma_{x_2y_1} + \mu_{x_2}\mu_{y_1})(\mu_{y_1})^{n-1} - \mu_{x_2}(\mu_{y_1})^n \\ &= \sigma_{x_2y_1}(\mu_{y_1})^{n-1} \\ &= \rho_{x_2y_1}^{(1)} \sigma_{x_2} \sigma_{y_1} (\mu_{y_1})^{n-1} \end{aligned} \tag{12}$$

Then the *n*-period correlation coefficient is as follows:

$$\begin{aligned} \rho_{x_2y_1}^{(n)} &= \frac{Cov(X_2^{(n)}, Y_1^{(n)})}{\sqrt{Var(X_2^{(n)})}\sqrt{Var(Y_1^{(n)})}} \\ &= \frac{\rho_{x_2y_1}^{(1)} \sigma_{x_2} \sigma_{y_1} (\mu_{y_1})^{n-1}}{\sigma_{x_2} \sqrt{(\sigma_{y_1}^2 + \mu_{y_1}^2)^n - \mu_{y_1}^{2n}}}. \\ &= \frac{\rho_{x_2y_1}^{(1)} \sigma_{y_1} (\mu_{y_1})^{n-1}}{\sqrt{(\sigma_{y_1}^2 + \mu_{y_1}^2)^n - \mu_{y_1}^{2n}}} \end{aligned} \tag{13}$$

Form Equation (13), we find that the *n*-period correlation coefficient $\rho_{x_2 y_1}^{(n)}$ depends on the parameter of the multiplicative variable (i.e., μ_{y_1} and σ_{y_1}) and the one-period correlation $\rho_{x_2 y_1}^{(1)}$, but not on the parameters of the systematic sampling variable X_2.

Proposition 1. Let $\rho_{x_2 y_1}^{(n)}$ be the *n*-period correlation coefficient as defined in Equation (13). We obtain the following results:

1. $\left(\rho_{x_2 y_1}^{(n)}\right)^2$ is monotonically decreasing in *n*.

2. As *n* approaches infinity, $\lim_{n\to\infty} \rho_{x_2 y_1}^{(n)} = 0$.

The proof is shown in the Appendix A.1.

3.6 One Is Multiplicative, the Other Is Additive

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 ($\rho_n^{\ 2}$) monotonically decreases in *n*, and approaches zero when *n* goes to infinity.

3.7 One Is Multiplicative, the Other Is Temporal Aggregated

Let $X_3^{(n)}$ is a temporal aggregated variable. Because $X_{31}, X_{32}, \ldots, X_{3n}$ are i.i.d., we have

$$E(X_3^{(n)}) = E(\sum_{j=1}^{n} X_{3j} / n) = \mu_{x_3} \tag{14}$$

and

$$Var(X_3^{(n)}) = Var(\sum_{j=1}^{n} X_{3j} / n) = \sigma_{x_3}^2 / n. \tag{15}$$

Similarly, the n-period covariance and correlation coefficient of $X_3^{(n)}$ and $Y_1^{(n)}$ can be obtained, respectively, by

$$
\begin{aligned}
Cov(X_3^{(n)}, Y_1^{(n)}) &= Cov(\sum_{i=1}^{n} X_{3i}/n, \prod_{j=1}^{n} Y_{1j}) \\
&= (E(\sum_{i=1}^{n} X_{3i} \prod_{j=1}^{n} Y_{1j}) - E(\sum_{i=1}^{n} X_{3i})E(\prod_{j=1}^{n} Y_{1j}))/n \\
&= (\sigma_{x_3y_1} + \mu_{x_3}\mu_{y_1})(\mu_{y_1})^{n-1} - \mu_{x_3}(\mu_{y_1})^{n} \\
&= \sigma_{x_3y_1}(\mu_{y_1})^{n-1} \\
&= \rho_{x_3y_1}^{(1)} \sigma_{x_3} \sigma_{y_1} (\mu_{y_1})^{n-1}
\end{aligned}
\tag{16}
$$

and

$$
\begin{aligned}
\rho_{x_3y_1}^{(n)} &= \frac{Cov(X_3^{(n)}, Y_1^{(n)})}{\sqrt{Var(X_2^{(n)})}\sqrt{Var(Y_1^{(n)})}} \\
&= \frac{\sqrt{n}\rho_{x_3y_1}^{(1)} \sigma_{x_3} \sigma_{y_1} (\mu_{y_1})^{n-1}}{\sigma_{x_3}\sqrt{(\sigma_{y_1}^2 + \mu_{y_1}^2)^n - \mu_{y_1}^{2n}}} \\
&= \frac{\sqrt{n}\rho_{x_3y_1}^{(1)} \sigma_{y_1} (\mu_{y_1})^{n-1}}{\sqrt{(\sigma_{y_1}^2 + \mu_{y_1}^2)^n - \mu_{y_1}^{2n}}}
\end{aligned}
\tag{17}
$$

Equation (17) is the same as the correlation coefficient between multiplicative and additive variables (see Levy et al., 2001, p.1152). It depends only on the parameters of the multiplicative variable $Y_1^{(n)}$ except the one-period correlation coefficient $\rho_{x_3y_1}^{(1)}$. We directly have the limiting properties that the squared correlation coefficient $(\rho_{x_3y_1}^{(n)})^2$ decreases monotonically as n increases and approaches zero when n goes to infinity.

3.8 One Is Temporal Aggregated, the Other Is Systematically Sampled

Let $Y_3^{(n)}$ and $X_2^{(n)}$ be a temporal aggregated and systematically sampled variable, respectively. The n-period correlation coefficient between them can be obtained by

$$\rho^{(n)}_{x_2y_3} = \frac{Cov(X_2^{(n)}, Y_3^{(n)})}{\sqrt{Var(X_2^{(n)})}\sqrt{Var(Y_3^{(n)})}} = \frac{\sigma_{x_2y_3}/n}{\sigma_{x_2}\sigma_{y_3}/\sqrt{n}} = \frac{1}{\sqrt{n}}\rho^{(1)}_{x_2y_3}.$$

Proposition 2. Let $\rho^{(n)}_{x_2y_3}$ be the *n*-period correlation coefficient as defined in the above equation. We obtain the following results:

1. $\left(\rho^{(n)}_{x_2y_3}\right)^2$ is monotonically decreasing in *n*.

2. As *n* approaches infinity, $\lim_{n\to\infty}\rho^{(n)}_{x_2y_3} = 0$.

Proof. We can directly obtain these results from the above equation.

3.9 One Is Additive, the Other Is Systematically Sampled

Term $Y_4^{(n)}$ is an additive random variable and $X_2^{(n)}$ is a random variable from systematic sampling. Then the *n*-period correlation coefficient is as follows:

$$\rho^{(n)}_{x_2y_4} = \frac{Cov(X_2^{(n)}, Y_4^{(n)})}{\sqrt{Var(X_2^{(n)})}\sqrt{Var(Y_4^{(n)})}} = \frac{\sigma_{x_2y_4}}{\sigma_{x_2}\sqrt{n}\sigma_{y_4}} = \frac{1}{\sqrt{n}}\rho^{(1)}_{x_2y_4}.$$

The above equation is the same as the result of Section 3.8. Hence, the squared correlation coefficient $(\rho^{(n)}_{x_2y_4})^2$ decreases monotonically as *n* increases and approaches zero when *n* goes to infinity.

3.10 One Is Additive, the Other Is Temporal Aggregated

Similarly, $Y_4^{(n)}$ and $X_3^{(n)}$ are additive and temporal aggregated variables, respectively. We have

$$\rho^{(n)}_{x_3y_4} = \frac{Cov(X_3^{(n)}, Y_4^{(n)})}{\sqrt{Var(X_3^{(n)})}\sqrt{Var(Y_4^{(n)})}} = \frac{\sigma_{x_3y_4}n}{\sigma_{x_3}\sigma_{y_4}} = \rho^{(1)}_{x_3y_4}.$$

It is unaffected by the selected time interval. This result is the same as Sections 3.2, 3.3 and 3.4.

From the above cases, the effect of the selected time interval on the various types of correlation coefficients can be summarized as follows:

1. Using the four variables that are multiplicative (M), systematically sampled (S), temporal aggregated (T) and additive (A), we can obtain ten kinds of coefficient coefficients as shown in Table 3.1.

2. It is interesting to note that $\rho_{SS}^{(n)}$, $\rho_{TT}^{(n)}$, $\rho_{TA}^{(n)}$ and $\rho_{AA}^{(n)}$ is unaffected by the time interval employed, the squared others is decreasing in *n*. In empirical study, we often transfer the multiplicative variables to additive variables by taking logarithms, especially in economics and finance. The simplification really has the better behavior over time.

3. The *n*-period correlation coefficients $\rho_{ST}^{(n)}$ and $\rho_{SA}^{(n)}$ is equal. Similarly, $\rho_{MS}^{(n)}$ and $\rho_{MA}^{(n)}$ also have the same results.

4. If one of the variables is multiplicative, the squared correlation coefficients are decreasing as time interval increases. As a result, to study the behavior of multiplicative variables, this subject deserves more than a passing notice.

Numerical Results. Here is a figure that shows the change of the various correlation coefficients by the selected time interval in the above cases. The data used in this numerical example are obtained from the Center for Research in Security Prices (CRSP) of the University of Chicago. We consider the monthly rates of returns of IBM stock and the S&P500 index from January 1926 to December 1999 for 888 observations. Dividend payments are included in the returns. (Tsay, 2002). We assume that the one-period correlation coefficients are equal to 0.6297. Figure 3.1 indicates that the *n*-period correlation coefficients show a horizontal line, which is the same as the results of Sections 3.2-3.4 and Section 3.10. That is to say they are independent of the differencing interval. In addition, when one of variables is multiplicative, $\rho_{MM}^{(n)}$, $\rho_{MT}^{(n)}$, $\rho_{MS}^{(n)}$ and $\rho_{MA}^{(n)}$, the n-period correlation coefficients decrease monotonically as n increases. Figure 3.1 also reveals the decreasing speed of $\rho_{MS}^{(n)}$ is far faster than $\rho_{MT}^{(n)}$. These results deserve more than a passing notice.

Table 3.1. The *n*-period Correlation Coefficients of Different Kinds of Variables

Correlation coefficients	M	S	T	A
M	$\rho_{MM}^{(n)}$ 1	$\rho_{MS}^{(n)}$ 2	$\rho_{MT}^{(n)}$ 3	$\rho_{MA}^{(n)}$ 4
S	-	$\rho_{SS}^{(n)}$ 5	$\rho_{ST}^{(n)}$ 6	$\rho_{SA}^{(n)}$ 7
T	-	-	$\rho_{TT}^{(n)}$ 8	$\rho_{TA}^{(n)}$ 9
A	-	-	-	$\rho_{AA}^{(n)}$ 10

1. The result of Section 3.1.
2. The result of Section 3.5.
3. The result of Section 3.7.
4. The result of Section 3.6.
5. The result of Section 3.2.
6. The result of Section 3.8.
7. The result of Section 3.9.
8. The result of Section 3.4.
9. The result of Section 3.10.
10. The result of Section 3.3.

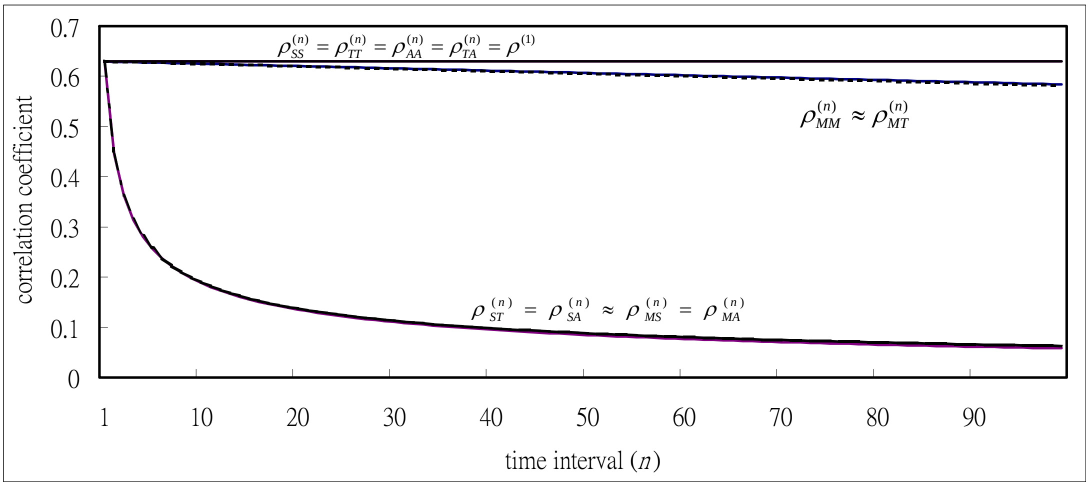

Figure 3.1. The Multi-period Correlation Coefficients

Chapter 4

Simple Regression Models:

Multiplicative-Time Aggregated Framework

From what has been mentioned above, we know the effect of the selected time interval when one random variable is systematically sampled or temporal aggregated and the other variable is multiplicative. Here, we would like to analyze the impact of the time interval on the regression coefficient. We consider the subject under the following cases: (1) the dependent variable is multiplicative; (2) the dependent variable is systematically sampled or temporal aggregated.

4.1 The Dependent Variable Is Multiplicative

In the regression model, the dependent variable $Y_1^{(n)}$ is multiplicative and the regressor $X_2^{(n)}$ is systematically sampled. We can then construct the following *n*-period simple regression model:

$$Y_1^{(n)} = \alpha_{0n} + \alpha_{1n} X_2^{(n)} + \varepsilon \,, \tag{18}$$

where $Y_1^{(n)}$ and $X_2^{(n)}$ are as defined in Section 3. Terms α_{0n} and α_{1n} are the regression coefficients corresponding to the *n*-period 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.

By the results of the OLS estimation, we get

$$\hat{\alpha}_{1n} = r_{x_2 y_1}^{(n)} \cdot \frac{s_{y_1^{(n)}}}{s_{x_2^{(n)}}} = r_{x_2 y_1}^{(1)} (\bar{y}_1)^{n-1} \frac{s_{y_1}}{s_{x_2}} = \hat{\alpha}_{11} (\bar{y}_1)^{n-1} \tag{19}$$

where s_{y_1} and s_{x_2} are the standard errors of $Y_1^{(n)}$ and $X_2^{(n)}$, respectively. The limiting properties of $\hat{\alpha}_{1n}$ are the same as the regression coefficient where the independent variable is additive and the dependent variable is multiplicative (see Levy et al., 2001).

Now substituting the systematically sampled variable $X_2^{(n)}$ with the temporal aggregated variable $X_3^{(n)}$, we can rewrite Equation (18) to become

$$Y_1^{(n)} = \alpha'_{0n} + \alpha'_{1n} X_3^{(n)} + \varepsilon \,. \tag{20}$$

Similarly, we have

$$\hat{\alpha}'_{1n} = r_{x_3 y_1}^{(n)} \frac{s_{y_1^{(n)}}}{s_{x_3^{(n)}}} = n r_{x_3 y_1}^{(1)} (\bar{y}_1)^{n-1} \frac{s_{y_1}}{s_{x_3}} = n \hat{\alpha}'_{11} (\bar{y}_1)^{n-1} \,. \tag{21}$$

Comparing Equation (20) and (21), we find Equation (21) is Equation (19) times n and both the limiting results are the same except the speed of convergence.

Proposition 3. Let $\hat{\alpha}_{1n}$ and $\hat{\alpha}'_{1n}$ be the regression coefficients as defined in Equations (19) and (21), respectively. When the one-period regression coefficients, $\hat{\alpha}_{11}$ and $\hat{\alpha}'_{11}$, is greater than zero, we obtain the following results:

1. If $|\bar{y}_1| > 1$, $|\hat{\alpha}_{1n}|$ and $|\hat{\alpha}'_{1n}|$ are monotonically increasing in n, $\lim_{n\to\infty} |\hat{\alpha}_{1n}| = \infty$ and $\lim_{n\to\infty} |\hat{\alpha}'_{1n}| = \infty$.

2. If $0 \le |\bar{y}_1| < 1$, $|\hat{\alpha}_{1n}|$ is monotonically decreasing in n, $\lim_{n\to\infty} |\hat{\alpha}_{1n}| = 0$ and $\lim_{n\to\infty} |\hat{\alpha}'_{1n}| = 0$.

3. If $|\bar{y}_1| = 1$, $|\hat{\alpha}_{1n}|$ and $|\hat{\alpha}'_{1n}|$ are always equal to $|\hat{\alpha}_{11}|$ and $|\hat{\alpha}'_{11}|$.

These results are obtained directly from Equations (19) and (21).

4.2 The Dependent Variable Is Systematically Sampled or Temporal Aggregated

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

$$Y_2^{(n)} = \beta_{0n} + \beta_{1n} X_1^{(n)} + \varepsilon\,, \tag{22}$$

where $Y_2^{(n)}$ and $X_1^{(n)}$ are as defined in Section 3. Terms β_{0n} and β_{1n} are the regression coefficients corresponding to Equation (22). Here, ε is a random error component. By the results of the OLS estimation, we get

$$\hat{\beta}_{1n} = r^{(n)}_{x_1 y_2} \cdot \frac{s_{y_2^{(n)}}}{s_{x_1^{(n)}}} = \frac{r^{(1)}_{x_1 y_2} \left(\bar{x}_1\right)^{n-1} s_{y_2} s_{x_1}}{(s_{x_1}^2 + \bar{x}_1^{\,2})^n - \bar{x}_1^{2n}} = \hat{\beta}_{11} \frac{\left(\bar{x}_1\right)^{n-1} s_{x_1}^2}{(s_{x_1}^2 + \bar{x}_1^{\,2})^n - \bar{x}_1^{2n}} \tag{23}$$

where s_{y_2} and s_{x_1} are the standard errors of $Y_2^{(n)}$ and $X_1^{(n)}$, respectively.

Proposition 4. Let $\hat{\beta}_{1n}$ be the regression coefficients as defined in Equation (23). As *n* approaches infinity, we obtain the following results:

1. If $|\bar{x}_1| \geq 1$ then $\lim\limits_{n\to\infty} \hat{\beta}_{1n} = 0$.

2. If $|\bar{x}_1| < 1$ and $\dfrac{|\bar{x}_1|}{s_{x_1}^2 + \bar{x}_1^2} \geq 1$, then $\lim\limits_{n\to\infty} \hat{\beta}_{1n} = \begin{cases} +\infty & \text{if } \hat{\beta}_{11} > 0 \\ -\infty & \text{if } \hat{\beta}_{11} < 0 \end{cases}$.

3. If $|\bar{x}_1| < 1$ and $\dfrac{|\bar{x}_1|}{s_{x_1}^2 + \bar{x}_1^2} < 1$, then $\lim\limits_{n\to\infty} \hat{\beta}_{1n} = 0$.

The above results can be proved by Levy et al. (2001).

Alternatively, when the dependent variable is temporal aggregated, the regression model is as the following

$$Y_3^{(n)} = \beta'_{0n} + \beta'_{1n} X_1^{(n)} + \varepsilon\,. \tag{24}$$

We find $\hat{\beta}'_{1n}$ is equal to $\hat{\beta}_{1n}$ defined in Equation (23). They have the same properties.

4.3 Numerical Example

Table 4.1 and Table 4.2 illustrate the effect of the selected time interval on the correlation coefficients in the regression models corresponding to the U.S. stock market. We use the same data and assumption (the one-period regression coefficient is equal to one) as Levy et al. (2001) in order to help to compare the difference between them. The multi-period regression coefficients in the regression models when the independent variable is systematically sampled or temporal aggregated and the dependent variable is multiplicative, are shown in Table 4.1 as a numerical example. In Table 4.1, Columns (1)-(4) show the monotonically increasing property of regression coefficients $\hat{\alpha}_{1n}$ and $\hat{\alpha}'_{1n}$ when $\bar{y}_1$ is greater than one, which agrees with Proposition 3. Columns (5) and (6) indicate the cases of $0 < \bar{y}_1 < 1$. Term $\hat{\alpha}_{1n}$ is monotonically decreasing in *n* and approaches zero when *n* goes to infinity. However, term $\hat{\alpha}'_{1n}$ is monotonically increasing as *n* is small. When *n* is large, $\hat{\alpha}'_{1n}$ is monotonically decreasing in *n* and approaches zero when *n* goes to infinity. Table 4.1 also tells us that the increasing speed of $\hat{\alpha}'_{1n}$ in *n* is faster and the decreasing speed of $\hat{\alpha}'_{1n}$ in *n* is slower.

Table 4.2 shows the numerical results of Proposition 4. We assume that the dependent variable is systematically sampled or temporal aggregated and the independent variable is multiplicative in a simple regression model. Columns (1)-(4) of Table 4.2 indicate that the regression coefficient $\hat{\beta}_{1n}$ decreases monotonically and approaches zero as *n* increases. Columns (5) and (6) reveal the case where $|\bar{x}_1| < 1$. They also show the claim of Case 3 and Case 2 of Proposition 4, respectively.

4.4 Concluding Remarks

In many applications, the relationship between two time series is of major interest. The association between variables is often measured by regression and correlation coefficients, which is widely applicable in economics and finance.

In time series studies in which direction of causation is an issue, the data collection interval is critical. The data for empirical studies are sometimes limited, and aggregation is usually used in empirical study. 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. On the other hand, the multiplicative variables are such as the rate of return on stocks, population size, economic growth, etc. Here, we use the three kinds of variables to express the impact of time interval on a simple regression model. The effect on the regression coefficients is substantial. Hence, the time interval of the data for such analyses cannot be selected arbitrarily.

Table 4.1. The Multi-period Regression Coefficients When the Dependent Variable Is Multiplicative

s_{y_1}	0.0260	0.0963	0.2110	0.6000	0.1000	0.0100
$\bar{y}_1$	1.0022	1.0287	1.1200	1.1200	0.9920	0.9500
	(1)	(2)	(3)	(4)	(5)	(6)
n			$\hat{\alpha}_{1n}$			
1	1	1	1	1	1	1
2	1.0022	1.0287	1.1200	1.1500	0.9920	0.9500
3	1.0044	1.0582	1.2544	1.3225	0.9841	0.9025
4	1.0066	1.0886	1.4049	1.5209	0.9762	0.8574
5	1.0088	1.1198	1.5735	1.7490	0.9684	0.8145
6	1.0110	1.1520	1.7623	2.0114	0.9606	0.7738
7	1.0133	1.1850	1.9738	2.3131	0.9529	0.7351
8	1.0155	1.2190	2.2107	2.6600	0.9453	0.6983
9	1.0177	1.2540	2.4760	3.0590	0.9378	0.6634
10	1.0200	1.2900	2.7731	3.5179	0.9303	0.6302
50	1.1137	4.0008	258.0377	942.3108	0.6746	0.0810
100	1.2430	16.4659	7.46E+04	1.02E+06	0.4515	0.0062
500	2.9940	1.36E+06	3.63E+24	1.94E+30	0.0182	0.0000
1000	8.9835	1.89E+12	1.48E+49	4.34E+60	0.0003	0.0000
n			$\hat{\alpha}'_{1n}$			
1	1	1	1	1	1	1
2	1.4173	1.4548	1.5839	1.6263	1.4029	1.3435
3	1.7397	1.8329	2.1727	2.2906	1.7044	1.5632
4	2.0132	2.1772	2.8099	3.0418	1.9524	1.7148
5	2.2558	2.5040	3.5185	3.9109	2.1654	1.8213
6	2.4766	2.8218	4.3168	4.9268	2.3531	1.8954
7	2.6809	3.1353	5.2222	6.1198	2.5213	1.9449
8	2.8723	3.4480	6.2528	7.5237	2.6738	1.9752
9	3.0532	3.7621	7.4279	9.1771	2.8133	1.9903
10	3.2254	4.0794	8.7692	11.1245	2.9417	1.9930
50	7.8750	28.2900	1824.6019	6663.1437	4.7704	0.5727
100	12.4304	164.6589	7.46E+05	1.02E+07	4.5150	0.0623
500	66.9468	3.03E+07	8.11E+25	4.34E+31	0.4063	0.0000
1000	284.0822	5.98E+13	4.66E+50	1.37E+62	0.0104	0.0000

Table 4.2. The Multi-period Regression Coefficient When the Independent Variable Is Multiplicative

s_{x_1}	0.0260	0.0963	0.2110	0.6000	0.1000	0.0100
$\bar{x}_1$	1.0022	1.0287	1.1200	1.1200	0.9920	0.9500
$\bar{x}_1/(s_{x_1}^2+\bar{x}_1^2)$	0.9971	0.9637	0.8623	0.6835	0.9979	1.0525
	(1)	(2)	(3)	(4)	(5)	(6)
n			$\hat{\beta}_{1n}$			
1	1	1	1	1	1	1
2	0.4987	0.4839	0.4386	0.3827	0.5015	0.5263
3	0.3316	0.3122	0.2565	0.1943	0.3353	0.3693
4	0.2481	0.2267	0.1687	0.1105	0.2522	0.2915
5	0.1980	0.1755	0.1184	0.0667	0.2024	0.2455
6	0.1646	0.1415	0.0865	0.0418	0.1691	0.2153
7	0.1407	0.1174	0.0650	0.0268	0.1454	0.1943
8	0.1228	0.0994	0.0499	0.0175	0.1276	0.1789
9	0.1089	0.0855	0.0389	0.0115	0.1138	0.1674
10	0.0977	0.0745	0.0307	0.0077	0.1027	0.1586
11	0.0886	0.0656	0.0244	0.0051	0.0936	0.1518
12	0.0810	0.0582	0.0196	0.0034	0.0861	0.1464
13	0.0746	0.0520	0.0159	0.0023	0.0797	0.1423
14	0.0691	0.0467	0.0129	0.0016	0.0742	0.1390
15	0.0643	0.0422	0.0106	0.0011	0.0694	0.1366
20	0.0476	0.0268	0.0041	0.0002	0.0528	0.1324
25	0.0376	0.0182	0.0017	0.0000	0.0428	0.1368
50	0.0177	0.0040	0.0000	0.0000	0.0229	0.2463
100	0.0078	0.0004	0.0000	0.0000	0.0129	1.5958
500	0.0006	0.0000	0.0000	0.0000	0.0036	2.54E+08
1000	0.0001	0.0000	0.0000	0.0000	0.0013	1.70E+19
1500	0.0000	0.0000	0.0000	0.0000	0.0004	1.51E+30

(1) Correspond to weekly data of Common Stock Index.
(2) Correspond to quarterly data of Common Stock Index.
(3) Correspond to yearly data of Common Stock Index.
(4) Correspond to individual stocks or to other phenomena not necessarily taken from the stock market.
(5) A case in Proposition 4, item 3.
(6) A case in Proposition 4, item 2.

Chapter 5

Multiple Regression Models:

Additive-Multiplicative Framework

This chapter presents the time interval effect of multiple regression models in which some of the variables are additive and some are multiplicative. It is to complement and extend the results in Levy and Schwarz (1997) and Levy et al. (2001). 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. 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, 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.

5.1 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:

$$Y_4^{(n)} = \alpha_{0n} + \alpha_{1n} X_4^{(n)} + \alpha_{2n} X_1^{(n)} + \varepsilon, \tag{25}$$

where $Y_4^{(n)}$, $X_4^{(n)}$, and $X_1^{(n)}$ are as defined in Section 3. 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

$$V_{4n} = \frac{Y_4^{(n)} - \overline{Y}_4^{(n)}}{\sqrt{n-1}S_{y_4^{(n)}}}, \quad U_{4n} = \frac{X_4^{(n)} - \overline{X}_4^{(n)}}{\sqrt{n-1}S_{x_4^{(n)}}} \quad \text{and} \quad U_{1n} = \frac{X_1^{(n)} - \overline{X}_1^{(n)}}{\sqrt{n-1}S_{x_1^{(n)}}}. \tag{26}$$

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

$$V_{4n} = \alpha_{0n}^* + \alpha_{1n}^* U_{4n} + \alpha_{2n}^* U_{1n} + \varepsilon, \tag{27}$$

where $\alpha_{0n}^* = 0$, $\alpha_{1n}^* = \alpha_{1n} \dfrac{S_{x_4^{(n)}}}{S_{y_4^{(n)}}}$ and $\alpha_{2n}^* = \alpha_{2n} \dfrac{S_{x_1^{(n)}}}{S_{y_4^{(n)}}}$.

We can denote here $\boldsymbol{\alpha}_n^* = \begin{pmatrix} \alpha_{1n}^* \\ \alpha_{2n}^* \end{pmatrix}$, $\mathbf{V}_n = (V_{4n})$ and $\mathbf{U}_n = (U_{4n} \quad U_{1n})$.

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

$$\hat{\boldsymbol{\alpha}}_n^* = (\mathbf{U}_n{}'\mathbf{U}_n)^{-1}(\mathbf{U}_n{}'\mathbf{V}_n) = \begin{bmatrix} 1 & r_{41}^{(n)} \\ r_{14}^{(n)} & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_{4y_4}^{(n)} \\ r_{1y_4}^{(n)} \end{bmatrix} = \begin{bmatrix} \dfrac{r_{4y_4}^{(n)} - r_{41}^{(n)} r_{1y_4}^{(n)}}{1-(r_{41}^{(n)})^2} \\ \dfrac{r_{1y_4}^{(n)} - r_{14}^{(n)} r_{1y_4}^{(n)}}{1-(r_{41}^{(n)})^2} \end{bmatrix}, \tag{28}$$

where $r_{ij}^{(n)}$ is the simple correlation between regressor $x_i^{(n)}$ and $x_j^{(n)}$. Similarly, $r_{jy_4}^{(n)}$ is the simple correlation between the regressor $x_j^{(n)}$ and the response $y_4^{(n)}$.

Proposition 5. Let $\hat{\alpha}_{1n}$ be the n-period partial regression coefficient of the regression as defined in (25). We obtain the following results:

1. As n approaches infinity, $\lim\limits_{n\to\infty} \hat{\alpha}_{1n} = \hat{\alpha}_{11}$ (for the properties of the partial regression coefficient $\hat{\alpha}_{2n}$, see Levy et al., 2001).
2. If the regressor variables, $X_4^{(n)}$ and $X_1^{(n)}$, are independent, then $\hat{\alpha}_{1n} = \hat{\alpha}_{11}$.

Proof.

1. Applying the results of Section 3.3 to Equation (28), we know that $\lim_{n\to\infty} r_{1y_4}^{(n)} = 0$ and $\lim_{n\to\infty} r_{41}^{(n)} = \lim_{n\to\infty} r_{14}^{(n)} = 0$. Hence, as n approaches infinity, the standardized regression coefficients $\hat{\boldsymbol{\alpha}}_n^*$ can be obtained

$$\lim_{n\to\infty} \hat{\boldsymbol{\alpha}}_n^* = \begin{bmatrix} r_{4y_4}^{(n)} \\ 0 \end{bmatrix} = \begin{bmatrix} r_{4y_4}^{(1)} \\ 0 \end{bmatrix}, \tag{29}$$

where $r_{4y_4}^{(n)} = r_{4y_4}^{(1)}$ is shown in Section 3.1. Using the relationship between the original and standardized regression coefficients, we achieve

$$\hat{\alpha}_{1n} = \hat{\alpha}_{1n}^* \frac{S_{y_4^{(n)}}}{S_{x_4^{(n)}}}, \qquad \hat{\alpha}_{2n} = \hat{\alpha}_{2n}^* \frac{S_{y_4^{(n)}}}{S_{x_1^{(n)}}}, \tag{30}$$

and

$$\hat{\alpha}_{0n} = \bar{y}_4^{(n)} - \hat{\alpha}_{1n}\bar{x}_4^{(n)} - \hat{\alpha}_{2n}\bar{x}_1^{(n)}.$$

Using Equations (29), and (30), the n-period partial regression coefficient $\hat{\alpha}_{1n}$ is as follows:

$$\lim_{n\to\infty} \hat{\alpha}_{1n} = \lim_{n\to\infty} \hat{\alpha}_{1n}^* \frac{S_{y_4^{(n)}}}{S_{x_4^{(n)}}} = r_{4y_4}^{(1)} \frac{\sqrt{n}S_{y_4^{(1)}}}{\sqrt{n}S_{x_4^{(1)}}} = \hat{\alpha}_{11},$$

which completes the proof.

2. Because $X_4^{(n)}$ and $X_1^{(n)}$ are independent, it is obvious that $r_{41}^{(n)} = r_{14}^{(n)} = 0$. Similarly, using Equation (30), we obtain $\hat{\alpha}_{1n} = \hat{\alpha}_{11}$.

Proposition 6. Let $r_{y_4 4.1}^{(n)}$ and $r_{y_4 1.4}^{(n)}$ be the partial correlation coefficients of the regression as defined in (25). Therefore,

1. $\lim_{n\to\infty} r_{y_4 4.1}^{(n)} = r_{y_4 4}^{(1)}$ (if $X_4^{(n)}$ and $X_1^{(n)}$ are independent, then $r_{y_4 4.1}^{(n)} = r_{y_4 4}^{(1)}$).

2. $\lim_{n\to\infty} r_{y_4 1.4}^{(n)} = 0$.

Proof.

1. The partial correlation coefficient $r_{y_4 4.1}^{(n)}$ can be expressed by

$$r_{y_4 4.1}^{(n)} = \frac{r_{y_4 4}^{(n)} - r_{y_4 1}^{(n)} r_{41}^{(n)}}{\sqrt{1-(r_{y_4 1}^{(n)})^2}\sqrt{1-(r_{41}^{(n)})^2}}.$$

Because $\lim_{n\to\infty} r_{41}^{(n)} = 0$ and $\lim_{n\to\infty} r_{y_4 1}^{(n)} = 0$ (see Section 3.6), we achieve $\lim_{n\to\infty} r_{y_4 4.1}^{(n)} = r_{y_4 4}^{(n)}$. Using the relationship $r_{y_4 4}^{(n)} = r_{y_4 4}^{(1)}$ (see Section 3.3), we obtain $\lim_{n\to\infty} r_{y_4 4.1}^{(n)} = r_{y_4 4}^{(1)}$. In particular, if $X_4^{(n)}$ and $X_1^{(n)}$ are independent, then $r_{y_4 4.1}^{(n)} = r_{y_4 4}^{(n)} = r_{y_4 4}^{(1)}$.

2. The partial correlation coefficient $r_{y_4 1.4}^{(n)}$ can be expressed by

$$r_{y_4 1.4}^{(n)} = \frac{r_{y_4 1}^{(n)} - r_{y_4 4}^{(n)} r_{41}^{(n)}}{\sqrt{1-(r_{y_4 4}^{(n)})^2}\sqrt{1-(r_{41}^{(n)})^2}}.$$

Since $\lim_{n\to\infty} r_{41}^{(n)} = 0$ (see Section 3.6) and $r_{y_4 4}^{(n)} = r_{y_4 4}^{(1)}$ (see Section 3.3), we directly obtain that $\lim_{n\to\infty} r_{y1.4}^{(n)} = 0$, which completes the proof.

5.2 The Dependent Variable Is Multiplicative

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

$$Y_1^{(n)} = \beta_{0n} + \beta_{1n} X_4^{(n)} + \beta_{2n} X_1^{(n)} + \varepsilon, \qquad (31)$$

where $Y_1^{(n)}$, $X_4^{(n)}$, and $X_1^{(n)}$ are as defined in Section 3. Terms β_{0n}, β_{1n}, and β_{2n} are the regression coefficients corresponding to Equation (31). Here, ε is a random error component.

We similarly let:

$$V_{1n}=\frac{Y_1^{(n)}-\overline{Y}_1^{(n)}}{\sqrt{n-1}S_{y_1^{(n)}}},\quad U_{4n}=\frac{X_4^{(n)}-\overline{X}_4^{(n)}}{\sqrt{n-1}S_{x_4^{(n)}}}\quad \text{and}\quad U_{1n}=\frac{X_1^{(n)}-\overline{X}_1^{(n)}}{\sqrt{n-1}S_{x_1^{(n)}}}. \tag{32}$$

The regression model then becomes

$$V_{1n}=\beta_{0n}^{*}+\beta_{1n}^{*}U_{4n}+\beta_{2n}^{*}U_{1n}+\varepsilon\,, \tag{33}$$

where $\beta_{0n}^{*}=0$, $\beta_{1n}^{*}=\beta_{1n}\dfrac{S_{x_4^{(n)}}}{S_{y_1^{(n)}}}$, and $\beta_{2n}^{*}=\beta_{2n}\dfrac{S_{x_1^{(n)}}}{S_{y_1^{(n)}}}$.

We can denote $\boldsymbol{\beta}_n^{*}=\begin{pmatrix}\beta_{1n}^{*}\\ \beta_{2n}^{*}\end{pmatrix}$, $\mathbf{V}_n=\left(V_{1n}\right)$ and $\mathbf{U}_n=\left(U_{4n}\quad U_{1n}\right)$.

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

$$\hat{\boldsymbol{\beta}}_n^{*}=(\mathbf{U}_n{}'\mathbf{U}_n)^{-1}(\mathbf{U}_n{}'\mathbf{V}_n)=\begin{bmatrix}1 & r_{41}^{(n)}\\ r_{14}^{(n)} & 1\end{bmatrix}^{-1}\begin{bmatrix}r_{4y_1}^{(n)}\\ r_{1y_1}^{(n)}\end{bmatrix}=\begin{bmatrix}\dfrac{r_{4y_1}^{(n)}-r_{41}^{(n)}r_{1y_1}^{(n)}}{1-(r_{41}^{(n)})^2}\\ \dfrac{r_{1y_1}^{(n)}-r_{14}^{(n)}r_{4y_1}^{(n)}}{1-(r_{41}^{(n)})^2}\end{bmatrix}.$$

Proposition 7. Let $\hat{\beta}_{2n}$ be the *n*-period partial regression coefficient of the regression as defined in (31). 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 7 appears in Appendix A.2.

Proposition 8. Let $r_{y_1 4.1}^{(n)}$ and $r_{y_1 1.4}^{(n)}$ be the partial correlation coefficients of the regression as defined in (31). Therefore:

1. $\lim_{n\to\infty} r_{y_1 4.1}^{(n)}=0$.

2. $\lim_{n\to\infty} r_{y_1 1.4}^{(n)}=0$.

Proof.

1. The partial correlation coefficient $r_{y_1 4.1}^{(n)}$ can be expressed by

$$r_{y_1 4.1}^{(n)} = \frac{r_{y_1 4}^{(n)} - r_{y_1 1}^{(n)} r_{41}^{(n)}}{\sqrt{1-(r_{y_1 1}^{(n)})^2}\sqrt{1-(r_{41}^{(n)})^2}}.$$

Since $\lim_{n\to\infty} r_{41}^{(n)} = 0$ and $\lim_{n\to\infty} r_{y_1 4}^{(n)} = 0$ (see Section 3.6), and $\lim_{n\to\infty} r_{y_1 1}^{(n)} = 0$ (see Section 2.1), we obtain that $\lim_{n\to\infty} r_{y_1 4.1}^{(n)} = 0$.

2. The partial correlation coefficient $r_{y_1 1.4}^{(n)}$ can be expressed by

$$r_{y_1 1.4}^{(n)} = \frac{r_{y_1 1}^{(n)} - r_{y_1 4}^{(n)} r_{41}^{(n)}}{\sqrt{1-(r_{y_1 4}^{(n)})^2}\sqrt{1-(r_{41}^{(n)})^2}}.$$

Similarly, because $\lim_{n\to\infty} r_{41}^{(n)} = 0$, $\lim_{n\to\infty} r_{y_1 4}^{(n)} = 0$ (see Section 3.6) and $\lim_{n\to\infty} r_{y_1 1}^{(n)} = 0$ (see Section 3.1), we obtain that $\lim_{n\to\infty} r_{y_1 4.1}^{(n)} = 0$, which completes the proof.

5.3 Numerical Example

Table 5.1 illustrates the effect of the selected time interval on the partial regression and correlation coefficients in the multiple regression models corresponding to the U.S. stock market. We use the monthly rates of returns of IBM stock and the S&P500 index shown in Table 5.1 as a numerical example. The sample period is from January 1926 to December 1999. In Table 5.1, 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 5.1, 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 5 where the dependent variable is additive. Column (4) of Table 5.1 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, $r^{(n)}_{y_4 4.1}$ approaches $r^{(1)}_{y_4 4}$ and $r^{(n)}_{y_4 1.4}$ approaches zero as n increases (see Columns (5) and (6)). The results also confirm with the claim of Proposition 6. 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 7. The claim of Proposition 8 is shown in Columns (8) and (9).

5.4 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 chapter 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 chapter 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.

Table 5.1. The Multi-period Partial Regression and Correlation Coefficients

n	Corr (n)	Corr (n)	Corr (n)	$\hat{\alpha}_{1n}$	$r^{(n)}_{y_4 4.1}$	$r^{(n)}_{y_4 1.4}$	$\hat{\beta}_{2n}$	$r^{(n)}_{y_1 4.1}$	$r^{(n)}_{y_1 1.4}$
	A&A	M&M	A&M						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(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 5.
(5) The partial correlation coefficient as defined in Proposition 6.
(6) The partial correlation coefficient as defined in Proposition 6.
(7) The partial regression coefficient as defined in Proposition 7.
(8) The partial correlation coefficient as defined in Proposition 8.
(9) The partial correlation coefficient as defined in Proposition 8.

Chapter 6

Multiple Regression Models: Aggregated-Systematically Sampled Framework

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 chapter we investigate the time interval effect of multiple regression models in which the variables are additive or systematically sampled. Levy and Schwarz (1997) and Levy et al. (2001) 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 modeling the statistical behavior of some variables over time.

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

$$Y_4^{(n)} = \alpha_{0n} + \alpha_{1n} X_4^{(n)} + \alpha_{2n} X_2^{(n)} + \varepsilon \, , \tag{34}$$

where $Y_4^{(n)}$, $X_4^{(n)}$, and $X_2^{(n)}$ are as defined in Section 3. 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

$$V_{4n}=\frac{Y_4^{(n)}-\overline{Y}_4^{(n)}}{\sqrt{n-1}S_{y_4^{(n)}}},\ U_{1n}=\frac{X_4^{(n)}-\overline{X}_4^{(n)}}{\sqrt{n-1}S_{x_4^{(n)}}}\ \text{ and }\ U_{2n}=\frac{X_2^{(n)}-\overline{X}_2^{(n)}}{\sqrt{n-1}S_{x_2^{(n)}}}. \tag{35}$$

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

$$V_{4n}=\alpha_{0n}^*+\alpha_{1n}^*U_{4n}+\alpha_{2n}^*U_{2n}+\varepsilon\,, \tag{36}$$

where $\alpha_{0n}^*=0$, $\alpha_{1n}^*=\alpha_{1n}\dfrac{S_{x_4^{(n)}}}{S_{y_4^{(n)}}}$ and $\alpha_{2n}^*=\alpha_{2n}\dfrac{S_{x_2^{(n)}}}{S_{y_1^{(n)}}}$.

We can denote here $\boldsymbol{\alpha}_n^*=\begin{pmatrix}\alpha_{1n}^*\\ \alpha_{2n}^*\end{pmatrix}$, $\mathbf{V}_n=(V_{4n})$ and $\mathbf{U}_n=(U_{4n}\quad U_{2n})$.

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

$$\hat{\boldsymbol{\alpha}}_n^*=(\mathbf{U}_n{}'\mathbf{U}_n)^{-1}(\mathbf{U}_n{}'\mathbf{V}_n)=\begin{bmatrix}1 & r_{42}^{(n)}\\ r_{24}^{(n)} & 1\end{bmatrix}^{-1}\begin{bmatrix}r_{4y_4}^{(n)}\\ r_{2y_4}^{(n)}\end{bmatrix}=\begin{bmatrix}\dfrac{r_{4y_4}^{(n)}-r_{42}^{(n)}r_{2y_4}^{(n)}}{1-(r_{12}^{(n)})^2}\\ \dfrac{r_{2y_4}^{(n)}-r_{24}^{(n)}r_{4y_4}^{(n)}}{1-(r_{42}^{(n)})^2}\end{bmatrix}, \tag{27}$$

where $r_{ij}^{(n)}$ is the simple correlation between regressor $x_i^{(n)}$ and $x_j^{(n)}$ (see Neter, 1989, p.290). Similarly, $r_{jy_4}^{(n)}$ is the simple correlation between the regressor $x_j^{(n)}$ and the response $y_4^{(n)}$.

Proposition 9. Let $\hat{\alpha}_{1n}$ be the *n*-period partial regression coefficient of the regression as defined in Equation (34). We obtain the following results:

1. As *n* approaches infinity, $\lim\limits_{n\to\infty}\hat{\alpha}_{1n}=\hat{\alpha}_{11}$ and $\lim\limits_{n\to\infty}\hat{\alpha}_{2n}=\hat{\alpha}_{21}-r_{24}^{(1)}r_{4y_4}^{(1)}S_{y_4^{(1)}}/S_{x_4^{(1)}}$.

2. If the regressor variables, $X_4^{(n)}$ and $X_2^{(n)}$, are independent, then $\hat{\alpha}_{1n}=\hat{\alpha}_{11}$ and $\lim\limits_{n\to\infty}\hat{\alpha}_{2n}=\hat{\alpha}_{21}$.

Proof.

1. Using the relationship between the original and standardized regression coefficients, we achieve

$$\hat{\alpha}_{1n} = \hat{\alpha}^*_{1n} \frac{S_{y_4^{(n)}}}{S_{x_4^{(n)}}}, \qquad \hat{\alpha}_{2n} = \hat{\alpha}^*_{2n} \frac{S_{y_4^{(n)}}}{S_{x_2^{(n)}}} \tag{38}$$

and

$$\hat{\alpha}_{0n} = \bar{y}_4^{(n)} - \hat{\alpha}_{1n}\bar{x}_4^{(n)} - \hat{\alpha}_{2n}\bar{x}_2^{(n)}.$$

Using Equation (38) and applying the results of Section 3.3 to Equation (37), the n-period partial regression coefficient $\hat{\alpha}_{1n}$ is as follows:

$$\lim_{n\to\infty} \hat{\alpha}_{1n} = \lim_{n\to\infty} \frac{r^{(1)}_{4y_4} - r^{(1)}_{42} r^{(1)}_{2y_4}/n}{1-(r^{(1)}_{42}/\sqrt{n})^2} \cdot \frac{\sqrt{n}S_{y_4^{(1)}}}{\sqrt{n}S_{x_4^{(1)}}} = \hat{\alpha}_{11} \tag{39}$$

Similarly,

$$\lim_{n\to\infty} \hat{\alpha}_{2n} = \lim_{n\to\infty} \frac{r^{(1)}_{2y_4}/\sqrt{n} - r^{(1)}_{24} r^{(1)}_{4y_4}/\sqrt{n}}{1-(r^{(1)}_{42}/\sqrt{n})^2} \cdot \frac{\sqrt{n}S_{y_4^{(1)}}}{S_{x_2^{(1)}}} = \hat{\alpha}_{21} - r^{(1)}_{24} r^{(1)}_{4y_4} S_{y_4^{(1)}} / S_{x_4^{(1)}}, \tag{40}$$

which completes the proof.

2. Because $X_4^{(n)}$ and $X_2^{(n)}$ are independent, it is obvious that $r^{(n)}_{42} = r^{(n)}_{24} = 0$ for all n. Similarly, using Equations (37) and (38), then $\hat{\alpha}_{1n} = \hat{\alpha}_{11}$. The result $\lim_{n\to\infty} \hat{\alpha}_{2n} = \hat{\alpha}_{21}$ is obtained by Equation (40).

Proposition 10. Let $r^{(n)}_{y_4 4.2}$ and $r^{(n)}_{y_4 2.4}$ be the partial correlation coefficients of the regression as defined in Equation (33). Therefore,

1. $\lim_{n\to\infty} r^{(n)}_{y_4 4.2} = r^{(1)}_{y_4 4}$ (if $X_4^{(n)}$ and $X_2^{(n)}$ are independent, then $r^{(n)}_{y_4 4.2} = r^{(1)}_{y_4 4}$).
2. $\lim_{n\to\infty} r^{(n)}_{y_4 2.4} = 0$.

Proof.

1. The partial correlation coefficient $r_{y_4 4.2}^{(n)}$ can be expressed by

$$r_{y_4 4.2}^{(n)} = \frac{r_{y_4 4}^{(n)} - r_{y_4 2}^{(n)} r_{42}^{(n)}}{\sqrt{1-(r_{y_4 2}^{(n)})^2}\sqrt{1-(r_{42}^{(n)})^2}}.$$

Because $\lim_{n\to\infty} r_{42}^{(n)} = 0$ and $\lim_{n\to\infty} r_{y_4 2}^{(n)} = 0$ (see Section 3.9), we achieve $\lim_{n\to\infty} r_{y_4 4.2}^{(n)} = r_{y_4 4}^{(n)}$. Using the relationship $r_{y_4 4}^{(n)} = r_{y_4 4}^{(1)}$ (see Section 3.6), we obtain $\lim_{n\to\infty} r_{y_4 4.2}^{(n)} = r_{y_4 4}^{(1)}$. In particular, if $X_4^{(n)}$ and $X_2^{(n)}$ are independent, then $r_{y_4 4.2}^{(n)} = r_{y_4 4}^{(n)} = r_{y_4 4}^{(1)}$.

2. The partial correlation coefficient $r_{y_4 2.4}^{(n)}$ can be expressed by

$$r_{y_4 2.4}^{(n)} = \frac{r_{y_4 2}^{(n)} - r_{y_4 4}^{(n)} r_{42}^{(n)}}{\sqrt{1-(r_{y_4 4}^{(n)})^2}\sqrt{1-(r_{42}^{(n)})^2}}.$$

Since $\lim_{n\to\infty} r_{42}^{(n)} = 0$ (see Section 3.9) and $r_{y_4 4}^{(n)} = r_{y_4 4}^{(1)}$ (see Section 3.6), we directly obtain that $\lim_{n\to\infty} r_{y_4 2.4}^{(n)} = 0$, which completes the proof.

6.2 The Dependent Variable Is Systematically Sampled

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

$$Y_2^{(n)} = \beta_{0n} + \beta_{1n} X_4^{(n)} + \beta_{2n} X_2^{(n)} + \varepsilon, \tag{41}$$

where $Y_2^{(n)}$, $X_4^{(n)}$, and $X_2^{(n)}$ are as defined in Section 3. Terms β_{0n}, β_{1n}, and β_{2n} are the regression coefficients corresponding to Equation (41). Here, ε is a random error component.

We similarly let:

$$V_{2n} = \frac{Y_2^{(n)} - \overline{Y}_2^{(n)}}{\sqrt{n-1}S_{y_2^{(n)}}},\ U_{4n} = \frac{X_4^{(n)} - \overline{X}_4^{(n)}}{\sqrt{n-1}S_{x_4^{(n)}}} \text{ and } U_{2n} = \frac{X_2^{(n)} - \overline{X}_2^{(n)}}{\sqrt{n-1}S_{x_2^{(n)}}}. \tag{42}$$

The regression model then becomes

$$V_{2n} = \beta^*_{0n} + \beta^*_{1n} U_{4n} + \beta^*_{2n} U_{2n} + \varepsilon \,, \tag{43}$$

where $\beta^*_{0n} = 0$, $\beta^*_{1n} = \beta_{1n} \dfrac{S_{x_4^{(n)}}}{S_{y_2^{(n)}}}$, and $\beta^*_{2n} = \beta_{2n} \dfrac{S_{x_2^{(n)}}}{S_{y_2^{(n)}}}$.

We can denote $\boldsymbol{\beta}^*_n = \begin{pmatrix} \beta^*_{1n} \\ \beta^*_{2n} \end{pmatrix}$, $\mathbf{V}_n = (V_{2n})$ and $\mathbf{U}_n = (U_{4n} \quad U_{2n})$.

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

$$\hat{\boldsymbol{\beta}}^*_n = (\mathbf{U}_n{}'\mathbf{U}_n)^{-1}(\mathbf{U}_n{}'\mathbf{V}_n) = \begin{bmatrix} 1 & r^{(n)}_{42} \\ r^{(n)}_{24} & 1 \end{bmatrix}^{-1} \begin{bmatrix} r^{(n)}_{4y_2} \\ r^{(n)}_{2y_2} \end{bmatrix} = \begin{bmatrix} \dfrac{r^{(n)}_{4y_2} - r^{(n)}_{42} r^{(n)}_{2y_2}}{1-(r^{(n)}_{42})^2} \\ \dfrac{r^{(n)}_{2y_2} - r^{(n)}_{24} r^{(n)}_{4y_2}}{1-(r^{(n)}_{42})^2} \end{bmatrix}. \tag{44}$$

Proposition 11. Let $\hat{\beta}_{2n}$ be the *n*-period partial regression coefficient of the regression as defined in (41).

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, $X^{(n)}_4$ 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 11 appears in Appendix A.3.

Proposition 12. Let $r^{(n)}_{y_2 4.2}$ and $r^{(n)}_{y_2 2.4}$ be the partial correlation coefficients of the regression as defined in (41). Therefore:

1. $\lim_{n\to\infty} r^{(n)}_{y_2 4.2} = 0$.
2. $\lim_{n\to\infty} r^{(n)}_{y_2 2.4} = r^{(1)}_{y_2 2}$.

Proof.

1. The partial correlation coefficient $r_{y_2 4.2}^{(n)}$ can be expressed by

$$r_{y_2 4.2}^{(n)} = \frac{r_{y_2 4}^{(n)} - r_{y_2 2}^{(n)} r_{42}^{(n)}}{\sqrt{1-(r_{y_2 2}^{(n)})^2}\sqrt{1-(r_{42}^{(n)})^2}}.$$

Since $\lim_{n\to\infty} r_{42}^{(n)} = 0$ and $\lim_{n\to\infty} r_{y_2 4}^{(n)} = 0$ (see Section 3.9), and $\lim_{n\to\infty} r_{y_2 2}^{(n)} = 0$ (see Section 3.2), we obtain that $\lim_{n\to\infty} r_{y_2 4.2}^{(n)} = 0$.

2. The partial correlation coefficient $r_{y_2 2.4}^{(n)}$ can be presented by

$$r_{y_2 2.4}^{(n)} = \frac{r_{y_2 2}^{(n)} - r_{y_2 4}^{(n)} r_{42}^{(n)}}{\sqrt{1-(r_{y_2 4}^{(n)})^2}\sqrt{1-(r_{42}^{(n)})^2}}.$$

Similarly, because $\lim_{n\to\infty} r_{42}^{(n)} = 0$, $\lim_{n\to\infty} r_{y_2 4}^{(n)} = 0$ (see Section 3.9) and $\lim_{n\to\infty} r_{y_2 2}^{(n)} = r_{y_2 2}^{(1)}$ (see Section 3.2), we obtain that $\lim_{n\to\infty} r_{y_2 2.4}^{(n)} = r_{y_2 2}^{(1)}$, which completes the proof.

6.3 Numerical Example

Table 6.1 illustrates the effect of the selected time interval on the partial regression and correlation coefficients in the multiple regression models corresponding to the U.S. stock market. We use the daily simple returns of the S&P 500 index and American Express stock shown in Table 6.1 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 Section 3 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 $\rho_{xy}^{(1)} \cong 0.5828$), the other parameters (Table 6.1, Columns (1)-(8)) can be easily obtained.

To begin with, we claim that $\lim_{n\to\infty} \hat{\alpha}_{1n} = \hat{\alpha}_{11}$ in Proposition 9 where the dependent variable is additive. Column (1) of Table 6.1 reveals that $\hat{\alpha}_{1n}$ becomes closer to $\hat{\alpha}_{11}$ (=1.3527) as n increases and $\hat{\alpha}_{1n} = 1.3527$ (i.e., $\hat{\alpha}_{1n} = \hat{\alpha}_{11}$) as $n = 5000$. Also, $r_{y_4 4.2}^{(n)}$ approaches $r_{y_4 4}^{(1)}$ and $r_{y_4 2.4}^{(n)}$ approaches zero as n increases (see Columns (3) and (4)). The results also confirm the result of Proposition 10. 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 11 and 12 are illustrated in the Columns (6), (7) and (8).

6.4 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 (e.g. 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 analyzed 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 chapter 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.

Table 6.1. The Multi-period Partial Regression and Correlation Coefficients

Time interval (n)	$\hat{\alpha}_{1n}$ (1)	$\hat{\alpha}_{2n}$ (2)	$r^{(n)}_{y_4 4.2}$ (3)	$r^{(n)}_{y_4 2.4}$ (4)	$\hat{\beta}_{1n}$ (5)	$\hat{\beta}_{2n}$ (6)	$r^{(n)}_{y_2 4.2}$ (7)	$r^{(n)}_{y_2 2.4}$ (8)
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

(1) & (2) The partial regression coefficient as defined in Proposition 9.
(3) & (4) The partial correlation coefficient as defined in Proposition 10.
(5) & (6) The partial regression coefficient as defined in Proposition 11.
(7) & (8) The partial correlation coefficient as defined in Proposition 12.

Chapter 7

Conclusions

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 aggregation, systematic sampling or multiplication depending on whether the variables are flow variables, stock variables or the growth rate of specific variables, respectively. This dissertation aims to study the effect of time interval on the association between the four types of variables.

In time series studies in which the association between variables is an issue, the collection interval is critical. If each of the variables exists serial correlation, a variety of papers have documented the fact that correlations change over time. This dissertation considers the impact of such analyses even if they are independent, identically distributed (i.i.d.) variables over time. It follows from what has been said that the squared correlation coefficients are decreasing as time interval increases except $\rho_{SS}^{(n)}$, $\rho_{TT}^{(n)}$, $\rho_{TA}^{(n)}$ and $\rho_{AA}^{(n)}$. 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 the impact of the employed time interval.

In addition to these, the important point to note is the effect of the selected time interval on regression coefficients and partial regression coefficients in simple regression models and multiple regression models respectively. They are also used to measure the association between variables. We study the simple and multiple linear regressions models with the above mentioned four types of variables. It can be also shown that regression coefficients and partial regression coefficients are affected by the selected time interval. The longer time intervals will decrease the relevant association between variables, particularly for the multiplicative dependent variable.

All these things make it clear that we should not overlook these phenomena in such empirical analyses or it might lead to making incorrect decisions and misguided actions. Their broad appeal and usefulness are applied to studies conducted in economics and finance. At the same time, we can also examine the impact of the selected data frequency on the demand correlation and the value of information sharing in a supply chain. An extension to the supply chain inventory management is the subject of future research.

Appendix

A.1. Proof of Proposition1.

$$(\rho_{x_2y_1}^{(n)})^2 = \frac{(\rho_{x_2y_1}^{(1)})^2 \sigma_{y_1}^2 (\mu_{y_1}^2)^{n-1}}{(\sigma_{y_1}^2 + \mu_{y_1}^2)^n - \mu_{y_1}^{2n}}$$

Differentiating $(\rho_{x_2y_1}^{(n)})^2$ with respect to *n*, we get $\frac{\partial(\rho_{x_2y_1}^{(n)})^2}{\partial n} \le 0$ (Levy et al., 2001). It is shown that $(\rho_{x_2y_1}^{(n)})^2$ is monotonically decreasing in *n*. Now turn to the second claim of Proposition 1. Because $0 \le (\rho_{x_2y_1}^{(n)})^2 \le n \cdot (\rho_{x_2y_1}^{(n)})^2$ and $\lim_{n\to\infty} n \cdot (\rho_{x_2y_1}^{(n)})^2 = 0$ (Levy et al., 2001), we can obtain $\lim_{n\to\infty}(\rho_{x_2y_1}^{(n)})^2 = 0$ (by Sandwich Theorem), which completes the proof.

A.2. Proof of Proposition 7.

Here we demonstrate the results of Proposition 7 from Levy and Schwarz (1997). Substituting the variable *B* (see Levy and Schwarz (1997) Equation (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).

A.3. Proof of Proposition 11.

1. The approach here is similar to that of Proposition 9. Substituting the variable $Y_4^{(n)}$ with the variable $Y_2^{(n)}$, we obtain

$$\lim_{n\to\infty}\hat{\beta}_{1n} = \lim_{n\to\infty}\frac{r_{4y_2}^{(1)}/\sqrt{n} - r_{42}^{(1)}r_{2y_2}^{(1)}/\sqrt{n}}{1-(r_{42}^{(1)}/\sqrt{n})^2}\cdot\frac{S_{y_2^{(1)}}}{\sqrt{n}S_{x_4^{(1)}}} = 0$$

and

$$\lim_{n\to\infty}\hat{\beta}_{2n} = \lim_{n\to\infty}\frac{r_{2y_2}^{(1)} - r_{24}^{(1)}r_{4y_2}^{(1)}/n}{1-(r_{42}^{(1)}/\sqrt{n})^2}\cdot\frac{S_{y_2^{(1)}}}{S_{x_2^{(1)}}} = \hat{\beta}_{21}.$$

2. If $X_4^{(n)}$ and $X_2^{(n)}$ are independent, then we have $r_{42}^{(n)} = r_{24}^{(n)} = 0$, 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.

References

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

Campbell, J. Y., A. W. Lo and A. C. MacKinlay (1997), *The Econometrics of Financial Markets*, Princeton University Press: New Jersey.

Cunningham, S. R. and J. R. Vilasuso (1995), Time Aggregation and Causality Tests: Results from a Monte Carlo Experiment, *Applied Economics Letters* 2: 403-405.

Cunningham, S. R. and J. R. Vilasuso (1997), Time Aggregation and the Money -Real GDP Relationship, *Journal of Macroeconomics* 19 (4): 675-695.

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

Levhari, D. and H. Levy (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. and G. Schwarz (1997), Correlation and the time interval over which the variables are measured, *Journal of Econometrics* 76: 341-350.

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

Neter, J., W. Wasserman, and M. H. Kutner (1989), Applied linear regression models, 2nd 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.

Sims, C. A. (1971), Discrete Approximations to Continuous Time Distributed Lags in Econometrics, *Econometrica* 39: 545-563.

Souza, L. R. and J. Smith (2002), Bias in the memory parameter for different sampling rates, *International Journal of Forecasting* 18: 299-313.

Tiao, G. C. and W. S. Wei (1976), Effects of Temporal Aggregation on the Dynamic Relationship Between Two Time Series Variables, *Biometrika* 63: 513-523.

Tsay, R. S. (2002), Analysis of financial time series, John Wiley & Sons, Inc.

Wei, W. S. (1982), The Effects of Systematic Sampling and Temporal Aggregation on Causality—A Cautionary Note, *Journal of the American Statistical Association* 77: 316-319.

Wei, W. S. (1993), Time Series Analysis, Addison-Wesley Publishing Company, Inc.