Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

European Journal of Operational Research 162 (2005) 433–441

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/dsw

Stochastics and Statistics

Correlation and the time interval in multiple regression models

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

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

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

Abstract

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

Keywords: Correlation coefficient; Partial regression coefficient; Time interval

1. Introduction

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

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

* Corresponding author. Address: Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu, Taiwan. Tel.: +886-3-5731857; fax: +886-3-5722392.

E-mail address: ctsu@cc.nctu.edu.tw (C.-T. Su).

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

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

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

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

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

2. The correlation coefficients between two random variables

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

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

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

and

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

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

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

and

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

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