On the origin of the Epps effect

Bence Tóth[a,b,*], János Kertész[b,c]

[a]*ISI Foundation, Viale S. Severo, 65, I-10133 Torino, Italy*
[b]*Department of Theoretical Physics, Budapest University of Technology and Economics, Budafoki út. 8. H-1111 Budapest, Hungary*
[c]*Laboratory of Computational Engineering, Helsinki University of Technology, P.O. Box 9203, FI-02015, Finland*

Abstract

The Epps effect, the decrease of correlations between stock returns for short time windows, was traced back to the trading asynchronicity and to the occasional lead-lag relation between the prices. We study pairs of stocks where the latter is negligible and confirm the importance of asynchronicity but point out that alone these aspects are insufficient to give account for the whole effect.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Financial correlations; Epps effect; High frequency data

1. Introduction

In 1979 Epps [1] reported empirical results showing that stock return correlations decrease as the sampling frequency increases . Later this phenomenon has been found in several studies of different stock markets [2,3] and of foreign exchange markets [4,5]. Since the most important factors in classical portfolio management are the correlations between the individual assets, the understanding and the accurate description of these correlations on all timescales are of major importance.

Considerable effort has been devoted to uncover the relationship of the correlation coefficient and the sampling frequency [6–11]. So far the studies have revealed two factors causing the Epps effect. The first one is the lead-lag effect [12–14] between the stock returns, which can occur for pairs of stocks of very different capitalization and/or for some functional dependencies between them. In this case the maximum of the time-dependent correlation function is at non-zero time lag resulting in increasing correlations as the sampling timescale gets in the same order of magnitude as the characteristic lag (typically of the order of minutes) [3,6]. However, in a recent study [14] we showed that through the years this effect becomes less important since the characteristic lag shrinks, signalizing an increasing market efficiency.

The second, more important factor is the asynchronicity of trading [6,7,11,15]. Empirical results showed that taking into account only synchronous price ticks the Epps effect is reduced, i.e., the measured correlations increase. It is natural to assume that, for a given sampling frequency, increasing trading activity should

---

*Corresponding author. Department of Theoretical Physics, Budapest University of Technology and Economics, Budafoki út. 8. H-1111 Budapest, Hungary.

E-mail address: bence@maxwell.phy.bme.hu (B. Tóth).

enhance synchronicity, leading to a weaker Epps effect. Indeed, using Monte Carlo experiments, an inverse relation was found between trading activity and the correlation drop [6].

The aim of the present communication is to investigate empirically the role of the trading activity on the Epps effect. We find that in contrast to the simulations, the empirical data do not scale in the expected way indicating that a further, possibly human factor is also at play.

2. Methodology

In our analysis we used the trade and quote (TAQ) database of the New York stock exchange (NYSE) for the period of 4.1.1993–31.12.2003, containing tick-by-tick data. To avoid the problems occurring from splits in the prices of stocks, which cause large logarithmic return values in the time series, we applied a filtering procedure. In high-frequency data, we omitted returns larger than 5% of the current price of the stock. This retains all logarithmic returns caused by simple changes in prices but excludes splits which are usually half or one-third of the price. We computed correlations for each day separately and averaged over the set of days, this way avoiding large overnight returns and trades out of the market opening hours.

We computed the logarithmic returns of stock prices:

$$r_{\Delta t}^A(t) = \ln \frac{p^A(t)}{p^A(t - \Delta t)}, \tag{1}$$

where $p^A(t)$ stands for the price of stock A at time t. The time-dependent correlation function $C_{\Delta t}^{A,B}(\tau)$ of stocks A and B is defined by

$$C_{\Delta t}^{A,B}(\tau) = \frac{\langle r_{\Delta t}^A(t) r_{\Delta t}^B(t+\tau) \rangle - \langle r_{\Delta t}^A(t) \rangle \langle r_{\Delta t}^B(t+\tau) \rangle}{\sigma^A \sigma^B}. \tag{2}$$

The notion $\langle \cdots \rangle$ stands for the moving time average over the considered period:

$$\langle r_{\Delta t}(t) \rangle = \frac{1}{T - \Delta t} \sum_{i=\Delta t}^{T} r_{\Delta t}(i), \tag{3}$$

where time is measured in seconds and T is the time span of the data.

σ is the standard deviation of the return series:

$$\sigma = \sqrt{\langle r_{\Delta t}(t)^2 \rangle - \langle r_{\Delta t}(t) \rangle^2}. \tag{4}$$

For pairs of stocks with a lead–lag effect the function $C_{\Delta t}^{A,B}$ has a peak at non-zero τ.

The equal-time correlation coefficient is: $\rho_{\Delta t}^{A,B} \equiv C_{\Delta t}^{A,B}(\tau = 0)$. In our notations the Epps effect means that $\rho_{\Delta t}$ decreases as Δt decreases (see Fig. 1). The prices are defined as being constant between two consecutive trades (previous tick estimator), thus the Δt timescale of the sampling can be chosen arbitrarily.

3. Results

In order to separate the different origins of the Epps effect we consider only pairs of stocks where the lead–lag effect is negligible, i.e., for which the price changes are highly correlated with the peak position of $C_{\Delta t}^{A,B}$ of Eq. (2) being at $\tau \approx 0$. Good candidates are the pairs Coca-Cola Company (KO)/PepsiCo, Inc. (PEP), Caterpillar Inc. (CAT)/Deere & Company (DE), Wal-Mart Stores, Inc. (WMT)/Sprint Nextel Corporation (S), etc. We shall illustrate our findings on the example of KO/PEP but the other pairs show similar behavior.

In Fig. 1 we show the correlation coefficient as a function of the sampling timescale for the whole period, 1993–2003. One can see that the correlations increase as the sampling frequency decreases. The growth is very fast in the beginning but several hours are needed for the correlation to reach its asymptotic value.

To study the trading frequency dependence of the correlation drop we computed the Epps curve separately for different years. Fig. 2 shows the correlation coefficient as a function of the sampling timescale for the years 1993, 1997, 2000 and 2003.