On Epps Effect and Rebalancing of Hedged Portfolio in Multiple Frequencies

Mustafa U. Torun[1] and Ali N. Akansu[2]
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
University Heights, Newark, NJ 07102 USA
[1]mustafa.torun@njit.edu
[2]akansu@njit.edu

***Abstract*—Correlations of financial asset returns play a central role in designing investment portfolios by using Markowitz's modern portfolio theory (MPT). Correlations are calculated from asset prices that happen at various trading time intervals. Therefore, trading frequency dictates correlation values. This phenomenon is called the Epps effect in finance. We present variations of correlations as a function of trading frequency to quantify Epps effect. The results reiterate that portfolio rebalancing, particularly in multiple trading frequencies, requires good estimation of correlations in order to deliver reliable hedging.**

I. INTRODUCTION

Correlations of asset returns in an investment portfolio is an important aspect of MPT [1]. It also serves for various hedging and pairs trading strategies. A good estimation of correlation is crucial for good performance in all trading and risk management systems [2]. However, a good correlation estimation, especially in intra-day and high-frequency trading where sampling periods are typically below a minute, is of a major challenge [3–5]. It is known in finance that the correlations among financial asset returns decrease as the sampling (trading) period of prices decreases, and this phenomenon is called Epps effect [6]. In this paper, we calculate correlations of several assets and bring additional insights to this important phenomenon for improved hedging in multiple trading frequencies.

In Section II, we define pair-wise correlation of asset returns, and highlight its significance in finance. In Section III, we treat correlation as a function of trading interval with the help of Black-Scholes pricing model [7]. Then, we revisit the Epps effect and quantify the asynchronous trading as its usual suspect. We present closing remarks in Section IV.

II. CORRELATIONS OF ASSET RETURNS

Return of a financial asset over a time period of T is defined as

$$R(n) = [P(n)/P(n-1)] - 1, \tag{1}$$

where $P(n)$ is the last traded price at discrete time n that is obtained by sampling the continuous-time last traded price of the asset at $t = nT$ as given

$$P(n) = P(t)\,|_{t=nT}\ , \tag{2}$$

It is customary to refer the asset returns with respect to their sampling periods, e.g. 30-min returns, 1-hour returns, end of day (EOD) returns. The cross-correlation coefficient of the returns of two assets, $R_1(n)$ and $R_2(n)$, is defined as

$$\rho = \frac{E\left[R_1(n)R_2(n)\right] - \mu_1\mu_2}{\sigma_1\sigma_2}, \tag{3}$$

where $E\left[\cdot\right]$ is the expected value operator, $\mu = E\left[R\right]$ is the mean, and $\sigma^2 = E\left[R^2(n)\right] - \mu^2$ is the variance of a return process. Note that *expected return* μ, and *volatility* σ are the two widely used metrics in finance. Correlation between asset returns plays an important role especially in portfolio selection, hedging, and pairs trading. In Markowitz's MPT [1], the optimum investment amount in each asset of an N-asset portfolio is obtained by solving the following minimization problem [2]

$$\mathbf{q}^* = \underset{\mathbf{q}}{\operatorname{argmin}} \frac{1}{2}\mathbf{q}^{\mathrm{T}}\mathbf{C}\mathbf{q} + \lambda_1\left(\mu_T - \mathbf{q}^{\mathrm{T}}\boldsymbol{\mu}\right) + \lambda_2\left(1 - \mathbf{q}^{\mathrm{T}}\mathbf{1}\right), \tag{4}$$

where $\mathbf{q} = \left[\begin{array}{cccc} q_1 & q_2 & \cdots & q_N \end{array}\right]^{\mathrm{T}}$ is the $N \times 1$ investment vector, q_i is the investment proportion of the total capital into the i*th* asset, T is the matrix transpose operator, $\mathbf{C} = [C_{ij}] = E\left[R_i(n)R_j(n)\right] - \mu_i\mu_j$ is the covariance matrix, λ_1 and λ_2 are two Lagrangian multipliers corresponding to the constraints $\mathbf{q}^{\mathrm{T}}\boldsymbol{\mu} = \mu_T$ and $\mathbf{q}^{\mathrm{T}}\mathbf{1} = 1$, respectively, μ_T is the target expected portfolio return, $\boldsymbol{\mu} = \left[\begin{array}{cccc} \mu_1 & \mu_2 & \cdots & \mu_N \end{array}\right]^{\mathrm{T}}$ is the $N \times 1$ expected return vector, $\mu_i = E\left[R_i(n)\right]$ is the expected return for the i*th* asset, and $\mathbf{1}$ is an $N \times 1$ vector with its elements all equal to one. Note that the covariance matrix $\mathbf{C}$, cross-correlations between returns of the assets in the portfolio, has a direct impact on the portfolio selection problem stated in (4).

Another application that the cross-correlation of asset returns comes into play is *hedging*. In this type of trading, returns of an asset in time are regressed on the returns of another asset (usually an exchange traded fund (ETF) that tracks the index of the corresponding industry) such that

$$R_1(n) = \alpha + \beta R_2(n) + \xi(n). \tag{5}$$

where α, β, and ξ are commonly referred to as drift, systematic component, and idiosyncratic component, respectively. According to the model given in (5), return of the first asset is composed of a constant drift, β times the return of the second asset, and white noise. The idea in hedging is to invest \$1 in the first stock and invest -\$$\beta$ in the second stock (hedge) such that the *return on investment* becomes

$$R_I(n) = R_1(n) - \beta R_2(n) = \alpha + \xi(n), \tag{6}$$

and the expected return on investment is calculated as $\mu_I = E[R_I(n)] = E[\alpha] + E[\xi(n)]$. This strategy is (i) profitable if $E[\alpha] > 0$ with $E[\xi(n)] = 0$ (ii) expected to profit if $E[\alpha] = 0$ and

$$X(n) = \sum_n \xi(n) \ll E[X(n)],$$

since $X(n)$ is expected to return to its mean, and expected to provide more positive samples than negative samples of $\xi(n)$. The latter is called *pairs trading*. It isolates the return on investment from the market (industry) and bets against the excess returns of a specific asset. By assuming $E[R_2(n)\xi(n)] = 0$ and substituting (5) into (3) we obtain

$$\beta = \rho\sigma_1/\sigma_2. \tag{7}$$

It is seen from (6) and (7) that the performance of a pairs trading strategy is related to the cross-correlation coefficient of two assets, ρ. Therefore, better correlation estimation of asset returns is an important factor of good performance.

III. EPPS EFFECT

According to Black-Scholes price model [7] log-price, $S(n) = \ln P(n)$, of an asset follows a geometric Brownian motion path as expressed

$$S(n) = S(n-1) + G(n), \tag{8}$$

where $G(n) \sim \mathcal{N}(\mu, \sigma^2)$ is the log-return. $G(n)$ is a stationary noise process and white, i.e.

$$E[G(n-k)G(n-l)] - \mu^2 = \sigma^2\delta_{k-l}, \tag{9}$$

where δ_{k-l} is the Kronecker delta function as given

$$\delta_{k-l} = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases},$$

Moreover, for small values of $R(n)$, return process of (1) is an approximation to the log-return due to the Taylor series expansion of the logarithm, i.e.

$$\begin{aligned} G(n) &= \ln[P(n)/P(n-1)] \\ &\cong [P(n)/P(n-1)] - 1 \\ &= R(n). \end{aligned} \tag{10}$$

It is seen from (8) that we can write the log-price at discrete time n as a sum of initial log-price and all log-returns up to n as follows

$$S(n) = S(0) + \sum_{i=1}^{n} G(i). \tag{11}$$

Now, consider two discrete-time log-prices of the same asset, $S_{T_1}(n)$ and $S_{T_2}(n)$ sampled with sampling periods T_1 and T_2 according to (2), respectively. Let $T_2 = kT_1$. From (8) and (11) we write

$$\begin{aligned} G_{T_2}(n) &= S_{T_2}(n) - S_{T_2}(n-1) \\ &= S_{T_1}(kn) - S_{T_1}(kn-k) \\ &= S_{T_1}(0) + \sum_{i=1}^{kn} G_{T_1}(i) - S_{T_1}(0) - \sum_{i=1}^{k(n-1)} G_{T_1}(i) \\ &= \sum_{i=0}^{k-1} G_{T_1}(kn-i). \end{aligned}$$

where $G_T(n)$ is the log-return obtained from the log-price $S_T(n)$ that is sampled with period T. Note that if $G_{T_1}(n) \sim \mathcal{N}(\mu, \sigma^2)$, then $G_{T_2}(n) \sim \mathcal{N}(k\mu, k\sigma^2)$. Using (10) and assuming that the mean of log-returns is zero, i.e. $\mu = 0$, we rewrite (3) as a function of the sampling period as

$$\rho_T = \frac{E[G_{T,1}(n)G_{T,2}(n)]}{\sigma_{T,1}\sigma_{T,2}}, \tag{12}$$

where we assume that $E[G_{T,1}(n-k)G_{T,2}(n-l)] = \rho_T\sigma_{T,1}\sigma_{T,2}\delta_{k-l}$, i.e. the cross-correlation between different time periods of different asset log-returns is zero. Given (9), (12), and $\sigma_{T_2} = \sqrt{k}\sigma_{T_1}$ we conclude that

$$\rho_{T_1} = \rho_{T_2}, \tag{13}$$

It is shown that the cross-correlation coefficient, ρ, between the returns of two assets that follow geometric Brownian motion paths with pure Gaussian increments is *no*t related to the sampling period. However, Epps [6] was the first to show that the empirical data does not comply with (13). Moreover, Epps stated that the cross-correlation between two financial assets decreases as the sampling period gets smaller, i.e.

$$\rho_T \to 0 \text{ as } T \to 0.$$

Cross-correlation coefficient between the log-returns of Apple Inc. stock (AAPL) and PowerShares QQQ Trust ETF (QQQ) as a function of time is displayed in Fig. 1.a. The sample correlation estimator employed for Fig. 1 is written as

$$\hat{\rho}_T = \frac{1}{W}\sum_{i=0}^{W-1} \bar{G}_{T,1}(n-i)\bar{G}_{T,2}(n-i), \tag{14}$$

where $\bar{G}_T(n)$ is the normalized log-return with zero mean and unit variance as calculated

$$\begin{aligned} \bar{G}_T(n) &= [G_T(n) - \hat{\mu}_T]/\hat{\sigma}_T \\ \hat{\mu}_T &= \frac{1}{W}\sum_{i=0}^{W-1} G_T(n-i) \\ \hat{\sigma}_T^2 &= \frac{1}{W}\sum_{i=0}^{W-1} [G_T(n-i) - \hat{\mu}_T]^2, \end{aligned} \tag{15}$$

and W is the measurement window size. Since AAPL is an important member of NASDAQ100 index and QQQ mimics the behavior of the index, one expects to have a significant correlation between the returns of these two relevant assets.

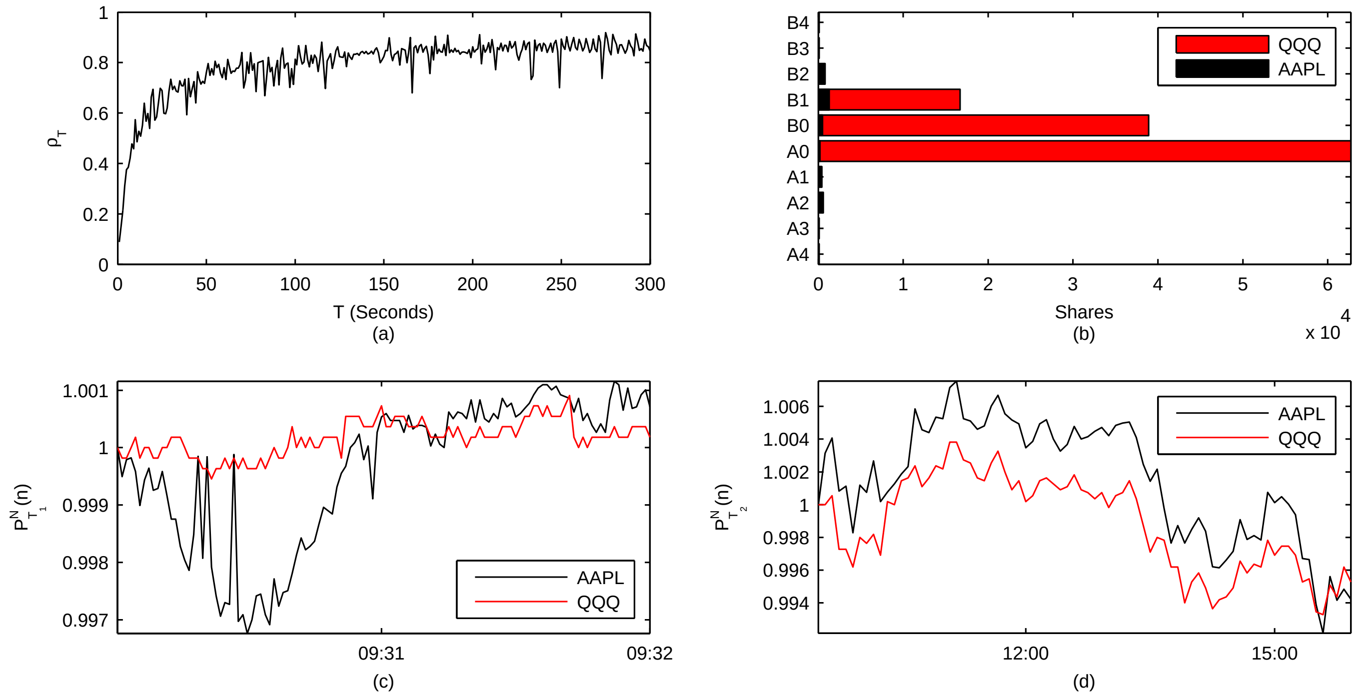

Figure 1. (a) Cross-correlation between AAPL and QQQ stocks estimated using 60 days of historical data between Apr 1, 2010 and Jun 30, 2010 as a function of sampling period, T, (b) A typical snapshot of the first five levels of the orderbooks (on both bid and ask sides) for AAPL and QQQ, Normalized last traded prices of both stocks on Mar 17, 2011 (c) between 9:30am and 9:32am sampled with $T = 1s$, (d) between 9:30am and 4:00pm sampled with $T = 300s$.

However, Fig. 1.a suggests a more complicated information. It is observed from the figure that assumption of (13) does not always hold, and the Black-Scholes model for the log-price of the assets (8) needs to be improved. This concern has been an active research topic where several people forwarded improved models (see [3, 4] and references therein.) Some researches have also proposed improved versions of estimators given in (14) and (15) [5].

The most widely accepted cause of the Epps effect in finance is the nature of *asynchronous trading*. Although prices of the assets within the same industry tend to behave similarly, and they respond to various intra-day economical and socio-political news in the same way, they are not traded at the same time points. The built-in asynchronicity affects correlation calculation since even the price of one asset of the pair does not change, the product of returns becomes zero, and that term is considered in the averaging operator. Similarly, a situation where the constant prices of both assets traded synchronously are also considered as perfectly uncorrelated pair of returns within this framework. Although these two zero-product cases for correlation calculations are distinct, it is more likely not to have price change at smaller time intervals. It is reasonable to expect higher correlation if correlation calculation only considers nonzero products with irregular sampling grid where price variations occur for both assets. Moreover, even though they were traded at the same time, their market structures, i.e. order books, volume and liquidity, are different and they

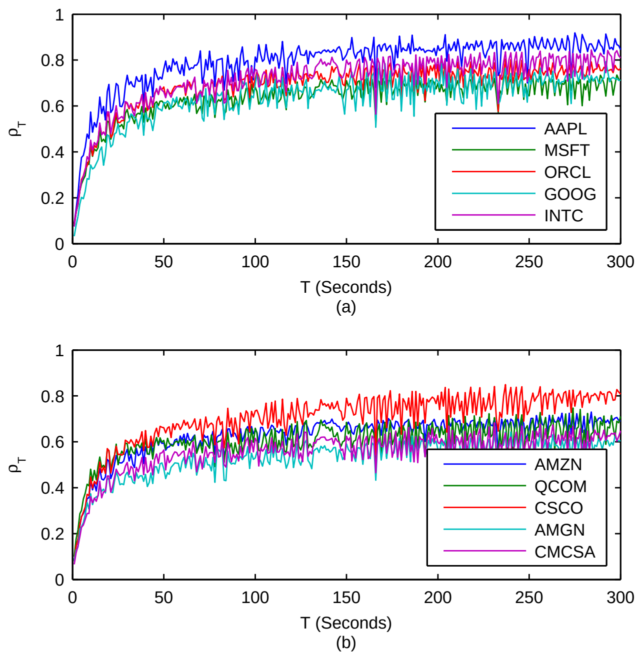

Figure 2. Cross-correlation between the QQQ ETF and (a) its first five largest holdings (AAPL: Apple Inc., MSFT: Microsoft Corp., ORCL: Oracle Corp., GOOG: Google Inc., INTC: Intel Corp.,) (b) its second five largest holdings (AMZN: Amazon.com Inc., QCOM: Qualcomm Inc., CSCO: Cisco Systems Inc., AMGN: Amgen Inc., CMCSA: Comcast Corp.) estimated using 60 days of historical data between Apr 1, 2010 and Jun 30, 2010 as a function of sampling period, T.

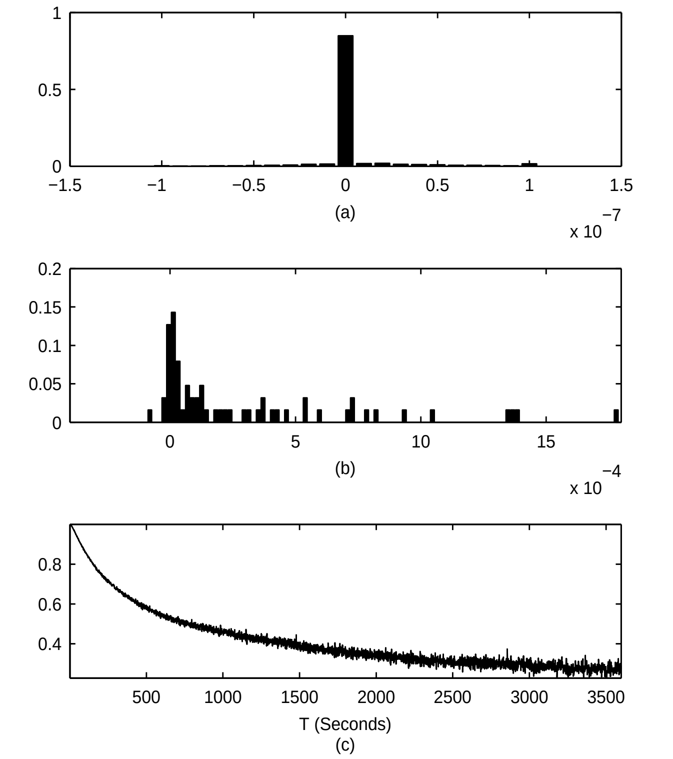

Figure 3. Histogram of pairwise products for the returns of AAPL and QQQ with sampling intervals (a) $T = 1s$, and (b) $T = 6.5h$ (EOD) (c) Probabilities of product terms being zero (between $-\epsilon$ and ϵ) as a function of sampling time interval, T, with $\epsilon = 3 \cdot 10^{-6}$. The data length is of 60 days between Apr 1, 2010 and Jun 30, 2010.

play a significant role in price formation. In our example, AAPL is not a liquid asset in the sense that the levels of the orderbook for AAPL do not offer too many shares available for selling or buying. However, QQQ is highly liquid compared to AAPL. A typical snapshot of the first five levels (on both bid and ask sides) of the orderbooks for AAPL and QQQ are displayed in Fig. 1.b. Since there are many QQQ shares resting at best bid and best ask prices, it is less likely for QQQ to provide a different last traded price print in short term than it is for AAPL. Normalized last traded prices, i.e. $P^N(n) = P(n)/P(0)$, of both stocks on Mar 17, 2011 between 9:30am and 9:32am with $T = 1s$, and between 9:30am and 4:00pm with $T = 300s$ are shown in Figs. 1.c and 1.d, respectively. It is observed from these figures that the good proxy between the prices of two stocks that exists at lower sampling rates (trading frequencies) disappears as the sampling rate increases. This is a very important phenomenon and a serious concern in particular for high-frequency trading since the traditional risk management framework does not hold to provide practical solutions in order to maintain an investment portfolio at high speeds. It is still an open research problem.

Holdings of QQQ are comprised of NASDAQ 100 technology stocks with their relevant investment factors. Hence, QQQ and those stocks are expected to be correlated. Correlation coefficients between the log-returns of QQQ and its first and second largest five holdings are displayed in 2.a and 2.b, respectively, as a function of sampling periods. It is observed from these figures that the correlation pattern and Epps Effect observed in 1.a. are not specific to AAPL and QQQ pair. The pairwise stock correlations are widely utilized in risk management methods and electronic trading algorithms. Therefore, their calculation and proper use in financial systems is of crucial importance.

Fig. 3.a displays histogram of pairwise product operations in the correlation calculation of AAPL and QQQ pair at $1s$ sampling period. Similarly, Fig. 3.b displays histogram for the case of EOD correlation. Fig. 3.c depicts probabilities of product terms being negligible (between $-\epsilon$ and ϵ), and used in the correlation calculations as a function of sampling period for $\epsilon = 3 \cdot 10^{-6}$. It is noted that the probability of having negligible product term in correlation calculation drops when sampling (trading) interval increases as depicted in this figure. This fact has direct impact on the values of pairwise correlations calculated through averaging that includes negligible ones. Hence, they drop significantly at higher sampling frequencies. This point needs further study.

IV. Conclusions

In this paper, we revisited Epps effect and pairwise correlation calculations of asset returns. The latter is a widely used metric in finance and plays a fundamental role in portfolio optimization employing Markowitz's MPT, in hedging, and pairs trading. We offered several insights with supporting evidence to explain Epps effect. We expect further improvements in modeling of pairwise correlations for assets in an investment portfolio in order to develop more sophisticated hedging methods in the future.

References

[1] H. M. Markowitz, *Portfolio selection: efficient diversification of investments*. Wiley, New York, 1959.

[2] M. U. Torun, A. N. Akansu, and M. Avellaneda, "Portfolio risk in multiple frequencies," *IEEE Signal Processing Magazine, Special Issue on Signal Processing for Financial Applications*, Sep 2011.

[3] M. U. Torun and A. N. Akansu, "On basic price model and volatility in multiple frequencies," in *Proc. IEEE Workshop on Statistical Signal Processing*, Jun 2011.

[4] E. Bacry, S. Delattre, M. Hoffmann, and J. Muzy, "Modeling microstructure noise using Hawkes processes," in *Proc. IEEE International Conference on Acoustics Speech and Signal Processing*, May 2011.

[5] L. Zhang, "Estimating covariation: Epps effect, microstructure noise," *Journal of Econometrics*, vol. 160, no. 1, pp. 33 – 47, 2011.

[6] T. W. Epps, "Comovements in stock prices in the very short run," *Journal of the American Statistical Association*, vol. 74, no. 366, pp. 291–298, 1979.

[7] F. Black and M. S. Scholes, "The pricing of options and corporate liabilities," *Journal of Political Economy*, vol. 81, no. 3, pp. 637–54, Jun 1973.