HENRYK GURGUL*, ARTUR MACHNO*

AGH University of Science and Technology, Poland

Asynchronous trading and Epps effect

Abstract:

Keywords: Market microstructure; Epps effect; nonsynchronous trading; lead-lag relationships; correlation estimation; covariance estimation; realized variance

JEL Classification: F36, G15

1. Introduction

The examination of dependence between stocks is one of the crucial problems in finance with respect to achievement of theoretical understanding of market structure and to schedule its relevant applications among others portfolio optimization. The collections of financial high frequency data allows for the correlation estimation on any time scales, down to the frequency of individual transactions. The dependence of correlations between stock prices on sampling frequency of time series is the phenomenon called Epps effect. The Epps effect has received considerable attention, not only from economists but also from mathematicians and theoretical physicists.

Epps effect may be unexpected on the first glance, however, it is simply explainable by asynchrony of trade times. For example, 5-minute returns are in fact returns for last trade prices before the appropriate times. Very short horizons or not liquid assets cause cases where some returns are calculated for the same trade, those returns are clearly zero and give no additional information to correlations. The situation where there is no traded in a time interval is utmost, however, the asynchrony causes Epps effect under any circumstances.

Financial transactions on stock markets play two roles. First of all they cause returns or price movements. They supply information about current prices. The second role of transactions is to assure fixing of the market value of a traded stocks until the next trade takes place. Therefore from this point

* Correspondence: AGH University of Science and Technology, Department of Applications of Mathematics in Economics, 10 Gramatyka street, 30-067 Krakow, Poland.
E–mails: henryk.gurgul@gmail.com; artur.machno@gmail.com

of view in financial literature are considered two main groups of contributions to the Epps effect. The first one relates the Epps effect to actual lagged correlations. The source of this phenomenon is a temporary or permanent effect of individual trades on the price dynamics. The second group of authors is concerned with the fact that price dynamics is not synchronous across stocks. This means that transactions are conducted at different times for different shares.

In this paper we propose the easy nonparametric method for removing the asynchrony effect on the correlation estimation for high frequency data. We apply the method to four chosen Polish stocks quoted on Warsaw Stock Exchange, which gives us six pairs for which we the perform correlation analysis. Our method requires some assumptions, which one can think of as simplifications. We inspect how the method works for the chosen time series and if it explains the Epps effect.

The content of this paper is as follows. Next section reviews most important contributions concerning the Epps Effect. In the third section main conjectures are presented. In the following fourth section the methodology is showed. Fifth section is concerned with dataset and empirical results. Finally, in the last section we summarize major conclusions and suggest directions for future research.

2.Literature overview

Epps (1979) was the first who supplied empirical evidence of a dramatic drop in correlations between stocks when decreasing the sampling time horizon. This observation has been noticed across different markets. The empirical evidence is reported e.g. by Bonanno et al. (2001), Zebedee (2009), Zebedee and Kasch-Haroutounian (2009) for stock prices and by Lundin et al. (1998) and Muthuswamy et al. (2001) for foreign exchange rates. However, Andersen et al. (2001a) and Andersen et al. (2001b) in their empirical studies, being concerned with stock prices and foreign exchange rates, respectively, detected that correlations are still significant when even it is computed using five minute returns.

In early literature, it is assumed that two main statistical features of the data may produce Epps effect: asynchronous trading and lead-lag relationships. The impact of asynchronous data on covariance measurement has been widely studied; two early prominent examples are here to mention: Lo and MacKinlay (1990) and, formerly, Scholes and Williams (1977). In addition to this, it is suggested that the Epps effect may depend on the fact that correlations are lagged. Therefore, when reducing the sampling frequency under time scales comparable to this lag, the correlation is underestimated.

Although, there is well known that lagged correlations and asynchronous sampling contribute to the Epps effect, the relative weight of these two effects is not always easy to assess (see Reno (2003) and Large (2007)). The first fact to be considered is the observation that trading is not synchronous.

Therefore, covariance estimation is intrinsically problematic at high frequency, as Scholes and Williams (1977) pointed.

The fundamental goal of the reviewed contributions was the assessment of the following effects (Reno (2003)):

・ What is the relative magnitude of the impact of asynchronous trading and lead-lag relationships?

・ Are these two features sufficient to explain the Epps effect?

Reno (2003) tried to address his questions by looking at market data. The first data sample under study was the collection of DEM-USD and JPY-USD exchange rate. The second data set under study consisted of the high frequency trades of the stock prices of Exxon and Mobil, from 9:30 to 16:00, from January 1995 to April 1995 for a total of 82 trading days at the NYSE. Reno (2003), based on his Monte Carlo experiments, concludes that even if other factors, apart from non-synchronicity and lead-lag relationships, concur in the Epps effect, their impact is negligible. In his opinion it is questionable if these factors can be explained in the framework of continuous-time models.

In the economic literature, one can find contradictory answers to these two questions. Zebedee (2009) claims that the Epps effect is mainly due to the lead-lag relationship. He argues that as frequency rises, correlation is shifting to other nearby time intervals. However, Lundin et al. (1998) represent other point of view. In their opinion different actors play different roles at different frequencies. Therefore it is not possible to recover the same correlation at different time scales. However, the authors claim that it is not obvious which type of price formation process is source of the Epps effect. In addition, Lundin et al. (1998) detected a significant inverse link between correlation and activity. They found that the growing quantity of assets traded reduces the Epps effect. This observation is in favor of the point of view that the role of synchronicity in explaining the correlation decrease at higher frequencies is dominant.

Burns *et al.* (1998), Martens and Poon (2001), taking into account the fact that daily stock prices are usually at slightly different closing hours, detected the impact of non- synchronicity on the correlation estimates of daily stock prices. They found that this impact, which may look negligible at first glance, can significantly impact the correlation estimates.

The knowledge of financial correlations and their behavior on time scale implies important practical consequences for market participants with respect to portfolio management. Pafka et al. (2007) suggested that for large portfolios, the estimation of risk measures at low frequency (e.g. one day) is biased due to instabilities, caused by the scarcity of data. The authors argued that estimates of financial correlations (which are related to risk) at high frequency is based on much longer time series. Therefore researchers are able to find structural changes in the process under study more reliable and accurately. High frequency data allow to compare the structure of correlations at longer time scales to that at shorter time scales. Long time series of high frequency data enable overcoming the information deficiency. The last phenomenon is main source of the instability of risk measures. However, Borghesi et al. (2007) detected that the structure of correlations in groups of very liquid stocks, is largely invariant across time scales ranging from 5 min to one day. This finding implies that estimates of correlations on long time scales on the basis high frequency correlations can in general succeed.

In the opinion of many contributors a number of financial mechanisms of stock price and trading volume formation is source of the Epps effect. In more recent contribution by Colon et al. (2009) the Epps effect is established using multivariate method and analyzed at longer scales than previously studied. A partition of the eigenvalue time series demonstrated, at very short scales, the emergence of negative returns when the largest eigenvalue is greatest. Finally, a portfolio optimization showed the importance of time–scale information in the context of risk management. Mastromatteo et al. (2011) identified two major causes of purely statistical origin. Their goal was not to derive a complete description of the Epps effect. The contributors stressed that they are rather aimed at identifying statistical causes that can be compensated directly, without the requirement of adjusting parameters, model calibrations or an optimal sampling frequency. They identified two major causes: the asynchrony of the time series and the impact of the discretisation by the tick-size.

3. Main conjectures

One of main contribution to Epps effect is due to asynchronous trading. I has been reported that the observed correlation consists of the actual correlation. The values of measures of this correlation one could calculate if which would be observed if prices are quoted continuously. The uncorrelated part is induces by asynchronous trading.

Conjecture 1. The effect of asynchronous trading on correlation can be justified and measured theoretically.

Our study concerns effect of asynchronous trading on size of correlation both from the theoretical and empirical point of view. The main aim is to quantify this effect for data from Warsaw Stock Exchange. Taking into account results reported in the literature for other stock exchanges we expect that:

Conjecture 2. The asynchronous trading has significant impact on correlation pattern
and is significant source for Epps effect.

Previous analysis consider several possible causes of the Epps effect. Under the condition that conjecture 2 holds true, we expect:

Conjecture 3. The asynchronous trading does not explain fully correlation pattern and Epps effect.

The model underlining our investigation is outlined in the next section.

4. Model for the compensation.

In this section we briefly present the method for correlation correction. Asynchronous times of trading effect on correlation is modeled using overlapping time intervals between trades of considered stocks. The goal is to determine the impact of this factor on the observed correlation between time series based on the intraday data such as 1- or 5-minute returns.

Our approach is based on the assumption that there exists the price of an asset at any time, however, it is observable only at the times of trade. Empirical study support this statement(cytowania), in fact observed Epps effect can suggest it. The series of returns from intervals of a certain length is in fact the series of returns from intervals between the time of the last trade before the original interval start to the time of the last trade before the original interval ends.

The analysis is based on [(18)]. The main concept of the approach is that the observed correlation consists of the real correlation (the coefficient which would be observed if prizes were quoted continuously) and the uncorrelated part which is present because of asynchronous trading.

Let $\tilde{r}^i(t)$ be the unobserved process of the *i*-th asset's real return at the time *t*, an advantage of this approach is that we do not need to specify nor estimate the model for the process. The model for the return process can be set either in discrete time or in continuous time. Additionally, let $r^i_{\Delta t}(t)$ be the observed return. Formally:

$$r^i_{\Delta t}(t) \coloneqq \frac{S^i(t+\Delta t) - S^i(t)}{S^i(t)}; \tag{a}$$

where $S^i(t)$ is the observed price of the *i*-th asset and Δt is the time interval for which we perform the analysis. In the sake of simplicity, we analyze normalized returns $\tilde{g}^i(t)$ and $g^i_{\Delta t}(t)$ with zero mean and unit variance. Formally:

$$\tilde{g}^i(t) \coloneqq \frac{\tilde{r}^i(t) - \langle \tilde{r}^i \rangle}{\tilde{\sigma}^i}; \tag{b}$$

$$g^i_{\Delta t}(t) \coloneqq \frac{r^i_{\Delta t}(t) - \langle r^i_{\Delta t} \rangle}{\sigma^i_{\Delta t}}; \tag{bb}$$

where $\tilde{\sigma}^i$ is the theoretical standard deviation of the return process; $\langle \dots \rangle$ means the average over the analyzed period $[0, \mathrm{T}]$ and $\sigma^i_{\Delta t}(t)$ is the realized standard deviation of the values $r^i_{\Delta t}(t)$.

We present the case of discrete time, because real data is discrete. However, rewriting the methodology into continuous time is straight. Let $\Delta\tilde{t}$ be the smallest observed time interval (e.g. 1 second). For the *i*-th asset, we define the time when the last trade before the time *t* occurred by $\gamma^i(t)$. When calculating the *i*-th asset's return in the interval $[\mathrm{t}, \mathrm{t}+\Delta t]$, one actually calculate the return in the interval $[\gamma^i(\mathrm{t}), \gamma^i(\mathrm{t}+\Delta t)]$. Hence, when calculating the correlation between two assets, one actually calculates correlation between numbers which correspond to returns in overlapping intervals, the overlapping interval for returns $r^1_{\Delta t}(t)$ and $r^2_{\Delta t}(t)$ equals $[\max\{\gamma^1(\mathrm{t}), \gamma^2(\mathrm{t})\}, \min\{\gamma^1(\mathrm{t}+\Delta t), \gamma^2(\mathrm{t}+\Delta t)\}]$. Let $\Delta t_0(t)$ stands for the length of the time interval of the actual overlap:

$$\Delta t_0(t) := \min\{\gamma^1(\mathrm{t}+\Delta t), \gamma^2(\mathrm{t}+\Delta t)\} - \max\{\gamma^1(\mathrm{t}), \gamma^2(\mathrm{t})\}. \tag{d}$$

Assuming that asynchronous parts of returns $r^1_{\Delta t}(t)$ and $r^2_{\Delta t}(t)$ are uncorrelated to each other and to the synchronous part of the other's asset return, we have:

$$\mathrm{corr}\left(r^1_{\Delta t}, r^2_{\Delta t}\right) = \frac{1}{T}\sum_{j=1}^{T} \mathrm{corr}_{t_j}\left(\tilde{g}^1, \tilde{g}^2\right)\frac{\Delta t_0(t_j)}{\Delta t}; \tag{e}$$

where t_j is the beginning of the *j*-th interval of the length Δt in the analyzed period $[0, \mathrm{T}]$. The coefficient $\mathrm{corr}\left(r^1_{\Delta t}(t), r^2_{\Delta t}(t)\right)$ is what we observe, however our goal is to extract the correlation for the unobserved series.

Under the assumption that $\Delta t_0(t_j) \neq 0$, we obtain the estimated corrected correlation $\mathrm{corr}_c(r^1_{\Delta t}(t), r^2_{\Delta t}(t))$ as:

$$\mathrm{corr}_c\left(r^1_{\Delta t}, r^2_{\Delta t}\right) = \frac{1}{T}\sum_{j=1}^{T} g^1_{\Delta t}(t) g^2_{\Delta t}(t)\frac{\Delta t}{\Delta t_0(t_j)}. \tag{f}$$

In fact, the assumption that $\Delta t_0(t_j) \neq 0$ can be omitted and the formula (f) takes slightly different form. However, obtaining formula (e) need some additional assumptions. The following theorem summarizes the estimate for the corrected correlation.

Theorem 4.1.
Notation:
Let, for $i = 1,2$, $\{r^i_{\Delta t}(t)\}_{t=0,\Delta t,\ldots,T}$ be observed return's time series for asset i an time interval Δt; let $\{\tilde{r}^i(t)\}_{t=0,\Delta\tilde{t},\ldots,T}$ be underlying true return's time series, where $\Delta\tilde{t}$ is a unit of time. Let $\tilde{g}^i(t)$ and $g^i_{\Delta t}(t)$ be normalized returns defined in (b) and (bb).
Assumptions:

A1. Series $N^{i_1}_{\Delta t}(t)$ are independent to each other, and to series $\tilde{r}^{i_2}(t)$, $i_1, i_2 = 1,2$.

A2. Variables $\tilde{r}^1(t)$ and $\tilde{r}^2(t+t_1)$ are independent, for $t_1 \neq 0$.

Then the corrected correlation is calculated as:

$$\mathrm{corr}_c\left(r^1_{\Delta t}, r^2_{\Delta t}\right) = \frac{1}{T}\sum_{j=1}^{T} g^1_{\Delta t}(t) g^2_{\Delta t}(t)\frac{\sqrt{\langle N^1_{\Delta t}\rangle\langle N^2_{\Delta t}\rangle}}{\bar{N}(t_j)}; \tag{g}$$

where $N^i_{\Delta t}(t)$ is the number of $\Delta\tilde{t}$ intervals in the interval $[\gamma^i(\mathrm{t}), \gamma^i_+(\mathrm{t})]$, where $\gamma^i(\mathrm{t})$ is the last trade of the asset i before time t;

$$\gamma^i_+(\mathrm{t}) := \min_{\substack{\mathrm{j}=1,2,\ldots \\ \gamma^i(\mathrm{t}+\mathrm{j}\Delta t)\neq\gamma^i(\mathrm{t})}} \{\gamma^i(\mathrm{t}+\mathrm{j}\Delta t)\};$$

and $\bar{N}(t)$ is number of $\Delta\tilde{t}$ intervals in the interval $[\max\{\gamma^1(\mathrm{t}), \gamma^2(\mathrm{t})\}; \min\{\gamma^1(\mathrm{t}+\Delta t), \gamma^2(\mathrm{t}+\Delta t)\}]$.

Proof: in the Appendix.

Remark 4.1.
The assumption A1 in the Theorem x.1. is weaker than common (cytowania) assumption that the price process is independent to trade time process.

Remark 4.2.
The assumption A2 does not exclude autocorrelation in either price series and the lack of autocorrelation does not ensure A2. However, A2 resembles autocorrelation and is related to it.

5. Empirical study.

In this section, we present the data and perform some descriptive statistics. The application of the compensation methods presented in the section 1 to several pairs of stocks from the Warsaw Stock Exchange (WSE) is presented. Additionally, we briefly describe the WSE and the stocks chosen to the analysis.

5.1. **Description of data set.**

Our data consist of every operation taken place in WSE from 1.01.2001 to 22.09.2014. The electronic system of the WSE has been changed since 1.08.2013, from the perspective of our analysis the crucial change is the extension exactness of the trade time from seconds to microseconds. This precision is very convenient from the researcher's point of view, in the period where microseconds are present have not been two transactions taken place at the same time. The trade hours on the WSE are 9:00 a.m. to 5:05 p.m. However, there is break in 4:50 p.m. to 5:00 p.m. and during the last 5 minutes trade is not conducted as it is during normal hours. Thereat, we exclude transaction which took place after 4:50 p.m. It leaves us with 470 minutes of trade each day.

All transactions in WSE are included in Polish zloty (PLN). At the time this publication is being prepared 1USD=3.75PLN and 1EUR=4.15PLN. In the text we present all values in PLN.

For the simplicity and clarity, we have chosen four stocks which were traded most frequently in the period 1.01.2014-22.09.2014. These are:

KGHM Polska Miedz (KGHM)- the biggest mining and metallurgy company in Poland. It's main product is copper. The value of the company is about 19 billion PLN and it employs about 30 thousand employees.

Powszechna Kasa Oszczednosci Bank Polski (PKOBP)- the biggest Polish Bank. The value of the bank is about 36 billion PLN and it employs about 26 thousand employees.

Polska Grupa Energetyczna (PGE)- the biggest polish energy producer and energy provider. Polish government owns more than 60% of its shares. The value of the company is about 33 billion PLN and it employs about 41 thousand employees.

Powszechny Zaklad Ubezpieczen (PZU)- the biggest Polish financial institution. It is specialized in insurance and dominates Polish insurance market. The value of the company is about 37 billion PLN and it employs about 17 thousand employees.

5.1 Some descriptive statistics

In this subsection, we present some tables and plots in order to visualize the data. We have randomly chosen two days, namely February 19 and July 7, 2014. Figure 1 presents all trades of the four analyzed stocks in the example day. The stocks are very liquid, therefore, it is hard to see durations. By duration we mean time between trades.

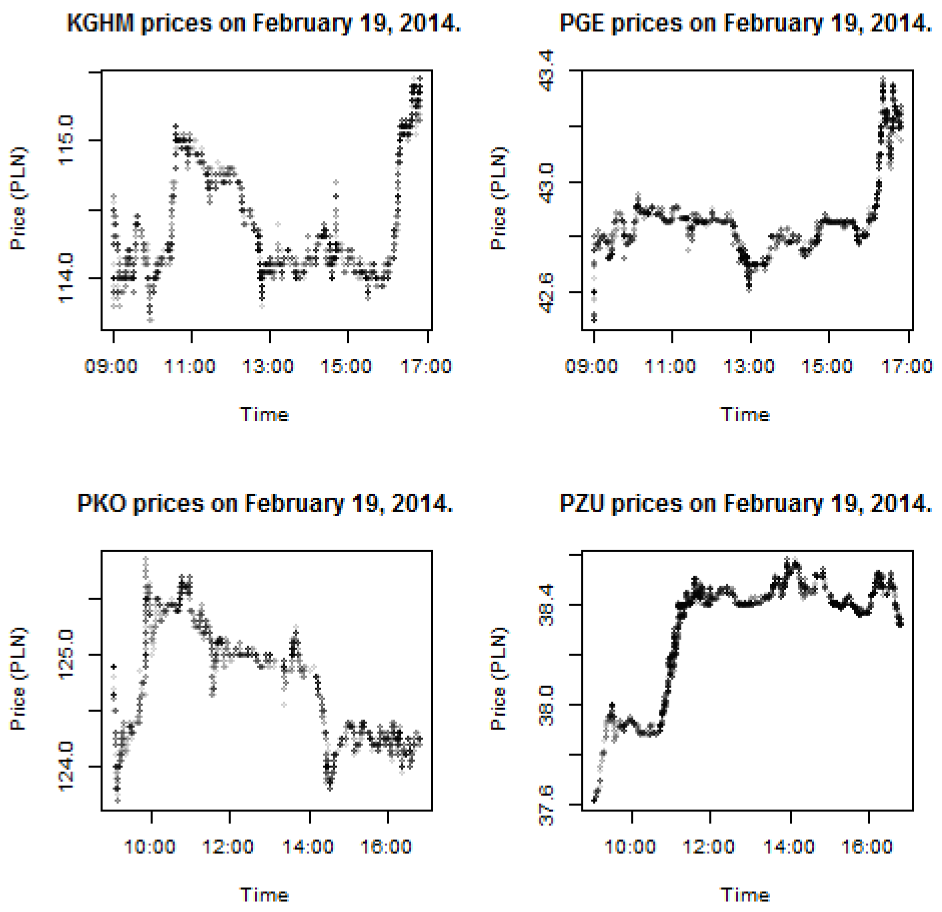

Figure 1. *All transactions of KGHM, PKO, PGE and PZU on Feb 19, 2014.*

In Table 1, we present number of trades of the chosen stocks and dates and summary statistics for duration. The daily pattern for the duration distribution is very interesting topic, especially on the market such as WSE which is strongly dependent on greater stock exchanges. However, we do not explore this topic in the paper thoroughly. At this point, we want to generally present data which we work with. In Table 1, we present number of trades of the chosen stocks and dates and summary statistics for duration. The statistics are quite similar for each stock and day. In fact minimums are the same (1 microsecond), which is probably caused by the electronic system. First quartiles varies from 4 to 6 microsecond which is extremely small interval and worth exploring how it is possible that the quarter of operations take place immediately after the preceding. Means and number of transactions are connected, a mean is a total trade time in the day (470 minutes, 28200 seconds) denominated by number of transactions. Third quartiles vary from 3.7 seconds to 18.2 second which is not very wide interval and maximums vary from 4.4 to 10.5 minutes. The only analyzed statistic which vary quite strongly for the analyzed datasets is median of the duration, namely from 42 to 849 milliseconds. The distribution of durations have “U” shape, most durations are either very low or relatively large. There is very little durations near median which may explain why median varies that strongly across stocks and days.

Table 1 Descriptive statistics for tick data

	Trades	Min.	Q_1	Median	Mean	Q_3	Max.

$KGHM_1$	2640	0.0000	0.0000	0.1681	10.680	7.660	266.40
$KGHM_2$	1990	0.0000	0.0000	0.1677	14.170	10.110	565.80
PKO_1	2886	0.0000	0.0000	0.0420	9.762	4.997	365.00
PKO_2	3261	0.0000	0.0000	0.0344	8.646	4.891	380.90
PGE_1	3352	0.0000	0.0000	0.0356	8.410	3.662	348.10
PGE_2	2135	0.0000	0.0000	0.0443	13.200	6.777	462.00
PZU_1	2102	0.0000	0.0000	0.7123	13.410	6.459	631.90
PZU_2	1251	0.0000	0.0000	0.8490	22.540	18.210	409.70

Note: The total number of transactions for a stock in particular day is in column "Trades". Descriptive statistics for time between trades are presented in seconds. Subscript 1 in the first column is for data on Feb. 19, 2014, subscript 2 for Jul. 7, 2014. Trades taken place between 9:00 and 16:50 each day are taken.

In the analysis of the Epps effect, we are particularly interested in the transformed data. We have transformed data into return series for chosen interval. The R code which process tick data into transformed data for whichever interval length is available, please contact author.

We have constructed plots in the sake of the visualization of transformed data. The main difference between the tick data and the transformed data is that the latter's length is fixed and labels (times) are fixed. In the sake of the visualization we present 1-minute data in Figure 2, and 10-minute data in Figure 3. The return is calculated as the difference between last trades before a certain time divided by the prior. The return in the first interval is calculated by taking the prior as the first transaction at a day. If there is no transactions in the last interval, we set a return to be zero, this situation occurs in narrow intervals, in the analyzed datasets the last trades took place 2 to 28 second before 4:50 p.m. Return series looks discrete for an asset which maximum and minimum prices at a day do not differ much.

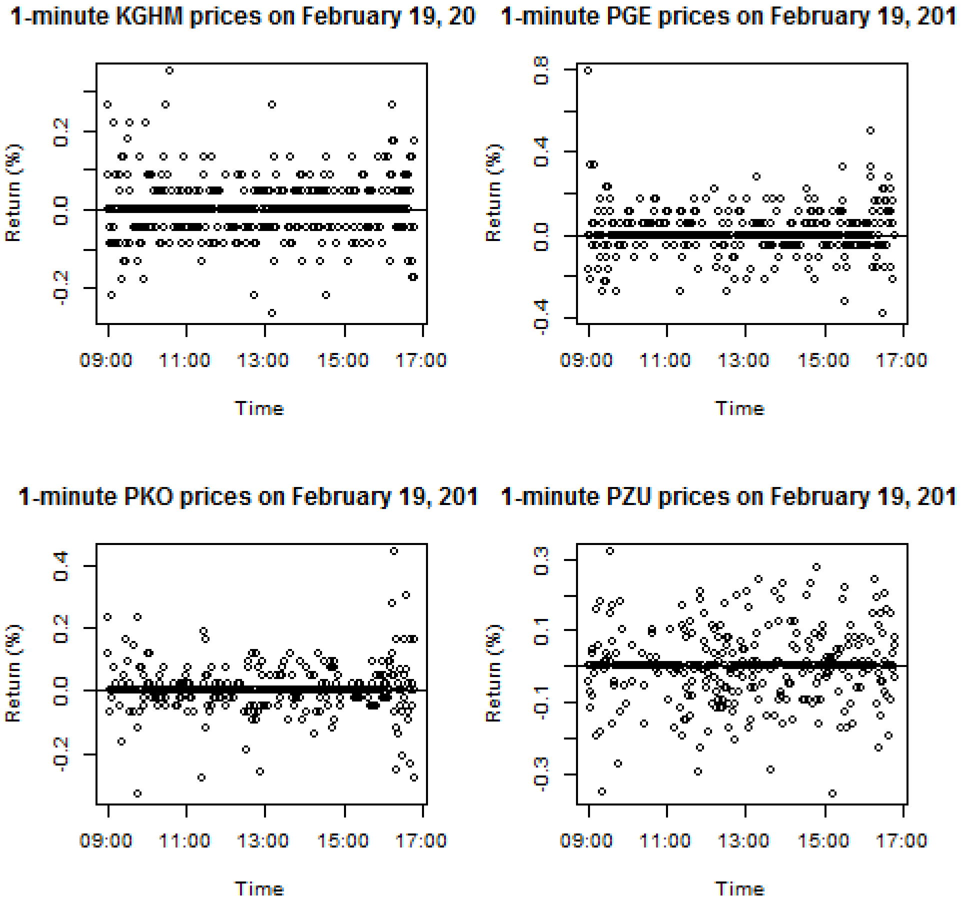

Figure 2. *1-minute returns of KGHM, PKO, PGE and PZU on Feb 19, 2014.*

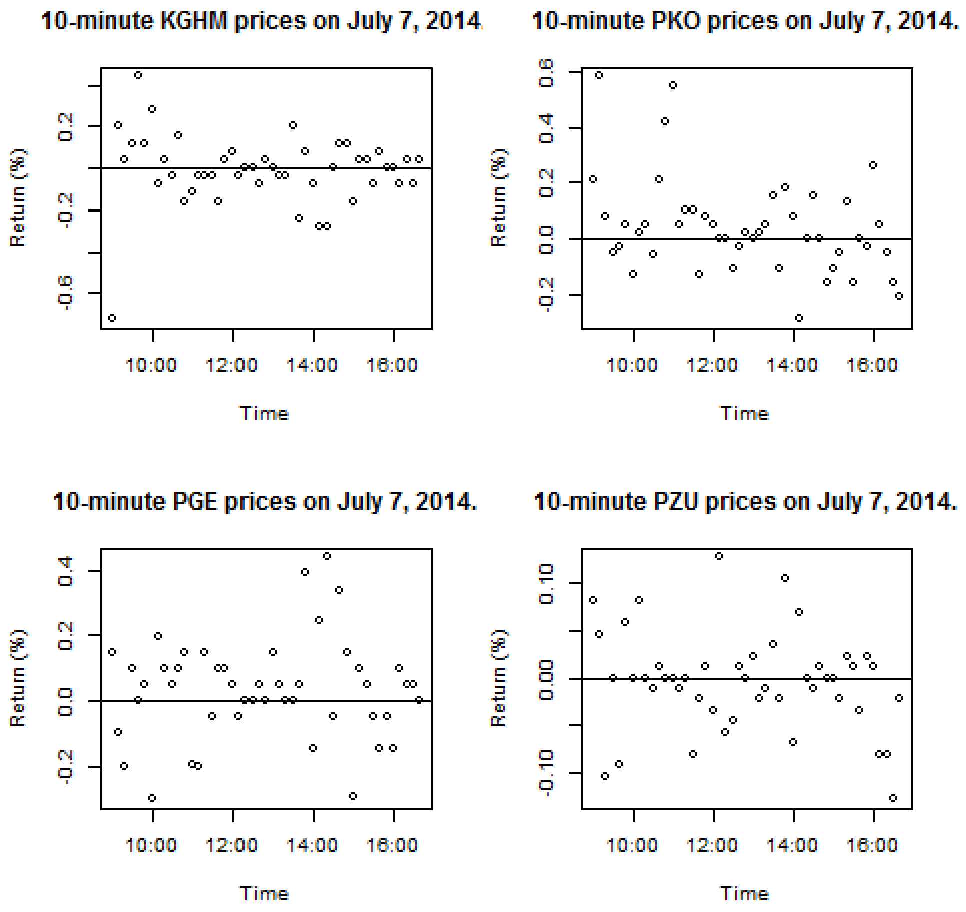

Figure 3. *10-minute returns of KGHM, PKO, PGE and PZU on Jul 7, 2014.*

Table 2 presents descriptive statisctics for one, five and ten minutes' interval data.

Table 2. Descriptive statistics for 1, 5 and 10-minute returns.

	Min.	Q_1	Median	Mean	Q_3	Max.	Sd.
$KGHM_1^{1min}$	-0.2622	-0.0435	0.0000	0.0022	0.0435	0.3487	0.0689
$KGHM_2^{1min}$	-0.4427	0.0000	0.0000	-0.0012	0.0000	0.2800	0.0616
PKO_1^{1min}	-0.3266	0.0000	0.0000	0.0028	0.0000	0.4402	0.0643
PKO_2^{1min}	-0.1560	0.0000	0.0000	0.0040	0.0260	0.2883	0.0487
PGE_1^{1min}	-0.3798	0.0000	0.0000	0.0087	0.0549	0.7861	0.0999
PGE_2^{1min}	-0.2910	0.0000	0.0000	0.0032	0.0000	0.4410	0.0677
PZU_1^{1min}	-0.3544	-0.0115	0.0000	-0.0010	0.0000	0.3185	0.0843
PZU_2^{1min}	-0.1735	0.0000	0.0000	-0.0005	0.0000	0.2778	0.0283
$KGHM_1^{5min}$	-0.4374	-0.0870	0.0000	0.0108	0.0872	0.4363	0.1331
$KGHM_2^{5min}$	-0.5204	-0.0699	0.0000	-0.0059	0.0800	0.4800	0.1360
PKO_1^{5min}	-0.3230	-0.0234	0.0000	0.0137	0.0467	0.8140	0.1320
PKO_2^{5min}	-0.2079	-0.0455	0.0000	0.0200	0.0780	0.3423	0.1040
PGE_1^{5min}	-0.6678	-0.0551	0.0544	0.0436	0.1105	1.2660	0.2313
PGE_2^{5min}	-0.3445	-0.0490	0.0000	0.0162	0.0984	0.3430	0.1229
PZU_1^{5min}	-0.3523	-0.1006	0.0000	-0.0048	0.0799	0.5469	0.1563
PZU_2^{5min}	-0.1158	-0.0231	0.0000	-0.0025	0.0116	0.1852	0.0459
$KGHM_1^{10min}$	-0.3061	-0.0874	0.0000	0.0215	0.0658	0.6999	0.1954
$KGHM_2^{10min}$	-0.7206	-0.0800	0.0000	-0.0118	0.0602	0.4419	0.1706
PKO_1^{10min}	-0.5075	-0.0467	0.0000	0.0275	0.0468	1.0250	0.2081
PKO_2^{10min}	-0.2853	-0.0524	0.0260	0.0399	0.0916	0.5836	0.1706
PGE_1^{10min}	-0.6125	-0.1100	0.0549	0.0872	0.2181	1.3760	0.3286
PGE_2^{10min}	-0.2956	-0.0485	0.0486	0.0324	0.0986	0.4390	0.1534
PZU_1^{10min}	-0.3399	-0.1487	-0.0114	-0.0097	0.0632	0.5813	0.1986
PZU_2^{10min}	-0.1274	-0.0232	0.0000	-0.0049	0.0116	0.1274	0.0520

Note: Descriptive statistics for returns. Subscript 1 in the first column is for data on Feb. 19, *2014, subscript 2 for Jul. 7, 2014. Trades taken place between 9:00 and 16:50 each day are taken. Superscript indicate the time interval length for the data*

5.1 Epps effect and correction

In this section, we present the visualization of the Epps effect and the impact of asynchronous trading times on it. We have transformed tick data into intraday data with intervals between 1 and 30 minutes by 1 minute. It is possible to transform data into intraday data of thinner intervals, however, it would not be more informative and it would be computationally heavy. Note that transformed data is larger for thinner time intervals, for example, there is more one minute returns that ten minute returns during the trading day.

The Epps effect is clearly visible among all six pairs of the stocks. Figure 4 consists of averaged correlations between analyzed stocks and averaged corrected correlations. The values corresponding to realized (not corrected) correlations are obtained as follow. For every trading day in the period 1/1/2014-9/22/2014, which is 182 days, we have transformed data into intraday data with the given interval. The length of the interval is the x-coordinate of the point. Then, we have calculated correlations between returns of appropriate stocks on every day. The average of those 182 correlations is the y-coordinate of the particular point.

The value corresponding to the corrected correlation is obtained analogously, instead of correlations between returns, values calculated using formula (g) is averaged.

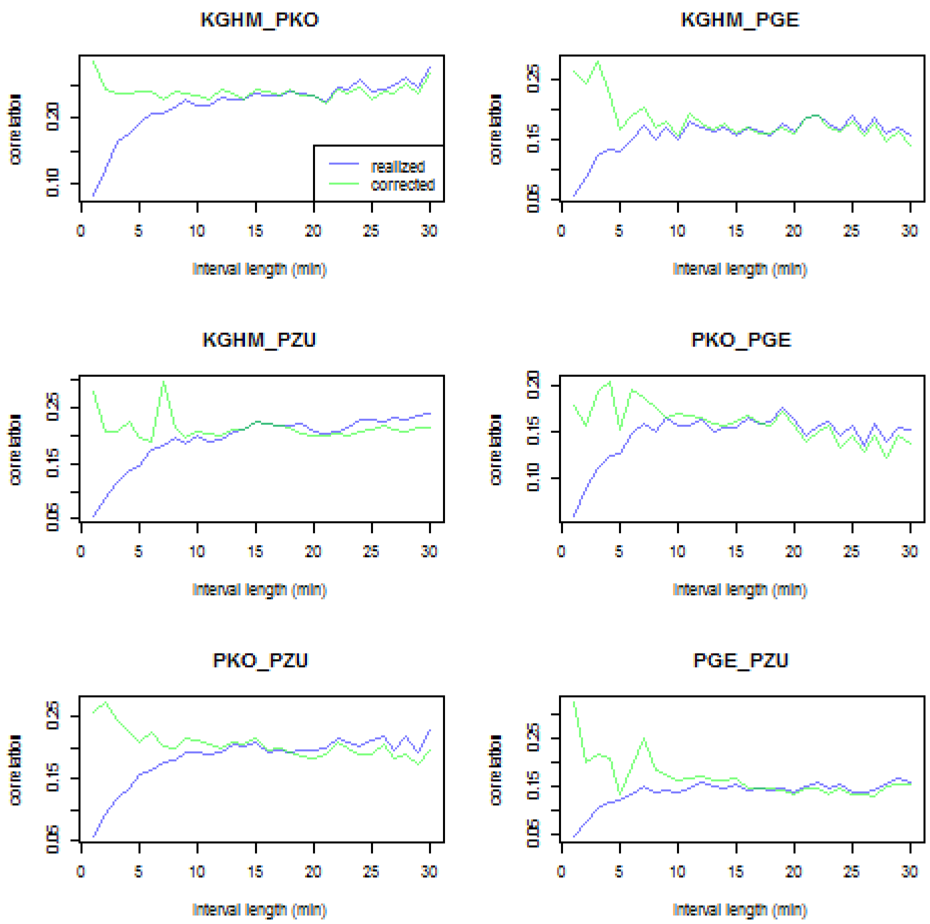

Figure 4. *Realized and corrected correlations between stock return according to data interval length.*

We see several interesting features on the Figure 4. The correction clearly works and the Epps effect is not visible after correction. However, it appears that corrected correlations possesses the opposite property. It appears that corrected correlation decreases for wider intervals, theoretically, the corrected correlation is constant under assumptions of Theorem 4.1. Naturally, the assumptions are not met, however, it would be expected that after the correction there is still visible smaller Epps effect caused by another factors such as discreteness of the data. Another disturbing fact is the high variability of the corrected correlations for thin intervals. I seems that the estimation strongly depends on the interval length, in another words, the estimated corrected correlations have high standard deviation.

Both, realized and corrected, correlations seems to have a cyclic component which is visible for wider intervals. It is caused by the last return in the series, note that the last return is calculated for interval which is not exactly of the length of the preceding ones. Therefore, it just a technical problem which can be solved by weighting the returns in the correlation computation or by just omitting the last return.

6. Concluding remarks

APENDIX

Proof of the theorem x.1:

Let us first calculate observed normalized process $g_{\Delta t}^{i}(t)$ in terms of the underlying process $\tilde{g}^{i}(t)$, where $i = 1,2$.

$$g_{\Delta t}^{i}(t) = \frac{r_{\Delta t}^{i}(t) - \langle r_{\Delta t}^{i}(t)\rangle}{\sigma_{\Delta t}^{i}} = \frac{\sum_{j=1}^{N_{\Delta t}^{i}(t)} \tilde{r}^{i}\left(\gamma^{i}(\mathrm{t}) + j\,\Delta\tilde{t}\right) - \langle r_{\Delta t}^{i}\rangle}{\sigma_{\Delta t}^{i}}$$

$$= \frac{\sum_{j=1}^{N_{\Delta t}^{i}(t)}\left(\tilde{\sigma}^{i}\tilde{g}^{i}\left(\gamma^{i}(\mathrm{t}) + j\,\Delta\tilde{t}\right) + \langle\tilde{r}^{i}\rangle\right) - \langle r_{\Delta t}^{i}\rangle}{\sigma_{\Delta t}^{i}}$$

$$= \frac{\sum_{j=1}^{N_{\Delta t}^{i}(t)}\left(\tilde{\sigma}^{i}\tilde{g}^{i}\left(\gamma^{i}(\mathrm{t}) + j\,\Delta\tilde{t}\right) + \langle\tilde{r}^{i}\rangle\right) - \langle r_{\Delta t}^{i}\rangle}{\sigma_{\Delta t}^{i}}$$

$$= \frac{\tilde{\sigma}^{i}\sum_{j=1}^{N_{\Delta t}^{i}(t)} \tilde{g}^{i}\left(\gamma^{i}(\mathrm{t}) + j\,\Delta\tilde{t}\right) + N_{\Delta t}^{i}(t)\langle\tilde{r}^{i}\rangle - \langle r_{\Delta t}^{i}\rangle}{\sigma_{\Delta t}^{i}}$$

Under the assumption A1:

$$\langle r_{\Delta t}^{i}(t)\rangle = \langle N_{\Delta t}^{i}\rangle\langle\tilde{r}^{i}\rangle;\ \left(\sigma_{\Delta t}^{i}\right)^{2} = \langle N_{\Delta t}^{i}\rangle\left(\tilde{\sigma}^{i}\right)^{2}$$

Thus,

$$g_{\Delta t}^{i}(t)=\frac{1}{\sqrt{\langle N_{\Delta t}^{i}\rangle}}\sum_{j=1}^{N_{\Delta t}^{i}(t)}\tilde{g}^{i}\left(\gamma^{i}(\mathrm{t})+j\,\Delta\tilde{t}\right)+\frac{\langle\tilde{r}^{i}\rangle\left(N_{\Delta t}^{i}(t)-\langle N_{\Delta t}^{i}\rangle\right)}{\sigma_{\Delta t}^{i}}$$

Let us now calculate the observed correlation $\mathrm{corr}\left(r_{\Delta t}^{1},r_{\Delta t}^{2}\right)$ in terms of the correlation of underlying processes $\mathrm{corr}_{t_j}(\tilde{g}^{1},\tilde{g}^{2})$.

$$\mathrm{corr}\left(r_{\Delta t}^{1},r_{\Delta t}^{2}\right)=\mathrm{corr}\left(r_{\Delta t}^{1},r_{\Delta t}^{2}\right)=\frac{1}{T}\sum_{j=1}^{T}g_{\Delta t}^{1}(t)g_{\Delta t}^{2}(t)$$

$$=\frac{1}{T}\sum_{j=1}^{T}\left(\left(\frac{1}{\sqrt{\langle N_{\Delta t}^{1}\rangle}}\sum_{j=1}^{N_{\Delta t}^{1}(t)}\tilde{g}^{1}(\gamma^{1}(\mathrm{t})+j\,\Delta\tilde{t})+\frac{\langle\tilde{r}^{1}\rangle\left(N_{\Delta t}^{1}(t)-\langle N_{\Delta t}^{1}\rangle\right)}{\sigma_{\Delta t}^{1}}\right)\right.$$

$$\left.\times\left(\frac{1}{\sqrt{\langle N_{\Delta t}^{2}\rangle}}\sum_{j=1}^{N_{\Delta t}^{2}(t)}\tilde{g}^{2}(\gamma^{2}(\mathrm{t})+j\,\Delta\tilde{t})+\frac{\langle\tilde{r}^{2}\rangle\left(N_{\Delta t}^{2}(t)-\langle N_{\Delta t}^{2}\rangle\right)}{\sigma_{\Delta t}^{2}}\right)\right)$$

The average of the second component in each factor equals zero, thus under the assumption A1 we have.

$$\mathrm{corr}\left(r_{\Delta t}^{1},r_{\Delta t}^{2}\right)=\frac{1}{T}\sum_{j=1}^{T}\left(\frac{1}{\sqrt{\langle N_{\Delta t}^{1}\rangle}}\sum_{j=1}^{N_{\Delta t}^{1}(t)}\tilde{g}^{1}(\gamma^{1}(\mathrm{t})+j\,\Delta\tilde{t})\times\frac{1}{\sqrt{\langle N_{\Delta t}^{2}\rangle}}\sum_{j=1}^{N_{\Delta t}^{2}(t)}\tilde{g}^{2}(\gamma^{2}(\mathrm{t})+j\,\Delta\tilde{t})\right)$$

The means of the process $\tilde{g}^{1}$ and $\tilde{g}^{2}$ equal zero, thus under the assumption A2 we have.

$$\mathrm{corr}\left(r_{\Delta t}^{1},r_{\Delta t}^{2}\right)=\frac{1}{T}\sum_{j=1}^{T}\left(\frac{1}{\sqrt{\langle N_{\Delta t}^{1}\rangle\langle N_{\Delta t}^{2}\rangle}}\sum_{j=1}^{\bar{N}(t_j)}\tilde{g}^{1}(\gamma^{1}(\mathrm{t})+j\,\Delta\tilde{t})\tilde{g}^{2}(\gamma^{2}(\mathrm{t})+j\,\Delta\tilde{t})\right)$$

$$=\frac{1}{T}\sum_{j=1}^{T}\left(\frac{\bar{N}\left(t_j\right)}{\sqrt{\langle N_{\Delta t}^{1}\rangle\langle N_{\Delta t}^{2}\rangle}}\mathrm{corr}_{t_j}(\tilde{g}^{1},\tilde{g}^{2})\right)$$

Thus,

$$\mathrm{corr}_{c}\left(r_{\Delta t}^{1},r_{\Delta t}^{2}\right)=\frac{1}{T}\sum_{j=1}^{T}g_{\Delta t}^{1}(t)g_{\Delta t}^{2}(t)\frac{\sqrt{\langle N_{\Delta t}^{1}\rangle\langle N_{\Delta t}^{2}\rangle}}{\bar{N}\left(t_j\right)}.$$

□

References

1. T. Andersen, T. Bollerslev, F. Diebold and P. Labys, The distribution of realized exchange rate volatility, *J. American Statistical Association* **96** (2001a) 42-55.
2. T. Andersen, T. Bollerslev, F. Diebold and H. Ebens, The distribution of realized stock volatility, *J. Financial Econ.* **61** (2001b) 43-76.
3. J. J. Angel, Tick size, share prices, and stock splits, *Journal of Finance* 52 (1997) 655-681.

4. O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde and N. Shephard, Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading, Economics Papers 2008-W10, Economics Group, Nuffield College, University of Oxford (2008).
5. O. E. Barndorff-Nielsen and N. Shephard, Econometric analysis of realised covariation: High frequency covariance, regression and correlation in financial economics, Economics Papers 2002-W13, Economics Group, Nuffield College, University of Oxford (2001).
6. M. Bartolozzi, C. Mellen, T. Di Matteo, and T. Aste, Multi-scale correlations in different futures markets, Eur. Phys. J. B **58**, 207–220 (2007)
7. T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31 (1986) 307-327.
8. G. Bonanno, F. Lillo and R. N. Mantegna, High-frequency cross-correlation in a set of stocks, *Quantitative Finance* ***1*** (2001) 96-104.
9. *J.P. Bouchaud, M. Potters, Theory of financial risk and derivative pricing: from statistical physics to risk management (Cambridge University Press, Cambridge, 2003)*
10. D. Bourghelle and F. Declerck, Why markets should not necessarily reduce the tick size, *Journal of Banking & Finance* 28 (2004) 373-398.
11. C. Borghesi, M. Marsili, S. Micciche, Phys. Rev. E 76, 026104 (2007)
12. P. Burns, R. Engle and J. Mezrich, Correlations and volatilities of asynchronous data, *J. Derivatives* **5** (1998) 7-18.
13. Conlon, T., Ruskin, H. J. and Crane, M., *Physica A* 388(5) (2009) 705-714.
14. R. Cont and P. Tankov, *Financial Modelling with Jump Processes*, Chap. 7 (CRC Press, Taylor and Francis, 2010), second edn.
15. E.J. Elton, M.J. Gruber, *Modern Portfolio theory and investment analysis* (J. Wiley & Sons, New York, 1995)
16. T. W. Epps, Comovements in stock prices in the very short run, *Journal of the American Statistical Association* 74 (1979) 291-298.
17. J. E. Griffin and R. C. Oomen, Covariance measurement in the presence of non- synchronous trading and market microstructure noise, *Journal of Econometrics* 160 (2011) 58-68.
18. P. R. Hansen and A. Lunde, Realized variance and market microstructure noise, *Journal of Business & Economic Statistics* 24 (2006) 127-161.
19. L. Harris, Stock price clustering and discreteness, *Review of Financial Studies* 4 (1991) 389-415.
20. T. Hayashi and N. Yoshida, On covariance estimation of non-synchronously observed diffusion processes, *Bernoulli* 11 (2005) 359-379.
21. R. D. Huang and H. R. Stoll, Tick size, bid-ask spreads, and market structure, *Journal of*

Financial and Quantitative Analysis 36 (2001) 503-522.

22. S. Kozicki and B. Hoffman, Rounding error: A distorting influence on index data, *Journal of Money, Credit and Banking* 36 (2004) 319-38.
23. J. Kwapien, S. Drozdz and J. Speth, Time scales involved in emergent market coherence, *Physica A* ***337*** (2004) 231-242.
24. J.. Large, Technical Report, Oxford-Man Institute, University of Oxford, 2007
25. A. Lo and C. MacKinlay, An econometric analysis of nonsynchronous trading, *J. Econometrics* **45** (1990) 181-211.
26. M. Lundin, M. Dacorogna and U. A. Muller, *Financial Markets Tick by Tick*, chapter Correlation of High Frequency Financial Time Series (Wiley, 1998).
27. M. Martens and S.-H. Poon, Returns synchronization and daily correlation dynamics between international stock markets, *Journal of Banking & Finance* 25 (2001) 1805—1827.
28. Mastromatteo, I., Marsili, M., and Zoi, P., Financial correlations at ultra-high frequency: theoretical models,and empirical estimation, Eur. Phys. J. B **80**, 243–253 (2011)
29. M. C. Munnix, R. Schafer and T. Guhr, Compensating asynchrony effects in the calculation of financial correlations, *Physica A* 389 (2010) 767-779.
30. M. C. Munnix, R. Schafer and T. Guhr, Impact of the tick-size on financial returns and correlations, *Physica A* ***389*** (2010) 4828-4843.
31. J. Muthuswamy, S. Sarkar, A. Low and E. Terry, Time variation in the correlation structure of exchange rates: High-frequency analyses, *Journal of Futures Markets* ***21*** (2001) 127-144.
32. J.-P. Onnela, J. Toyli and K. Kaski, Tick size and stock returns, *Physica A* ***388*** (2009) 441-454.
33. S. Pafka, I. Kondor, G. Nagy, J. Bank. Financ. 31, 1545 (2007)
34. R. Reno, A closer look at the Epps-effect, *International Journal of Theoretical and, Applied Finance* 6 (2003) 87-102.
35. M. Scholes and J. Williams, Estimating betas from non-synchronous data, *J. Financial Econ.* 5 (1977) 309-327.
36. G. G. Szpiro, Tick size, the compass rose and market nanostructure, *Journal of Banking & Finance* 22 (1998) 1559-1569.
37. Toth and J. Kertesz, The epps effect revisited, *Quantitative Finance* 9 (2009) 793-802.
38. V. Voev and A. Lunde, Integrated covariance estimation using high-frequency data in the presence of noise, *Journal of Financial Econometrics* 5 (2007) 68-104.
39. A. Zebedee, J. Econ. Bus. **61**, 279 (2009)
40. A. A. Zebedee and M. Kasch-Haroutounian, A closer look at co-movements among stock returns, *Journal of Economics and Business* 61 (2009) 279-294.
41. L. Zhang, Estimating covariation: Epps effect, microstructure noise, *Journal of*

Econometrics 160 (2011) 33-47.