Université Paris Dauphine - PSL
Master Ingénierie Economique et Financière

The Epps effect under alternative sampling schemes

Natanaël BINYEGUI KOUAMO
Julian Sierra Pena
Amine MOUNAZIL
Emmanuel ZHENG

Gestion Quantitative

May 2022

Abstract

Cross-correlations in high resolution data are substantially smaller than their asymptotic value in daily data. Litterature indicates that changing the trade frequency should change the phenomenon's characteristic time. This is not the case with empirical data: The Epps curves do not scale in response to market activity. The latter conclusion implies that the time scale of the phenomena is related to market participants' reaction time (which we refer to as human time scale), regardless of market activity.

Outline

Introduction

Epps (1979) reported empirical evidence that stock correlations decrease when sampling frequency increases. This phenomenon, named Epps effect, has been observed in several markets. Central to the effect is the relationship between correlations and the time scale of the sampling scheme.

The Epps effect is a long-studied phenomenon that still lacks a satisfactory and comprehensive explanation. The main sources that contribute towards this effect include:

- lead-lag effect between stock returns which can appear mainly between stocks of very different capitalisation and/or for some functional dependencies between them
- asynchronicity of ticks in case of different stocks

Time metrics

Calendar time

This is the most natural way to perceive time as it is the way humans experience time at low-frequency and it is common to all investors and markets. It is also the most common approach used to compute statistical properties of financial time series because it can be defined without reference to the system generating the data.

Even though the passage of time is most natural to human operators under this definition, it is not without challenges. These problems include:

- intraday seasonality
- asynchronicity across stocks
- different and often overlapping market calendars and time zones

However, calendar time can allow coordinated sampling in a unique way independent of the data generation process.

Time metrics

Event time

In this approach, we can count events and use the event arrivals to increment a time.
For example, if we consider the full order book then one may increment event time by one unit when an order book event occurs. This may be a trade, a cancellation, an order modification or any order book event that changes the order book in some way.
Different types of event times :

- **transaction price** : time is increased by one unit each time a unique transaction occurs
- **tick time** : time is increased by one unit each time the price changes

Time metrics

Volume time

Here, time is increased by one unit for each unit of volume transacted each time a single share or instrument is exchanged.The advantage of this count is that stocks can naturally be synchronised, irrespective of their liquidity, directly in terms of sequence of comparable volumes traded.

This is a practical data-science construct which is useful for the aggregation of real-time relative risk measures such as the Volume Synchronized Probability of Informed Trading (VPIN).

Estimators

Malliavin-Mancino

Malliavin and Mancino introduced a Fourier estimator that expresses the Fourier coefficients of the volatility process using the Fourier coefficients of the price process $X_t^i = ln(p_t^i)$ where p_t^i is the generic asset price at time t for asset i.
Following Malliavin and Mancino, by rescaling the transactions times between [0, T] to [0, 2π] and using the Bohr convolution product, we have that for all k $\in$ Z and N samples:

$$\mathcal{F}(\sum^{ij})(k) = \lim_{N\to\infty} \frac{2\pi}{2N+1} \sum_{|s|\leq N} \mathcal{F}(dX^i)(s)\mathcal{F}(dX^j)(k-s) \tag{1}$$

with $N = \frac{1}{2}(\frac{T}{\Delta t} - 1)$

Estimators

Malliavin-Mancino

Using previous tick interpolation and a simple function approximation for the Fourier coefficients, we obtain the Dirichlet representation of the integrated volatility/covolatility estimator:

$$\widehat{\sum_{n,N}^{ij}} = \frac{1}{2N+1} \sum_{|s|\leq N, h=0, l=0}^{n_i-1, n_j-1} e^{is(t_l^j - t_h^i)} \delta_i(I_h)\delta_j(I_l) \tag{2}$$

The advantage behind the estimator is its ability to investigate different time scales through the choice of N. This means that there is no need to re-sample the raw observations onto a homogeneous grid using previous tick interpolation, such as in the case when using the RV estimator.

Estimators

Hayashi-Yoshida

Hayashi and Yoshida introduced a cumulative covariance estimator defined as:

$$\sum_{T}^{\hat{ij}} = \sum_{h=1}^{\#U^i} \sum_{l=1}^{\#U^j} (X^i_{t^i_h} - X^i_{t^i_{h-1}})(X^j_{t^j_l} - X^j_{t^j_{l-1}}) w_{hl} \tag{3}$$

The advantage behind the estimator is its ability to ameliorate the statistical cause of the Epps effect arising from asynchrony. The weakness of this estimator is that it is unable to investigate different time scales, which is particularly problematic because asynchrony is only one of the various sources of the Epps effect.

Estimators

Realised Volatility

Both the Malliavin–Mancino and Hayashi–Yoshida estimators were designed to compute estimates in calendar time, while dealing with the issue of asynchronous arrival of transactions. Both estimators overcome the need to re-sample the process onto a synchronous and homogeneous grid, for example, by using previous tick interpolation. However, given that we plan to re-sample the process under different definitions of time, it becomes useful to consider how the two estimators relate to the Realised Volatility (RV) estimator.

In the case of the Hayashi–Yoshida estimator, when t_h^i and t_l^j are synchronous and homogeneously spaced, we see that the RV estimator is given as:

$$\widehat{\sum_{T}^{ij}} = \sum_{h=1}^{n} (X^i_{t^i_h} - X^i_{t^i_{h-1}})(X^j_{t^j_h} - X^j_{t^j_{h-1}}) \tag{4}$$

Hawkes process

To generate a price process based on events, a Hawkes toy model was used. Concretely, parameters from the fine-to-coarse model were borrowed from Bacry et al. We consider the bivariate log-price:

$$\left\{ \begin{array}{lcl} X_t^1 & = & X_0^1 + N_1(t) - N_2(t) \\ X_t^2 & = & X_0^2 + N_3(t) - N_4(t) \end{array} \right. \tag{5}$$

where $\{N_m(t)\}_{m=1}^{M}$ is a 4-dimensioal (M = 4) mutually exciting Hawkes process with the associated intensity $\lambda(t) = left\{\lambda_m(t)_{m=1}^{M}$ taking the form :

$$\lambda^m(t) = \mu + \sum_{n=1}^{M} \int_{-\infty}^{t} \phi^{mn}(t-s)dN_s^n \tag{6}$$

Hawkes process

The counting processes are coupled through:

$$\Phi = \begin{pmatrix} 0 & \phi^r & \phi^c & 0 \\ \phi^r & 0 & 0 & \phi^c \\ \phi^c & 0 & 0 & \phi^r \\ 0 & \phi^c & \phi^r & 0 \end{pmatrix} \tag{7}$$

where : $\phi^{(r)} = \alpha^{(r)} e^{-\beta t} \nVdash_{t \in \mathbb{R}_+}$ and $\phi^{(c)} = e^{-\beta t} \nVdash_{t \in \mathbb{R}_+}$

Empirical : Extension of the Malliavin-Mancino estimator

The Fourier coefficients of the volatility process is given as:.

$$\alpha_t(\sum_{n_i,n_j,N}^{ij}) = \frac{2\pi}{2N+1}\sum_{|s|\leq N}\mathcal{F}(dX^i)(s)\mathcal{F}(dX^j)(k-s) \tag{8}$$

Using this formula, the instantaneous volatility/co-volatility can be reconstructed as follows :

$$\hat{\sum}^{ij}_{n_i,n_j,N,M}(t) = \sum_{|k|\leq M}(1-\frac{|k|}{M})e^{itk}\alpha_k(\sum_{n_i,n_j,N}^{ij}) \tag{9}$$

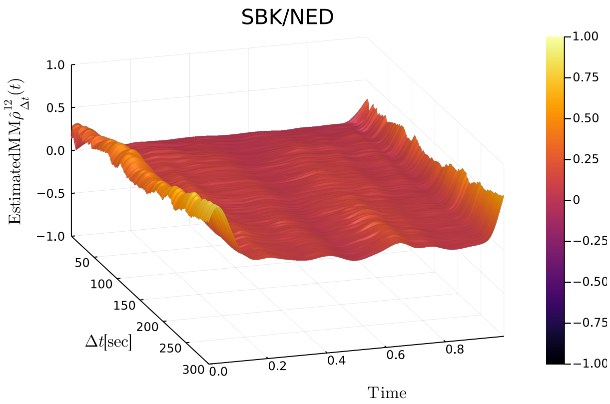

Figure: The Epps surface under calendar time for SBK/NED

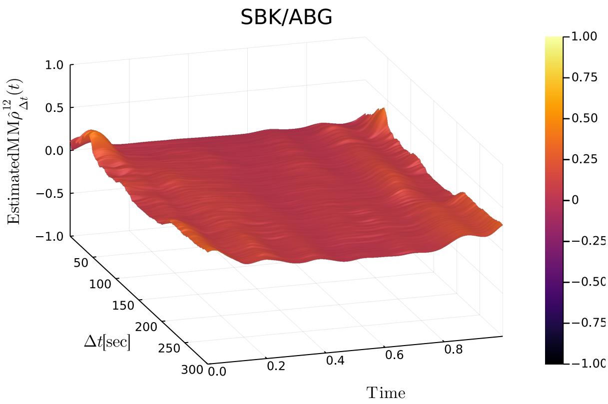

Figure: The Epps surface under calendar time for SBK/ABG

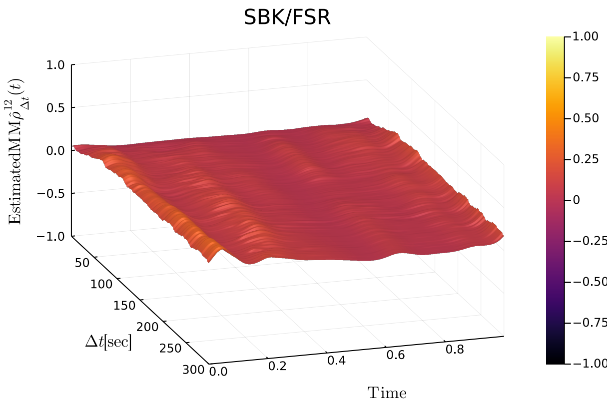

Figure: The Epps surface under calendar time for SBK/FSR

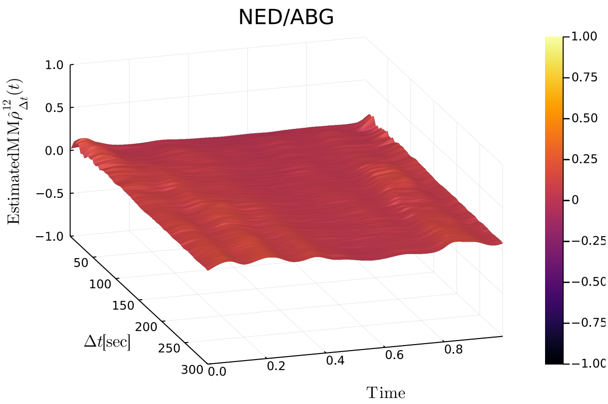

Figure: The Epps surface under calendar time for NED/ABG

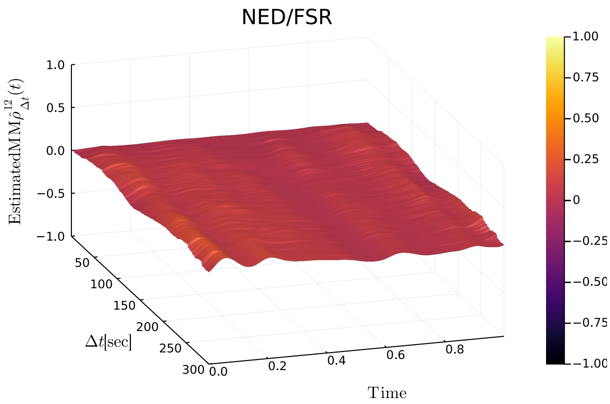

Figure: The Epps surface under calendar time for SBK/FSR

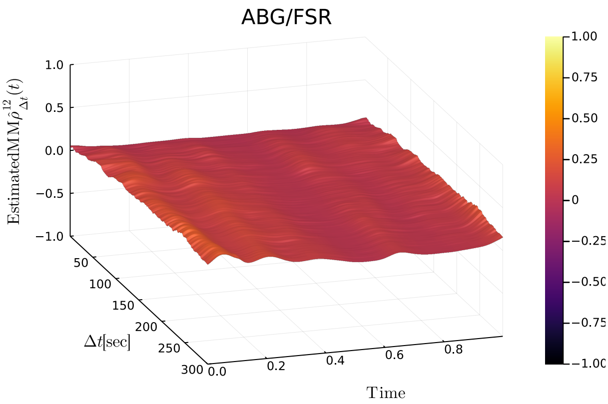

Figure: The Epps surface under calendar time for NED/ABG

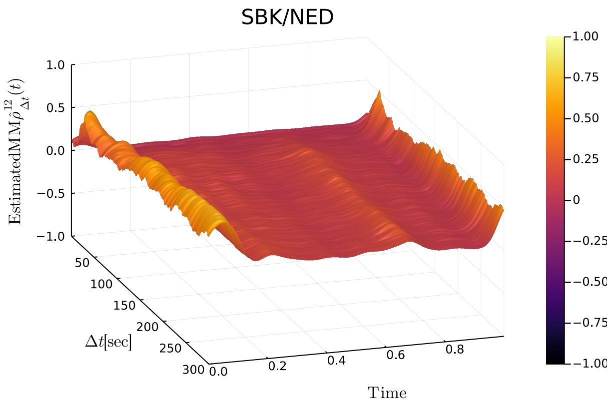

Figure: The Epps surface under event time for NED/FSR

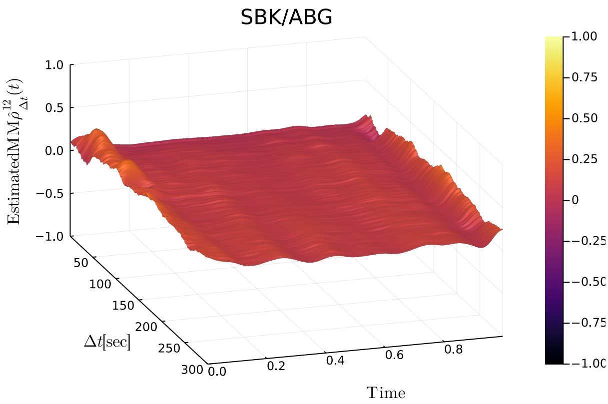

Figure: The Epps surface under event time for ABG/FSR

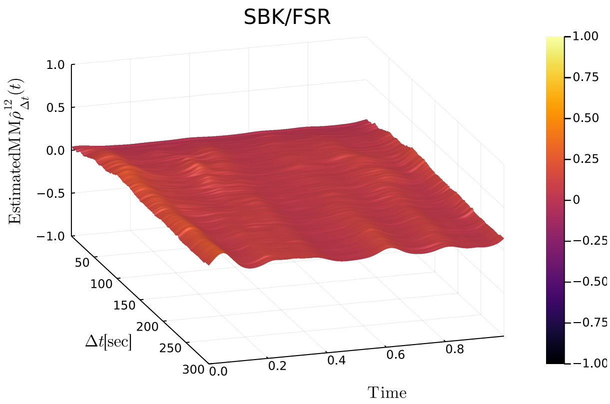

Figure: The Epps surface under event time for SBK/NED

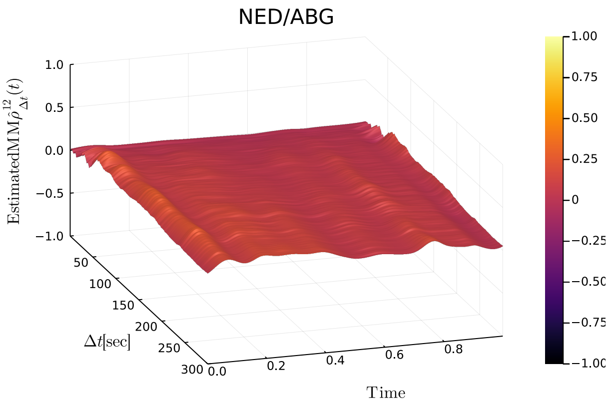

Figure: The Epps surface under event time for SBK/ABG

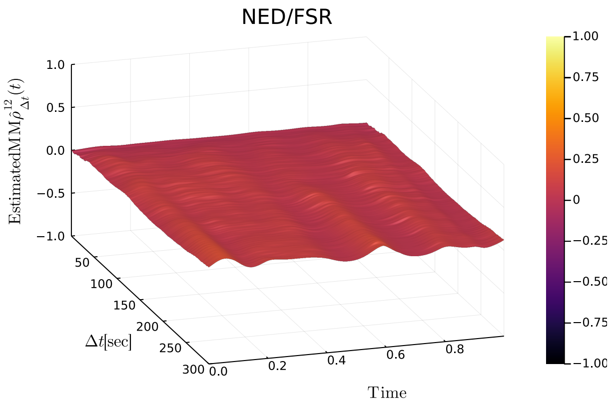

Figure: The Epps surface under event time for SBK/FSR

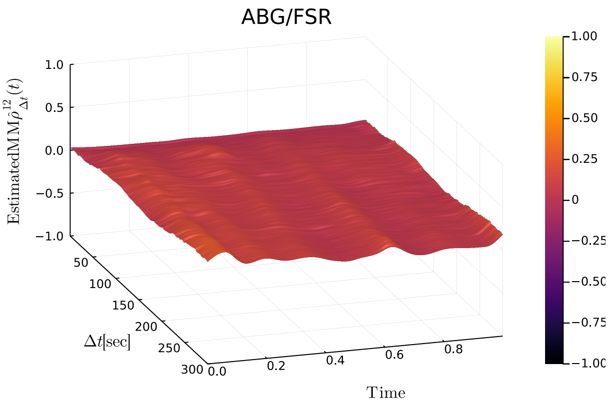

Figure: The Epps surface under event time for NED/ABG

Conclusion

In this study, we compared the Epps effect under three definitions of time: calendar time, trade time and volume. Furthermore, we investigated this phenomenon empirically on data from the original paper using the instantaneous volatility/co-volatility estimator which relies on non-uniform fast Fourier transforms to computationally speed up the evaluation.

References

- [T.W. Epps, Journal of the American Statistical Association 74, 291-298 (1979)]
- Bacry, E., Delattre, S., Hoffmann, M., Muzy, J.F., 2013a. Modelling microstructure noise with mutually exciting point processes. Quantitative Finance 13, 65–77.].
- Toth, B., Kertesz, J., 2009. The Epps effect revisited. Quantitative Finance 9, 793–802.
- Chang, P., 2020. Fourier instantaneous estimators and the Epps effect. PLOS ONE 15, 1–24.
- Ren'o, R., 2003. A closer look at the Epps effect. International Journal of Theoretical and Applied Finance 06, 87–102.