Imperial College London

MATH70125 — Market Microstructure

Imperial College London

Department of Mathematics

Assessed Coursework

Authors:
Max Dedieu (02289032),
Rayan Messekher (02299112),
Matthieu Baboulène (02299325),
Quentin Varrier (02280080)

Lecturer:
Mathieu Rosenbaum

May 15, 2023

Contents

1 Introduction

Market microstructure is a field that examines the mechanisms governing financial markets, particularly the execution and trading of securities. It provides insights into market functioning, price determination, liquidity, and overall market efficiency. This understanding is vital for traders, investors, regulators, and researchers seeking to navigate the complexities of financial markets and make informed decisions.
In this coursework, we will analyze key concepts in market microstructure by simulating correlated prices paths within the Black-Scholes model. We will study the dynamics between observed tick price and the underlying semi-martingale efficient price.
We will then study the different covariation and correlation estimators available for asynchronous and non-asynchronous time-series. Namely, we will explore the Epps effect through simulated Black-Scholes prices path, and see how it is a limiting factor in co-variation estimation.
Additionally, we will look at the Hayashi-Yoshida estimator, a statistical tool used in market microstructure to estimate the covariation of two asynchronous assets. We will also introduce the potential benefits of the corrected Hayashi-Yoshida estimator.

2 Black-Scholes paths for 2 correlated assets

To carry out the coursework, we first need to simulate correlated geometric Brownian motions to represent the efficient price of our assets. We recall the formula for two assets $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$:

$$\begin{aligned} dX_t &= X_t\mu_X dt + X_t\sigma_X dW_t \\ dY_t &= Y_t\mu_Y dt + Y_t\sigma_Y(\rho dW_t + \sqrt{1-\rho^2}dB_t) \\ [W_t, B_t]_t &= 0 \end{aligned} \tag{1}$$

where $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ are market parameters (most likely $\mu_X = \mu_Y = r$ where r is the risk-free rate). $\rho \in [-1, 1]$ is the correlation factor between X and Y.
We wish to simulate the two assets' variations on the interval $[0, 1]$. We select a number of steps $N \in \mathbb{N}^*$, and create the time points $(t_i = \frac{i}{N})_{i\in[0,N]}$. We can then use Ito's lemma and integrate 1 to get the following expression :

$$\begin{aligned} X_{t_i} &= X_0 \exp\left(\left(\mu_X - \frac{1}{2}\sigma_X^2\right)t_i + \sigma_X W_{t_i}\right) \\ Y_{t_i} &= Y_0 \exp\left(\left(\mu_Y - \frac{1}{2}\sigma_Y^2\right)t_i + \sigma_Y\left(\rho W_{t_i} + \sqrt{1-\rho^2}B_{t_i}\right)\right) \end{aligned} \tag{2}$$

We observe one pair of sample path below on figure 1.

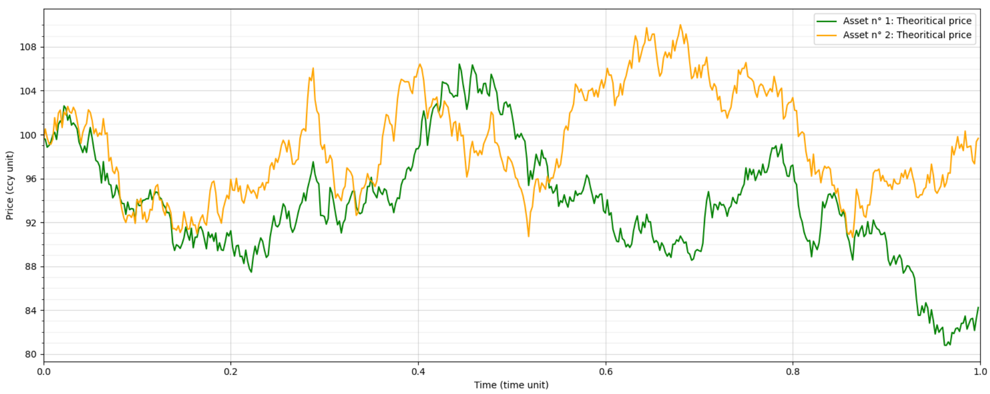

Figure 1: Sample Black-Scholes-Merton path for two assets with $r = 5\%, \sigma_X = \sigma_Y = 20\%, N = 500, \rho = 0.5, X_0 = Y_0 = 100$.

In the remaining tasks, we will always select the same volatility for X and Y for scaling purposes. The grid of the plot on figure 1 is divided in ticks, where we have $\alpha = 1$.

3 The Uncertainty zones model

We know the efficient prices are the underlying semi-martingale component of the observed price (P_{t_i}). The observed price lives on the tick grid, where each possible price is a multiple of the tick value α. To get the efficient price from the observed price we use below formula :

$$X_{\tau_i} = P_{t_i} + \alpha(\frac{1}{2} - \eta)\text{sign}(P_{t_i} - P_{t_{i-1}}) \tag{3}$$

Where η is the aversion-to-change parameter and $(t_i)_{i\geq 0}$ are the observed transaction time and $(\tau_i)_{i\geq 0}$ are random stopping times such that $\tau_i \geq t_i$. In our simulation we take $\tau_i = t_i, \forall i \in \mathbb{N}^*$ for simplification purposes.
In our problem's setting, we have the efficient prices, and we try to reverse 3 to get the observed prices $(P_{t_i})_{i\geq 0}$. To do so, we divide the price space in multiples of α, and create uncertainty zones of width $2\alpha\eta$ centered on the half-ticks, such that $U_k = [0, +\infty[\times[d_k, u_k]$ and $d_k = \alpha(k + \frac{1}{2} - \eta)$,$u_k = \alpha(k + \frac{1}{2} + \eta)$. We use below rules to update the observed price from the behavior of the efficient price :

- When the efficient price X_t enters an uncertainty zones U_n from below at time t_0, such that $X_{t_0} = d_n$ and exits it at time t_1 such that such that $X_{t_1} = u_n$, we set the observed price $P_{t_1} = X_{t_1} = u_n$
- When the efficient price X_t enters an uncertainty zones U_n from above at time t_0, such that $X_{t_0} = u_n$ and exits it at time t_1 such that such that $X_{t_1} = d_n$, we set the observed price $P_{t_1} = X_{t_1} = d_n$

Using the rules above, we can compute and graph the observed prices from the efficient prices, as done below in figure 2

Figure 2: Efficient (plain) and observed (dot) prices for two simulated assets with parameters $r = 0\%, \sigma_X = \sigma_Y = 1\%, N = 1000, \rho = -0.5, X_0 = Y_0 = 100$. The green lines represent the boundary of the uncertainty zones U_n.

4 Covaration and correlation estimator

We have previously simulated prices for two correlated assets under the Black-Scholes-Merton assumptions. We can use these samples to study different methods of covariance estimation on empirical data.

4.1 Non-asynchronous setting

We first study the covariance and correlation estimator in the non-asynchronous setting, meaning that we observe the price jumps for asset X and Y at the same times $t_i, i \geq 0$. We make the assumption of constant volatility (which is true since we simulate the assets' price with constant σ_X, σ_Y). We use below estimator of covariance :

$$\begin{aligned} \Delta_i^n X &= log(X_{\frac{i}{n}}) - log(X_{\frac{i-1}{n}}) \\ \Delta_i^n Y &= log(Y_{\frac{i}{n}}) - log(Y_{\frac{i-1}{n}}) \\ \hat{c_n} &= \sum_{i=1}^{n} \Delta_i^n X \Delta_i^n Y \end{aligned} \tag{4}$$

$\hat{c_n}$ is an estimator of $\int_0^1 \sigma_{X\,t}\sigma_{Y\,t}\rho_t dt$. Under the assumption of constant correlation and volatility, we have $\hat{\rho_n}$ the estimator of correlation ρ.

$$\hat{\rho_n} = \frac{\hat{c_n}}{\sqrt{\sum_{i=1}^{n}(\Delta_i^n X)^2 \sum_{i=1}^{n}(\Delta_i^n Y)^2}} \tag{5}$$

We can use Monte Carlo simulation to simulation multiple pairs of paths for the two assets, and use the Central Limit Theorem to derive a confidence interval for the estimator (with a convergence speed $\sqrt{m}$, where m is the number of simulation). We observe the performance of this estimator in figure 3

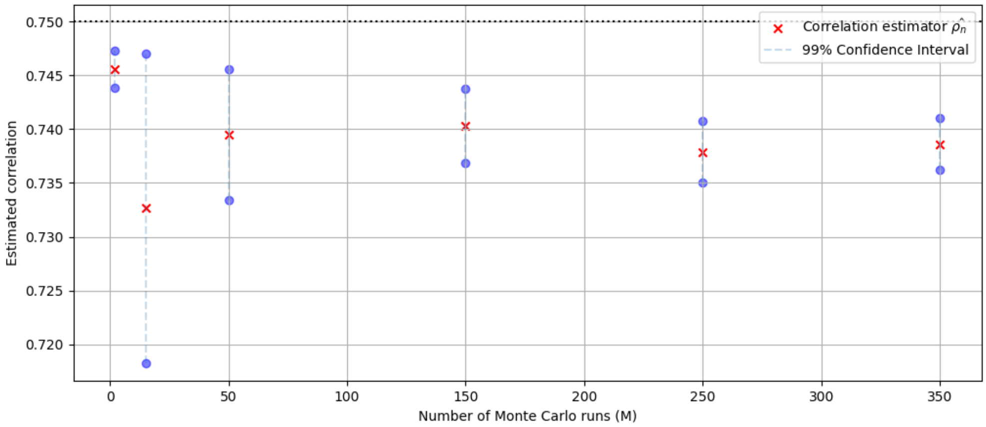

Figure 3: Correlation estimator for constant volatility assumptions for different number of Monte Carlo simulations. Two simulated assets with parameters $r = 30\%, \sigma_X = \sigma_Y = 80\% N = 1000, \rho = 0.8, X_0 = Y_0 = 100$.

In the case of non-constant volatility, we use the $\hat{c_n}$ to estimate the covariance, and the $\hat{a_n}$ estimator to estimate $\frac{\pi}{2}\int_0^1 \sigma_{X\,t}\sigma_{Y\,t} dt$, such that :

$$\hat{a_n} = \sum_{i=1}^{n-1} |\Delta_{i+1}^n X \Delta_i^n Y| \tag{6}$$

In that case we have :

$$\hat{\rho_n} = \frac{2}{\pi}\frac{\hat{c_n}}{\hat{a_n}} \tag{7}$$

We can simulate non-constant volatility in the Black-Scholes model by changing the volatility σ during the simulation in equation 1. We observe the result in figure 4.

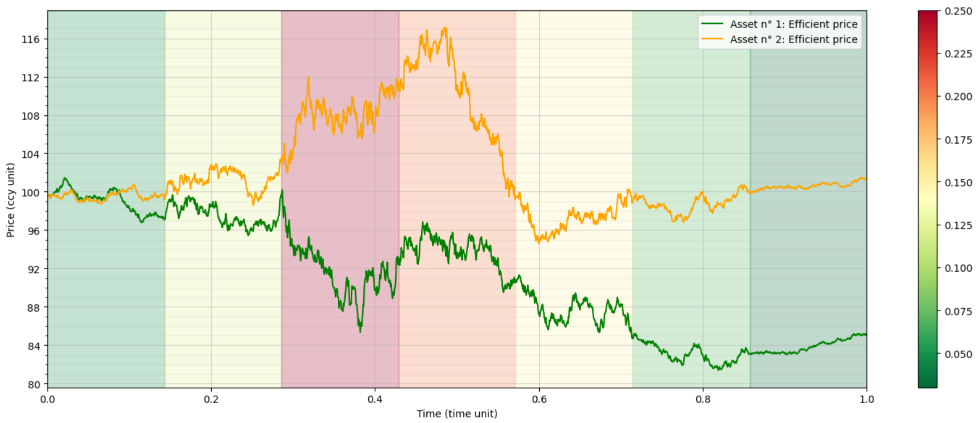

Figure 4: Black-Scholes path for two assets under non-constant volatility with parameters $r = 5\% N = 2000, \rho = 0.5, X_0 = Y_0 = 100$.

Finally, we can observe the corrected estimator in action below in figure 5. We see it successfully estimate the correct correlation between the two assets, while the previous estimator is biased.

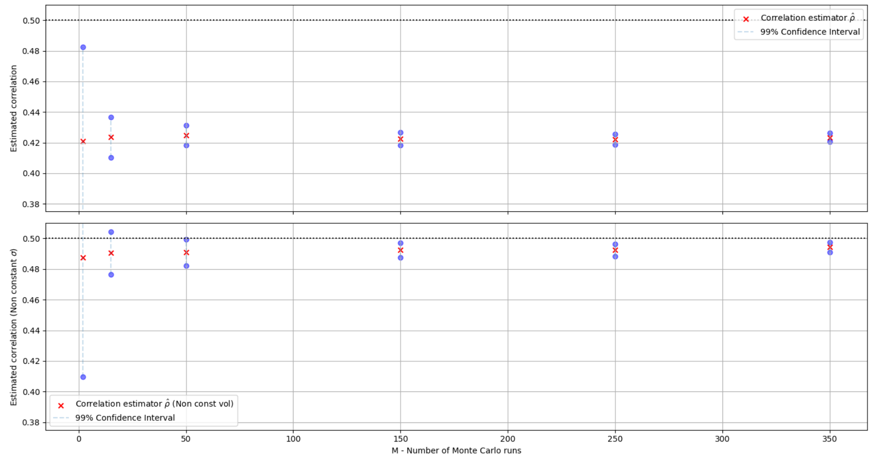

Figure 5: Correlation estimator for constant volatility assumptions (top) and non constant volatility assumption (bottom) for different numbers of Monte Carlo simulations. Two simulated assets with parameters $r = 5\%, N = 5000, \rho = 0.5, X_0 = Y_0 = 100$ and varying volatility

These estimators are biased and do not yield satisfying results in this setting. This is because the hypothesis of synchronicity of the series is not verified.

4.2 Asynchronous setting - The Epps effect

In our simulation, and more generally in the real world, we observe price jumps at different times for two different assets. In our setting, we have asset X and Y, and observe price jumps at time $(T_i^X)_{i\geq 0}$ for asset X and at time $(T_i^Y)_{i\geq 0}$ for asset Y. In order to compute covariance, we introduce the "extended"-prices $\bar{X}_t$ and $\bar{Y}_t$ such that :

$$\begin{aligned} \bar{X}_t &= X_{T_i^X}, \text{for } t \in [T_i^X, T_{i+1}^X[\\ \bar{Y}_t &= X_{T_i^Y}, \text{for } t \in [T_i^Y, T_{i+1}^Y[\end{aligned} \tag{8}$$

This notation allows us to deal with piecewise constant functions, and is equivalent to the last-traded price for each asset X and Y. We can now work with arbitrary precision since $(\bar{X}_t)_{t\geq 0}$ and $(\bar{Y}_t)_{t\geq 0}$ are piecewise continuous.
The previous tick covariation is then defined as :

$$V_h = \sum_i^m (\log \bar{X}_{ih} - \log \bar{X}_{(i-1)h})(\log \bar{Y}_{ih} - \log \bar{Y}_{(i-1)h}), \quad mh = 1 \tag{9}$$

This estimator allows us to compute the covariance of X and Y in an asynchronous setting. However, it has several shortcomings. This estimator is systematically biased, meaning we cannot obtain the true value of $\int_0^1 \sigma_{X t} \sigma_{Y t} dt$ using it. Moreover, we observe below property :

$$\lim_{h\to\infty} \mathbb{E}[V_h] = 0 \tag{10}$$

This phenomena is called the "Epps effect". This can be explained since the finer the mesh ($h \to 0$), the less chance we have to observe two jumps in the interval $[i, (i+1)h]$. This means that even if X has a jump in the interval $[i, (i+1)h]$, $\log \frac{\bar{X}_{ih}}{\bar{X}_{(i-1)h}} \neq 0$ but $\log \frac{\bar{Y}_{ih}}{\bar{Y}_{(i-1)h}} = 0$, and conversely for a jump of Y on $[i, (i+1)h]$.
In practice, we use the Central-Limit-Theorem to estimate confidence interval at a 99% level. We obtain below graph for 500 simulation each, on figure 6.

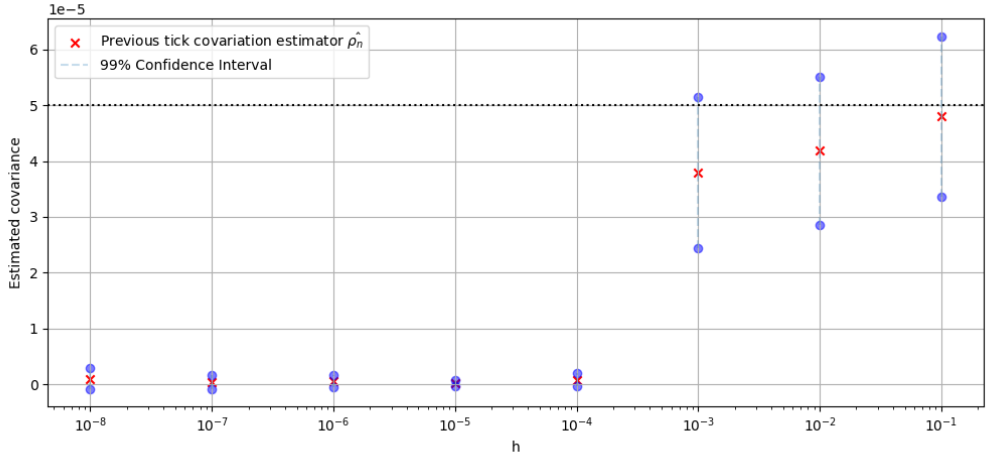

Figure 6: The Epps effect for covariance estimation of two assets' observed price with M=500 runs at each step-size h. Market parameters are $r = 0\%, \sigma = 1\%, N = 1000, \rho = 0.5, X_0 = Y_0 = 100$.

5 The Hayashi-Yoshida estimator

5.1 Non corrected estimator

The Hayashi-Yoshida estimator is an attempt at negating the Epps effect by being independent of a precision parameter h. Let $I_i^X = [T_i^X, T_{i+1}^X[$ and $I_i^Y = [T_i^Y, T_{i+1}^Y[$ the interval between jumps for X and Y. With $(P_{t_i}^X)_{i\geq 0}$ and $(P_{t_i}^Y)_{i\geq 0}$ the observed prices of X and Y at jump time $t = t_i$, we define the Hayashi-Yoshida estimator H_n :

$$H_n = \sum_{i,j} \Delta^{P^X}(I_i^X)\Delta^{P^Y}(I_j^Y)\mathbb{1}_{\{I_i^X \cap I_j^Y \neq \emptyset\}} \tag{11}$$

This estimator is independent of a precision parameter h, we sum the product of the log returns on each intersecting pair of jumps. Below is a graphical representation of the Hayashi-Yoshida sum, on figure 7.

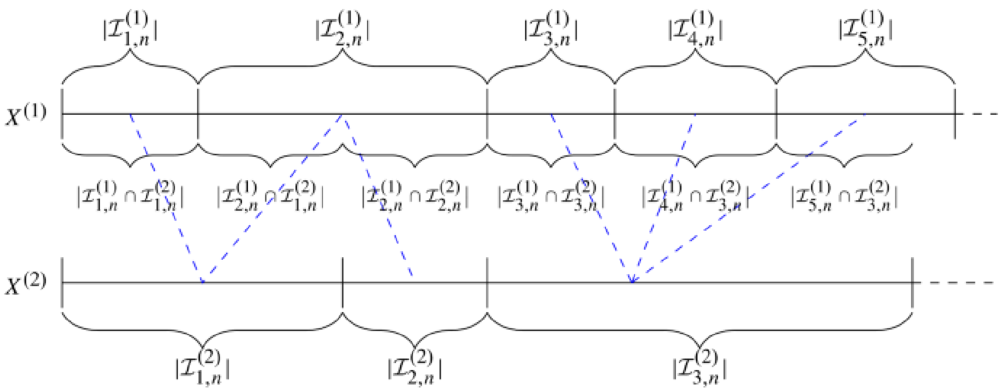

Figure 7: Representation of the Hayashi-Yoshida sum, directly taken from Laws of large numbers for Hayashi–Yoshida-type functionals - 10 May 2019 - Ole Martin & Mathias Vetter

We can compute the Hayashi-Yoshida estimator using the central limit theorem to gain information about the confidence interval at level 99%. We compile the results of the simulations in below table : 1.

Number of runs	Mean value $\bar{H}_n$	99% Confidence interval	% Error
5	0.09510	0.00075	3.51%
15	0.09286	0.00441	1.08%
50	0.09267	0.00402	0.86%
150	0.09270	0.00372	0.90%
500	0.09257	0.00402	0.76%

Table 1: Table of the results of the non-corrected Hayashi-Yoshida estimator for the observed price, for market parameters : $r = 5\%, N = 1500, \sigma = 30\%, \rho = 0.75, S_0 = 100., \alpha = 0.01, \eta = 0.2$ and covariance $= 0.0919$

As we can see, the estimator provides very satisfying results, the precision is increased when we increase the number of Monte Carlo runs.

5.2 Corrected estimator

As seen on 1, the Hayashi-Yoshida still has some level of bias to it when we use it on the observed price. We can however reduce this bias by using it on the efficient price instead. We create the corrected Hayashi-Yoshida estimator, where $(E_t^X)_{t\geq 0}$ and $(E_t^Y)_{t\geq 0}$ are the efficient price of X and Y.

$$H_n^{(c)} = \sum_{i,j} \Delta^{E^X}(I_i^X)\Delta^{E^Y}(I_j^Y)\mathbb{1}_{\{I_i^X \cap I_j^Y \neq \emptyset\}} \tag{12}$$

This estimator, ran in the same settings as 1, gives the results on table 2.

Number of runs	Mean value $H_n^{(c)}$	99% Confidence interval	% Error
5	0.08995	0.00286	-2.10%
15	0.09085	0.00311	-1.11%
50	0.09183	0.00341	-0.05%
150	0.09144	0.00428	-0.47%
500	0.09145	0.00389	-0.46%

Table 2: Table of the results of the non-corrected Hayashi-Yoshida estimator for the observed price, for market parameters : $r = 5\%, N = 1500, \sigma = 30\%, \rho = 0.75, S_0 = 100., \alpha = 0.01, \eta = 0.2$ and covariance $= 0.0919$

As we can see, this estimator yields more accurate results than the non-corrected version.

6 Conclusion

In conclusion, we have simulated the underlying efficient price of two correlated assets using the Black-Scholes model, and we have deduced the observed price using the η "aversion-to-price-change" parameter. We then studied the Epps effect, which is an issue with common covariance estimator for observed price. We saw how the effect was impacted by the step-size parameter h and how the theoretical result $[V_h] \to 0$ was verified in practice. We then introduced the more sophisticated Hayashi-Yoshida estimator, which enables us to compute the covariance without setting a step-size parameter. We also saw how we could reduced the bias by using the estimator on the underlying semi-martingale instead of on the directly observed price.
This coursework allowed us to have a first detailed look on some core concept of market microstructure, like aversion to price-change modelling, estimation of statistical component between series of observed price and more. If we were to extend the coursework, we could work on setting up a lead-lag trading strategy by estimating the covariance between two simulated "lagged" observed price series.