https://doi.org/10.26219/heal.aueb.5795

Modelling Realized and Implied Volatility with Lévy Processes in FX Markets

Ioannis Chouliaras

A THESIS

Submitted to the Department of Statistics

of the Athens University of Economics and Business

in partial fulfilment of the requirements for

the degree of Master of Science in Statistics

Athens, Greece
August 2023

https://doi.org/10.26219/heal.aueb.5795

Μοντελοποίηση κινδύνου και τεκμαρτού κινδύνου με τη χρήση στοχαστικών διαδικασιών Lévy στην αγορά συναλλαγμάτων

Ιωάννης Χουλιάρας

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Διπλώματος Μεταπτυχιακών Σπουδών στη Στατιστική

Αθήνα, Ελλάδα
Αύγουστος, 2023

https://doi.org/10.26219/heal.aueb.5795

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Professor Athanasios Yannacopoulos and Dr. Panagiotis Papaioannou, for their full guidance, patience, and understanding throughout this thesis. Finally, I am thankful to all my friends and family for their understanding and support during this period.

https://doi.org/10.26219/heal.aueb.5795

https://doi.org/10.26219/heal.aueb.5795

VITA

After completing my studies in Accounting and Finance (AUEB) I decided to continue in the Statistics field, and more specifically in Financial Mathematics domain. At the same time, I am currently working in financial industry and more specifically in Asset Management. I am looking forward to the next steps of my career either by continuing my studies or working in the financial mathematics domain.

https://doi.org/10.26219/heal.aueb.5795

https://doi.org/10.26219/heal.aueb.5795

ABSTRACT

Ioannis Chouliaras

Modelling Realized and Implied Volatility using Lévy Processes in FX Markets

August 2023

The purpose of the present thesis is the application of Lévy Processes in FX option pricing by comparing the Black-Scholes model with Merton Jump Diffusion model. In addition, the comparison of volatility measures such as realized and implied volatility will be implemented under the BS model in FX Option Pricing. Finally, financial time series predictive techniques will be used to forecast the respective inputs of the BS model, spot price, volatility and both interest rates.

Various stochastic processes, widely known in financial mathematics literature, will be presented theoretically and we will provide an extension of the Vasicek model by adding a jump component represented by a Compound Poisson process. Techniques from financial time series analysis and machine learning will also be presented in an introductive theoretical aspect.

The idea behind implementing Lévy Processes instead of standard geometric Brownian motion (GBM), which is widely used is finance literature, is that the first ones can capture different stylized facts that financial data possesses, such as fat tails, non-normality and asymmetry.

https://doi.org/10.26219/heal.aueb.5795

https://doi.org/10.26219/heal.aueb.5795

ΠΕΡΙΛΗΨΗ

Ιωάννης Χουλιάρας

Μοντελοποίηση κινδύνου και τεκμαρτού κινδύνου με τη χρήση στοχαστικών διαδικασιών Lévy στην αγορά συναλλαγμάτων

Αύγουστος 2023

Ο σκοπός της παρούσας διπλωματικής είναι η εφαρμογή των διαδικασιών Lévy στην τιμολόγηση δικαιωμάτων προαίρεσης στην αγορά συναλλαγμάτων συγκρίνοντας το Black-Scholes υπόδειγμα και το Merton Jump Diffusion υπόδειγμα. Επιπλέον, η σύγκριση των μέτρων μεταβλητότητας, όπως η ιστορική και η τεκμαρτή μεταβλητότητα, θα εφαρμοστεί σύμφωνα με το υπόδειγμα Black-Scholes στην τιμολόγηση δικαιωμάτων προαίρεσης. Τέλος, θα χρησιμοποιηθούν τεχνικές πρόβλεψης χρηματοοικονομικών χρονοσειρών για την πρόβλεψη των αντίστοιχων εισαγόμενων μονάδων του υποδείγματος Black-Scholes, όπως η ισχύουσα τιμή συναλλάγματος, του κινδύνου και των δύο επιτοκίων.

Διάφορες στοχαστικές διαδικασίες, ευρέως γνωστές στη βιβλιογραφία των χρηματοοικονομικών μαθηματικών, θα παρουσιαστούν σε θεωρητικό επίπεδο και θα παρέχουμε μια επέκταση του υποδείγματος Vasicek προσθέτοντας ένα άλμα που προκύπτει από μια σύνθετη διαδικασία Poisson. Τεχνικές από την ανάλυση χρηματοοικονομικών χρονοσειρών και τη μηχανική μάθηση θα παρουσιαστούν, επίσης, σε μια εισαγωγική θεωρητική πτυχή.

Η ιδέα πίσω από την εφαρμογή των διαδικασιών Lévy, αντί της τυπικής γεωμετρικής κίνησης Brown (GBM) που χρησιμοποιείται ευρέως στην χρηματοοικονομική βιβλιογραφία είναι ότι οι πρώτες μπορούν να συλλάβουν διαφορετικά ιδιαίτερα χαρακτηριστικά που διαθέτουν τα χρηματοοικονομικά δεδομένα, όπως οι παχιές ουρές, η μη κανονικότητα και η ασυμμετρία.

https://doi.org/10.26219/heal.aueb.5795

https://doi.org/10.26219/heal.aueb.5795

Table of Contents

https://doi.org/10.26219/heal.aueb.5795

https://doi.org/10.26219/heal.aueb.5795

List of Figures

https://doi.org/10.26219/heal.aueb.5795

https://doi.org/10.26219/heal.aueb.5795

https://doi.org/10.26219/heal.aueb.5795

List of Tables

https://doi.org/10.26219/heal.aueb.5795

Chapter 1

1.Introduction

A significant part of financial literature and, in particular, option pricing and its applications assume that the underlying prices follow diffusion type processes and more specifically a geometric Brownian motion (GBM). Black and Scholes (1973) in their revolutionary work presented the theoretical estimate of the price of European-style options, under several underlying assumptions. These underlying assumptions are also the weakness of this well-known model and, in real cases, lead to mispricing of the options. In that content, empirical studies have shown that there are some stochastic processes that may capture better the behavior of the underlying asset and its volatility, processes known as Lévy processes. In mathematical finance, Lévy processes are becoming widely known since they can describe the observed reality of the financial markets, given that empirical studies have shown that the asset price processes have jumps or spikes. In addition, while Black-Scholes assumed that the empirical distribution is the normal distribution, the empirical distribution of asset returns exhibit fat tails and skewness. Moreover, a vital concept in finance and mathematical finance is the volatility of the underlying assets and the optimal way to model its behavior. Traders need models that can capture implied volatility accurately, to handle the risk of their trades. Another significant concept of financial time-series is mean-reversion. In financial literature, stochastic processes, such as Ornstein Uhlenbeck, exhibit this characteristic and instruments, like interest rates, are usually modeled by these processes. The well-known Vasicek model (1977) that used to describe the evolution of the interest rate is, actually, an Ornstein Uhlenbeck process. Furthermore, financial time series data exhibit some similar characteristics, known as stylized facts. These specific characteristics lead to the use of specific models coming from different domains of mathematics, statistics and computer science. Forecasting financial time-series by implementing these techniques could be beneficial to predict the inputs of the Black-Scholes framework with an interest in calculating a

https://doi.org/10.26219/heal.aueb.5795

theoretical option price, under which a trader may execute a trade or hedge their position.

The main aim of this thesis is twofold. Firstly, the theoretical concepts of Lévy processes will be presented along with the simulation of various well-known Lévy processes (Chapter 2). Moving forward, the same procedure will be implemented in respect of Ornstein Uhlenbeck processes (Chapter 2) and the connection between these processes with the Vasicek model (Chapter 3). Techniques for forecasting financial time series data and their volatility will be introduced in Chapter 4. More specifically, Generalized Autoregressive Conditional Heteroscedasticity model (GACRH) and some of its variations for modelling volatility of asset returns, time series modelling will be implemented by using models such as autoregressive (AR) and autoregressive integrated moving average (ARIMA) models and finally a simple recurrent neural network (RNN)- long short-term memory (LSTM) will be presented. In Chapter 5, we will deal with the well-known framework of Black-Scholes and more specifically the adjusted Black-Scholes model in Foreign Exchange Markets, first presented by Garman & Kohlhagen (1983). In addition, the extension of Black-Scholes model introduced by Merton (1979) will be presented in Chapter 5, extending the Black-Scholes model to a jump diffusion model. Finally, an extension of the Vasicek model will be introduced, after adding a jump component in the standard model, assuming a compound Poisson process. This chapter will conclude the more theoretical aspect of the thesis. In Chapter 7, we will deal with FX markets and the application of some concepts presented in the theoretical part of the thesis. The main purpose of the empirical application will be to compare the option prices after assuming realized volatility in one case and implied volatility in the other. After calculating both call and put prices for both volatilities, we will combine the forecasting methods presented in Chapter 4 and the calculation of theoretical option prices for the next day will be made. GARCH models have been selected for volatility, time-series models for interest rate and RNN-LSTM for the spot exchange price. Consequently, the call and put price of Merton's jump diffusion model will be calculated to compare the option prices using the standard Black-Scholes and the jump-diffusion assumption-based model.

https://doi.org/10.26219/heal.aueb.5795

Chapter 2

2.Preliminaries

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a filtered probability space, where $\mathcal{F}$ is the filtration $(\mathcal{F}_t)$ with $t \in [0,T]$ satisfying the usual conditions and $T \in [0,\infty]$ denoting the time horizon, Ω is the set containing all the possible events, the elements of $\mathcal{F}$ are events than can be assigned a probability and $\mathbb{P}$ a probability measure, most usually mentioned as the physical measure.

2.1. Lévy processes

Definition 2.1.1. A stochastic process $\{L_t\}t \geq 0$ on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$ called a Lévy process if the following conditions are satisfied:

I. $L_0 = 0$ almost surely (a.s.)
II. L_t has independent increments meaning that, $L_t - L_s$ is independent of $\mathcal{F}_s$ for any $0 \leq s < t \leq T$.
III. L_t has stationary increments, i.e., the distribution of $L_{t+s} - L_t$ does not depend on t for any $0 \leq s, t \leq T$.
IV. L_t is stochastically continuous, that is, $\lim_{t \to 0}(|L_{s+t} - L_s| \geq \varepsilon) = 0, \forall \varepsilon \geq 0$.
V. The sample paths are right continuous with left limits a.s., a càdlàg process.

Some well-known examples of Lévy processes are Brownian motion, which is the only Lévy process with continuous sample paths, Poisson, and compound Poisson processes. A jump-diffusion process is often called the sum of a linear drift, a Brownian motion and a compound Poisson process. These examples will be further defined and illustrated in a later section.

Suppose $\varphi(u)$ is the characteristic function of a distribution and Px its law. We say that the distribution is infinitely divisible if for every positive integer n, φ(u) is also the n^{th} power of a characteristic function. For example, assume $L_t, t \geq 0$ a càdlàg Lévy process then this process has $(\varphi(u))^t$ as its

https://doi.org/10.26219/heal.aueb.5795

characteristic function. For a more formal definition of infinitely divisible property, see below definition 2.1.2.

Definition 2.1.2. The law P_x of a random variable X is infinitely divisible, if for all n ∈ ℕ there exist i.i.d. random variables $X_1^{(1/n)}$, …., $X_n^{(1/n)}$ such that

(2.1.1) $$X \stackrel{d}{=} X_1^{(1/n)} + \cdots + X_n^{(1/n)}.$$

Equivalently, the law P_x of a random variable X is infinitely divisible, if for all n ∈ ℕ there exists another law $P_x^{(1/n)}$ such that

(2.1.2) $$Px = Px(1/n) * \ldots * Px(1/n).$$

Alternatively, we can characterize an infinitely divisible random variable using its characteristic function.

Definition 2.1.3. The law Px of a random variable X is infinitely divisible, if for all $n \in \mathbb{N}$, there exists a r.v. $X_1^{(1/n)}$, such that:

(2.1.3) $$\varphi(u) = (\varphi(1/n)(u))\, n.$$

The next theorem, known as Lévy-Khintchine formula, provides a complete characterization of random variables with infinitely divisible distributions through their characteristic functions.

Theorem 2.1.1 (Lévy-Khintchine representation). The law Px of a random variable X is infinitely divisible, if and only if there exists a triple (b, c, v), with $b \in \mathbb{R}$, $c \in \mathbb{R} \geq 0$ and a measure satisfying $v(\{0\}) = 0$ and $\int_{\mathbb{R}} (1 \wedge |x|^2) v(dx) < \infty$, such that

(2.1.4) $$\mathbb{E}\left[e^{iux}\right] = \exp[ibu - \frac{u^2 c}{2} + \int_{\mathbb{R}} (e^{iux} - 1 - iux 1_{\{|x|<1\}}) v(dx)].$$

The triplet (b,c,ν) is called the Lévy or characteristic triplet and the exponent in the above formula

(2.1.5) $$\psi(u) = iub - \frac{u^2 c}{2} + \int_{\mathbb{R}} (e^{iux} - 1 - iux 1_{\{|x|<1\}}) v(dx)]$$

https://doi.org/10.26219/heal.aueb.5795

is called the Lévy or characteristic exponent. Furthermore, b ∈ $\mathbb{R}$ is called the drift term, c ∈ $\mathbb{R}_{\geq 0}$ the gaussian or diffusion coefficient and ν the Lévy measure.

After taking the characteristic function of the random variables underlying the Lévy jump-diffusion using the Lévy-Khintchine formula we can conclude to the following equation

$$\mathbb{E}\left[e^{iuL_t}\right] = \exp\left[t\left(iub - \frac{u^2\sigma^2}{2} + \int_{\mathbb{R}} (e^{iux} - 1 - iux)\lambda F(dx)\right)\right]. \tag{2.1.6}$$

After comparing equation (1.4) and (1.6), we can deduce that the random variable L_t of the Lévy jump-diffusion has a Lévy triplet b = b · t, c = σ^2 · t and ν = (λF) · t.

From the Lévy-Khintchine formula, in general, it is noticeable that a Lévy process consists of three independent parts: (i) a linear deterministic part, (ii) a Brownian part and (iii) a pure jump part. The Lévy measure ν(dx) prescribes how the jump will occur.

To finalize this short journey to Lévy processes, we must define a very important component of this process and an equally important result. First, we will proceed to the well-known Lévy-Itô decomposition of a Lévy process and finally we will define the Lévy measure ν(dx).

Theorem 2.1.2 (Lévy-Itô Decomposition). Let consider a Lévy triplet (b, c, ν) where b ∈ $\mathbb{R}$ (the drift term), c ∈ $\mathbb{R}_{\geq 0}$ (the diffusion coefficient) and ν a measure satisfying ν ({0}) = 0 and $\int_{\mathbb{R}} (1 \wedge |x|^2)\nu(dx) < \infty$. Then there exists a probability space on which four independent Lévy processes exist, $L^{(1)}$, $L^{(2)}$, $L^{(3)}$ and $L^{(4)}$. The first one is a constant drift, the second one a Brownian motion, the third is a compound Poisson process while the fourth is a square integrable martingale (a pure jump) with an a.s. countable number of jumps of magnitude less than 1 on each finite time interval. After taking the summation of these four independent Lévy processes and its characteristic exponent, we have the Lévy-Khintchine formula (2.6). In this paper we will not prove the Lévy-Itô decomposition, as it is a hard mathematical proof, but we will go

https://doi.org/10.26219/heal.aueb.5795

through some steps of the proof, because it contains some important information about the structure of the paths. Let consider the split of the Lévy exponent into four parts $\psi = \psi^{(1)} + \psi^{(2)} + \psi^{(3)} + \psi^{(4)}$where

- $\psi^{(1)} = iub$,
- $\psi^{(2)} = \frac{u^2 c}{2}$,
- $\psi^{(3)} = \int_{|x|\geq 1}\left(e^{iux} - 1\right)v(dx)$,
- $\psi^{(4)} = \int_{|x|<1}\left(e^{iux} - 1 - iux\right)v(dx)$.

In some books and papers, the first and second process presented as one combined linear Brownian motion and the Lévy exponent is divided into three parts, but in this analysis, we will proceed with the split into four parts. The first part is equivalent of a deterministic linear process (drift) with parameter b, the second one corresponds to a Brownian motion with coefficient $\sqrt{c}$ and the third part to a compound Poisson process with arrival rate $\lambda \equiv \nu(\mathbb{R}(-1,1))$ and jump magnitude $F(dx) \equiv \frac{\nu(dx)}{\nu(\mathbb{R}(-1,1))} 1_{\{|x|\geq 1\}.}$. The final part has some difficulties and need some extra clarification. Let $\Delta L^{(4)}$ denote the jumps of the fourth independent Lévy process and is equal to $\Delta L^{(4)} = L_t^{(4)} - L_{t-}^{(4)}$ and let $\mu^{L^{(4)}}$denote the random measure which counts the jumps. The next step is to construct a compensated compound Poisson process

(2.1.7) $$L_t^{(4,\varepsilon)} = \sum_{0\leq s\leq t} \Delta L_s^{(4)} 1_{\{1>\left|\Delta L_s^{(4)}\right|>\varepsilon\}} - t\left(\int_{1>|x|>\varepsilon} x\, \nu(dx)\right)$$

$$= \int_0^t \int_{1>|x|>\varepsilon} x\mu^{L^{(4)}}(dx, ds) - t\left(\int_{1>|x|>\varepsilon} x\, \nu(dx)\right)$$

and to showthat the jumps of $L^{(4)}$ form a Poisson process. Therefore, it is possible to decompose any Lévy process into four independent Lévy processes $L = L^{(1)} + L^{(2)} + L^{(3)} + L^{(4)}$ as follows

(2.1.8) $$L_t = bt + \sqrt{c}W_t + \int_0^t \int_{|x|\geq 1} x\mu^L(ds, dx) + \int_0^t \int_{|x|<1} x(\mu^L - v^L)(ds, dx)$$

https://doi.org/10.26219/heal.aueb.5795

Once again, the first part is a constant drift, the second one a Brownian motion, the third a compound Poisson process and the last one a pure jump martingale.

Definition 2.1.4. Let $L_{t\in[0,\infty)}$ be a Lévy process with a Lévy triplet (b, c, v) defined on a filtered probability space $(\Omega, \mathcal{F}_{t\in[0,\infty)}, \mathcal{P})$. The Lévy measure v is a unique positive measure on $\mathbb{R}$ which satisfies the following conditions v $(\{0\})$ = 0 and $\int_{\mathbb{R}} (1 \wedge |x|^2) v(dx) < \infty$. The Lévy measure describes the expected number of jumps of certain height in a time interval of length 1:

(2.1.9) $$v(A) = \mathbb{E}[\#\{t \in [0,1]: \Delta L_t = L_t - L_{t-} \neq 0, \Delta L_t \in A\}]$$

where $\Delta L_t \in A$ indicates that the jump size belongs to a set A and a set A is a Borel set. For more details, the reader may consult *Cont and Tankov (2004)* section 3.3.

Examples of Lévy Processes

In this final section, some examples of Lévy processes will be presented shortly. Brownian motion, Poisson process and a compound Poisson process will be defined and Lévy jump diffusion process as well, as the sum of a Brownian motion and a compound Poisson process.

Definition 2.1.5. A continuous-time stochastic process $\{W_t : 0 \leq t \leq T\}$ is called Brownian motion if it satisfies the following properties:

(i) $W_0 = 0$

(ii) The increments of W_t are independent meaning that for any set of times $0 \leq t_1 < t_2 < \cdots < t_n < T$ then the random variables $W_{t_2} - W_{t_1}, W_{t_3} - W_{t_2}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent.

(iii) For any 0 ≤ s < t ≤ T the increment $W_t - W_s$ has the Gaussian distribution with mean 0 and variance t-s.

(iv) For all ω in a set of probability one, $W_t(\omega)$ is a continuous function of t.

In Figure 2.1 below, a sample path of Brownian motion is exhibited, assuming all the properties mentioned above.

https://doi.org/10.26219/heal.aueb.5795

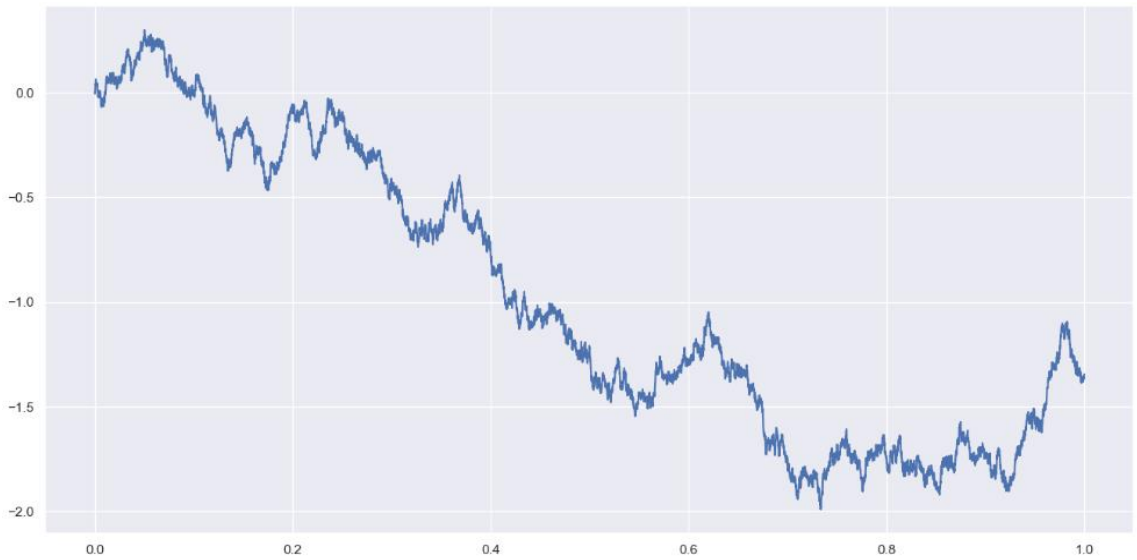


Figure 2.1 Simulation of a sample path of a Brownian Motion.

To define a Poisson process, it is necessary to first define the counting process.

Definition 2.1.6 A counting process is a stochastic process $\{C_t, t \geq 0\}$ with values that are non-negative, integer and non-decreasing:

(i) $C_t \geq 0$

(ii) C_t is an integer.

(iii) If s ≤ t, then $C_s \leq C_t$

Counting process is a stochastic process that keeps counts the number of events that have occurred up to time t.

Definition 2.1.7. A Poisson process $\{N_t, t \geq 0\}$, given any positive number λ, is a counting process with rate λ with the following additional properties:

(i) N (0) = 0

(ii) The process has stationary and independent increments.

(iii) $\mathbb{P}(N_t = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, n = 0,1,2, \ldots.$

The last property states that the number of events in any interval of length t follows a Poisson distribution with mean λt. Figure 2.2 demonstrates different Poisson processes with different intensity rates λ. It can be observed that the highest the intensity rate is, the more discontinuities the process has.

https://doi.org/10.26219/heal.aueb.5795

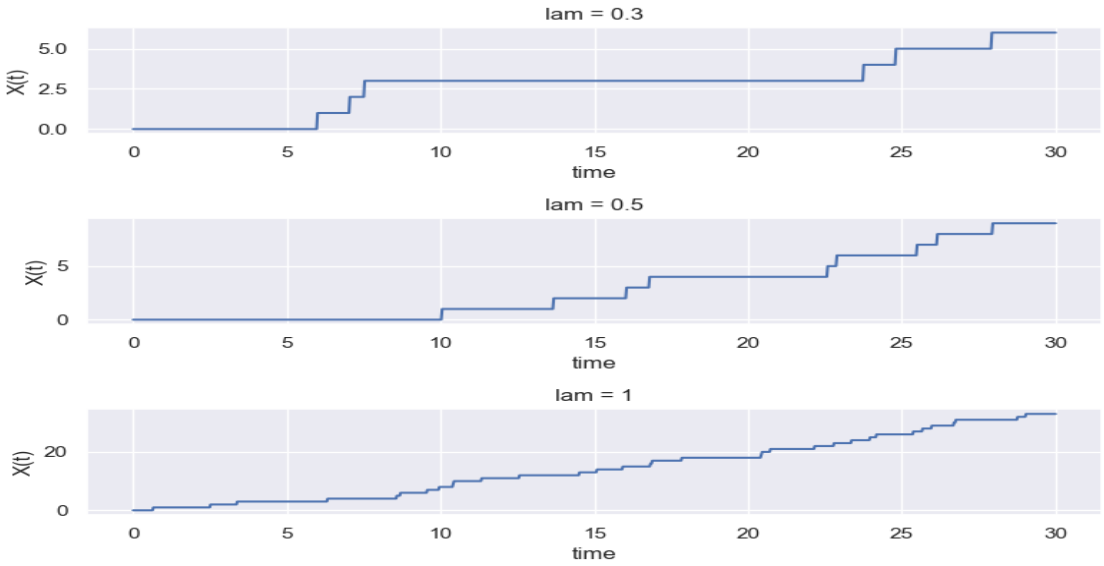


Figure 2.2 Simulation of Poisson Processes with different intensity rates λ.

Definition 2.1.8. A Compound Poisson process, associated with a given Poisson process N_t and a given sequence of independent, identically distributed (i.i.d) random variables taking values in $\mathbb{R}$ $\xi_k, k \in \mathbb{N}$ with common distribution and independent of the underlying Poisson process, is the stochastic process $\{J_t : t \in [0, \infty)\}$ where $J_t = \sum_{k=1}^{N_t} \xi_k$.

A compound Poisson process shares some properties with the underlying Poisson process, since J_t has stationary and independent increments, and the mean and variance of J_t are:

1. $\mathbb{E}(J_t) = \mu\lambda t$
2. $Var(J_t) = (\mu^2 + \sigma^2)\lambda t$

where μ and σ are the expectation and the variance for the random variable ξ_k. The graph in Figure 2.3 depicts the trajectory of a compound Poisson process assuming exponential jump sizes and how the value of the process evolves through time.

https://doi.org/10.26219/heal.aueb.5795

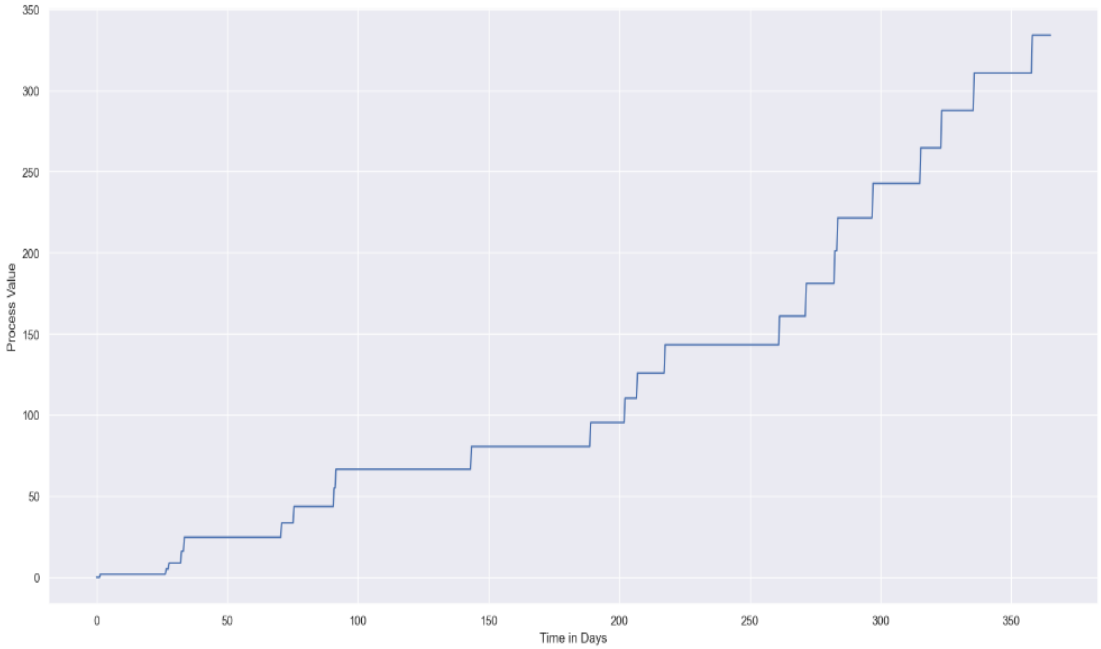


Figure 2.3 Simulation of a Compound Poisson Process with exponential jump sizes with intensity rate λ=0.05 and scale of exponential distribution alpha = 1.

Definition 2.1.9. A Lévy jump-diffusion process is a process of the following form

(2.1.10) $$L_t = bt + \sigma W_t + (\sum_{k=1}^{N_t} \xi_k - t\lambda\kappa)$$

where $b \in R$, $\sigma \in R_{\geq 0}$, W_t is a standard Brownian motion, N_t is Poisson process with parameter λ and ξ_k (with $k \geq 1$) is an i.i.d. sequence of random variable with probability distribution F and $E[J] = \kappa < \infty$. Hence, F describes the distribution of the jumps, which arrive according to the Poisson process. Figure 2.4 demonstrates a trajectory of a jump-diffusion process with a Brownian (gaussian) component and a finite jump intensity, through a Poisson process while the distribution of the jumps come from the normal distribution.

https://doi.org/10.26219/heal.aueb.5795

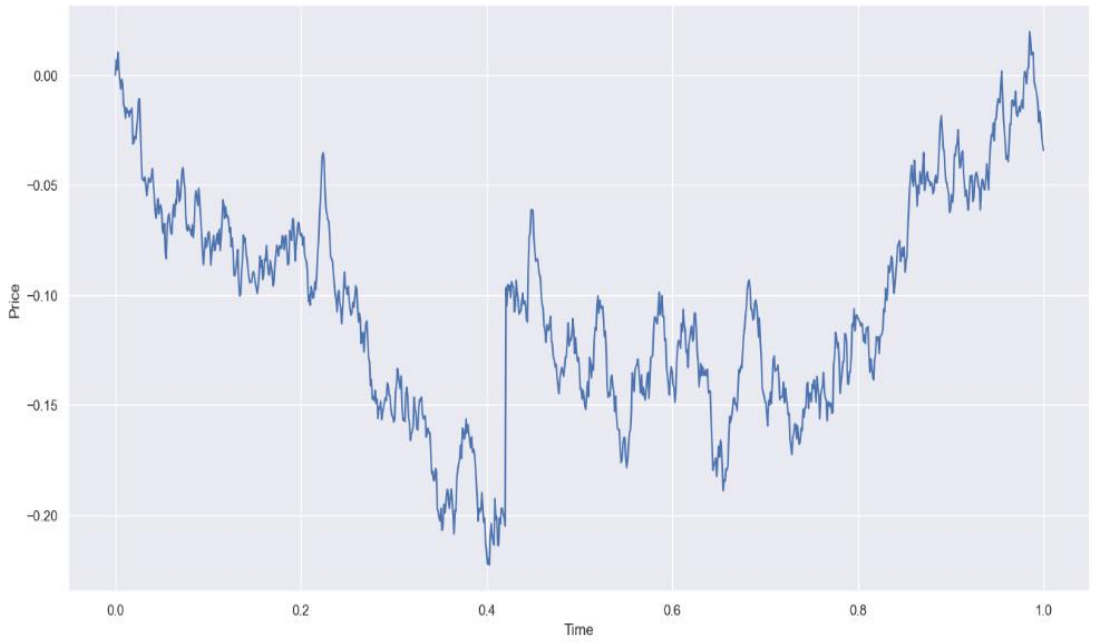


Figure 2.4 Simulation of a Lévy jump-diffusion process consisting of a Brownian part and a Compound Poisson process with normally distributed jumps (μ=0.1, σ=0.2, λ=1, μ_j=0.05, σ_j=0.1)

2.2. Ornstein-Uhlenbeck Process

The Ornstein-Uhlenbeck process (throughout: OU) was first proposed by *Uhlenbeck and Ornstein* (1930) with application in physics, modelling the velocity of a massive Brownian particle under the influence of friction. Since the original paper's appearance, the OU process has been used in a wide variety of applications and more specifically in financial mathematics, where it used to model interest rates stochastically. It is best known for their connection with the *Vasicek* model (1977).

Definition 2.2.1. A stochastic process $\{X_t: t \geq 0\}$ is an OU process if it is stationary, Gaussian, Markovian and continuous in probability. An OU process satisfies the following stochastic differential equation:

(2.2.1) $$dX_t = -\theta X_t dt + \sigma dW_t$$

where $\{W_t: t \geq 0\}$ is a Brownian motion and $\theta > 0$ and $\sigma > 0$ are constants. Sometimes an additional drift is added, following the below stochastic differential equation:

(2.2.2) $$dX_t = \theta(\mu - X_t)dt + \sigma dW_t$$

https://doi.org/10.26219/heal.aueb.5795

where κ again is a constant. In Figure 2.5, various OU processes have been shown with different parameters. In every three graphs it is assumed constant drift (μ) and the level of the volatility (σ) is changed.

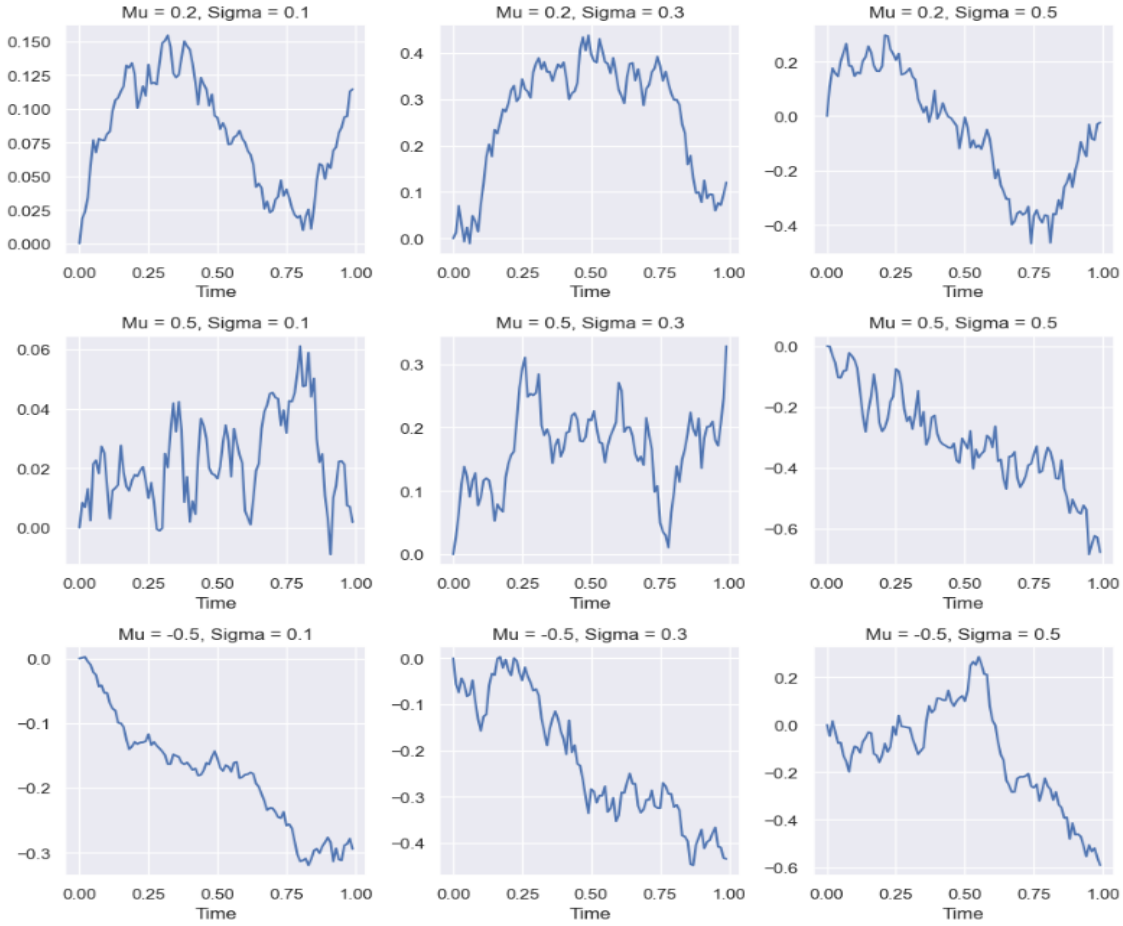


Figure 2.5 Simulation of Ornstein Uhlenbeck processes with different levels of mean and volatility with θ=0.5.

Over time, the process tends to drift towards its long-term mean and such process is called mean-reverting. Moreover, in several approaches in which the OU process is used to model interest rates, currency exchange rates and commodity prices stochastically, the parameter μ represents the mean value supported by fundamentals, σ the degree of volatility caused by fluctuations of the asset and θ the rate by which these fluctuations dissipate, and the variable reverts towards the mean. In the same context, in Figure 2.6, in every three graphs, we assume similar drift and varying volatility, but, in addition, different values of the mean-reverting parameter (θ) are assumed, too.

https://doi.org/10.26219/heal.aueb.5795

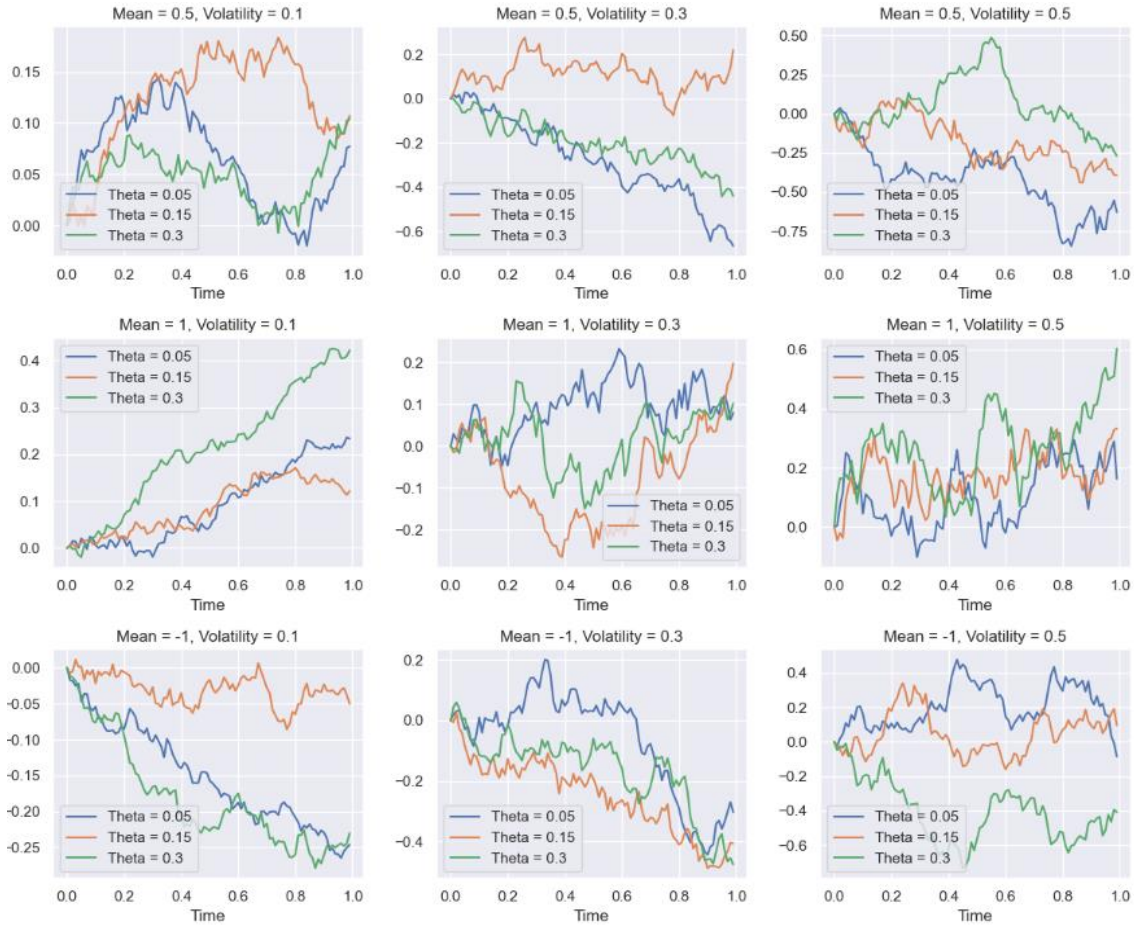


Figure 2.6 Simulation of Ornstein Uhlenbeck processes with different levels of mean, volatility, and mean reversion parameter (θ).

Mathematical Properties

The OU process is a Gaussian process that has a bounded variance, and its probability distribution is stationary. The difference between a Brownian process and an OU process is the drift term. While in a Brownian motion the drift term is constant, in the OU process it is dependent on the current value of the process.

The moments of the process conditioned (asymptotically stationary case) where X_0 is initially constant are given by:

$$E(X_t|X_0) = X_0 e^{-\theta t} + \mu(1 - e^{-\theta t}) \quad (2.2.3)$$

$$Cov(X_s, X_t|X_0) = \frac{\sigma^2}{\theta} (e^{-\theta|s-t|} - e^{-\theta(s+t)} \quad (2.2.4)$$

https://doi.org/10.26219/heal.aueb.5795

Additionally, the solution of the stochastic differential equation will be presented using variation of parameters method. As a first step, we are changing variable $f(x_t, t) = x_t e^{-\theta t}$ and we get $df(x_t, t) = \theta x_t e^{-\theta t} + e^{\theta t} dx_t = e^{\theta t}\, \theta\mu dt + \sigma e^{\theta t}\, dW_t$.

After integrating from 0 to t we get $x_t = x_0 e^{\theta t} + \mu\left(1 - e^{-\theta t}\right) + e^{-\theta t} \int_0^t \sigma\, e^{\theta s}\, dW_s$. Since the Itô integral of deterministic integrand is normally distributed, it follows that $x_t = x_0 e^{\theta t} + \mu\left(1 - e^{-\theta t}\right) + \frac{\sigma}{\sqrt{2\theta}} W_{1-e^{-2\theta t}}$.

Generalized OU Processes

There are many generalizations of Gaussian OU process but in this analysis, we will focus on a specific class, which has application in financial modelling. Our analysis will be restricted to a Lévy generalization, in which the Brownian motion component will be replaced by a Lévy process L_t.

Definition 2.2.2. Let $\{L_t : t \geq 0\}$ is a Lévy process and $\{X_t : t \geq 0\}$ a Lévy - driven OU process which satisfies the following stochastic differential equation given by

(2.2.5) $$dX_t = \theta(\mu - X_t)dt + \sigma dL_t$$

where θ is a constant and denotes the mean-reverting rate, μ is also a constant and denotes the long-run mean and finally σ denotes the volatility of X_t.

The solution of this stochastic differential equation can be obtained after applying Ito's formula on the function $f(t, X_t) = X_t e^{\theta t}$. Thus, the solution is

(2.2.6) $$X_t = X_0 e^{-\theta t} + \int_0^t \theta\mu e^{-\theta(t-s)} ds + \int_0^t \sigma\, e^{-\theta(t-s)}\, dL_s.$$

2.3 Affine Diffusion Processes

In this section, we describe a class of SDEs for which the characteristic function, $\varphi_X(u; t, T)$, most of the times, can be found in closed form. The class of affine diffusion processes (AD) is a part of the more general affine jump diffusion processes (AJD). Roughly speaking, an affine process is a Markov process whose characteristic exponents (the logarithm of the characteristic function) are affine with respect to the initial state. Before defining the class

https://doi.org/10.26219/heal.aueb.5795

of affine diffusion and affine jump diffusion processes, it is necessary to first define the Markov process.

As usual, our starting point is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration$(\mathcal{F}_s, s \in I)$, where I is an index set and the time set T is $[0, \infty)$ (continuous time) and S a state space. Let assume a stochastic process $X = \{X_t : t \in T\}$ adapted to the filtration $\mathcal{F}_s$.

Definition 2.3.1. The random process X is a Markov process if $\mathbb{P}(X_{s+t} \in A|\mathcal{F}_s) = \mathbb{P}(X_{s+t} \in A|X_s)$ for all $s, t \in I$ and for each $A \in S$ with $s < t$.

This defining condition is known as the Markov property. The Markov property states that given the present state of the process, the future state is independent of the past. This means that the conditional distribution of X_{s+t} given $\mathcal{F}_s$ is the same as the conditional distribution of X_{s+t} just given X_s.

Theorem 2.3.1. The random process X is a Markov process if and only if $\mathbb{E}[f(X_{s+t}|\mathcal{F}_s)] = \mathbb{E}[f(X_{s+t}|X_s)]$ for all $s, t \in I$ and every $f \in \mathcal{B}$, where $\mathcal{B}$ is the Borel set.

Under the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Markov process X(t), in some space $D \subset \mathbb{R}^n$ an affine process can be expressed by the following SDE.

$$dX(t) = \mu\big(t, X(t)\big)dt + \sigma\big(t, X(t)\big)dW(t), 0 \le t_0 \le t, \tag{2.3.1}$$

where W(t) is an independent Brownian motion. For processes in the AD class, a basic requirement is that the drift term and the covariance/diffusion term are of the affine form meaning that $\mu\big(t, X(t)\big) = \alpha_0 + \alpha_1 X(t)$ for any $(\alpha_0, \alpha_1) \in \mathbb{R}^n \times \mathbb{R}^{n\times n}$ and $\sigma\big(t, X(t)\big)\sigma\big(t, X(t)\big)^T = c_0 + c_1 X(t)$ with $(c_0, c_1) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n\times n}$. It can be shown that the discounted characteristic function of an AD process X(t) has the following form:

$$\varphi_X(u; t, T) = E\left(e^{izX(t)}\middle|X(s)\right) = e^{\varphi(s,t,z)+\psi(s,t,z)X(s)} \tag{2.3.2}$$

for some coefficients $\varphi(s, t, z)$ and $\psi(s, t, z)$.

To simply translate the above properties, an AD process is a real-valued Markov process which has the below properties.

https://doi.org/10.26219/heal.aueb.5795

- Whenever s<t the characteristic exponents of X_t and X_s are affine with respect to X_s
- There is affine dependence on the current state X_t of the drift, of the diffusion term and of the jump measure (if exists any).

A basic affine jump diffusion process (AJD) is a stochastic process X of the form

(2.3.3) $$dX_t = \kappa(\theta - X_t)dt + \sigma\sqrt{X_t}dW_t + dJ_t, t \geq 0, X_0 \geq 0$$

where W_t is a standard Brownian motion and J is an independent compound Poisson process. An AJD process possesses the same properties mentioned above, as the drift term, the diffusion term and the jump intensity should be of the affine form, i.e.

(2.3.4) $$J(t, X_t) = l_0 + l_1 X_t, with\ (l_0, l_1) \in \mathbb{R}^n \times \mathbb{R}^n$$

The above definitions are more intuitive definitions. The purpose of them is for the reader to comprehend AD and AJD processes to a level that would be adequate to use them in next chapters. For a strict mathematical definition of AD and AJD processes the reader may consult *D.Duffie, D. Fillipovic* and *W. Schachermayer Affine Processes and Applications in Finance (2003)* section 2.

2.4 Discretization Scheme

Typically, the dynamics of a stock price and interest rates are assumed to be driven by a continuous time stochastic process, which are given by a stochastic differential equation. The simulation of these processes is done at discrete time steps. To be more precise, the simulation methods are based on discrete approximations of the continuous solution to an SDE. To model a continuous time process, we must rely on N discrete realizations. Thus, the sample period [0, T] is divided into N intervals and then the continuous time process has been replaced with a piecewise-constant process and in each interval $[t_n, t_{n+1})$ with $n = 1,2 \dots, N-1$ it is assumed that the process is changing from one interval to the other. We will present the Euler scheme as the main discretization method used in this analysis.

https://doi.org/10.26219/heal.aueb.5795

Euler-Maruyama Method

Suppose we have an SDE of the form:

(2.4.1) $$dX_t = \mu(t, X_t) + \sigma(t, X_t)dW_t$$

with initial deterministic value $X_{t0} = X_0$ and the discretization N of the interval $[0, T]$: $0 = t_0 < t_1 < \cdots < t_N = T$ and we wish to simulate values of Xt assuming that we do not know its distribution. Thus, if we divide the time interval $[t_0, T]$ into N intervals and the Euler scheme is equal to

(2.4.2) $$Y_{n+1} = Y_n + \mu(t_n, Y_n)(t_{n+1} - t_n) + \sigma(t_n, Y_n)(W_{t_{n+1}} - W_{t_n})$$

for $n = 1,2 \ldots, N-1$ with initial value $Y_0 = X_0$. Y is a continuous stochastic process satisfying the above iterative scheme known as the Euler approximation of X. Usually, the time increment $\Delta t = t_{n+1} - t_n$ is taken to be constant. Additionally, $\Delta W_n = (W_{t_{n+1}} - W_{t_n}) \sim N(0, \Delta)$.

Let assume now a jump-diffusion SDE of the following form

(2.4.3) $$dX_t = \mu(t, X_t) + \sigma(t, X_t)dW_t + c(X_{t-})dJ_t$$

with initial deterministic value $X_{t0} = X_0$ and $W = \{Wt, t \in [0, T]\}$ is a Brownian motion and we assume $J = \{Jt, t \in [0, T]\}$ is a compound Poisson process. We construct an equidistant time discretization $0 = t_0 < t_1 < \cdots < t_N = T$ and we consider Y the solution of the jump-diffusion SDE using the Euler scheme given below.

(2.4.4) $$Y_{n+1} = Y_n + \mu(t_n, Y_n)(t_{n+1} - t_n) + \sigma(t_n, Y_n)\left(W_{t_{n+1}} - W_{t_n}\right) + c(t_n, Y_n)(J_{n+1} - J_n)$$

In the above solution $\Delta W_n = (W_{t_{n+1}} - W_{t_n}) \sim N(0, \Delta)$ is the nth increment of the Brownian motion and $\Delta J_n = (J_{t_{n+1}} - J_{t_n})$ is the nth increment of the compound Poisson process, which can be expressed as

(2.4.5) $$\Delta J_n = \sum_{i=N_{t_n}+1}^{N_{t_{n+1}}} \xi_i.$$

https://doi.org/10.26219/heal.aueb.5795

Here N is a Poisson process with intensity λ and ξ_i is a given sequence of independent, identically distributed (i.i.d) random variables with common distribution function.

https://doi.org/10.26219/heal.aueb.5795

Chapter 3

3.Vasicek Model: A Brief Review

In this section we will proceed with a brief introduction to the well-known Vasicek model (*Vasicek 1977*). The Vasicek model is one of the earliest stochastic models of the term structure of interest rates. It assumes that the instantaneous interest rate r_t evolves as an OU-process with constant coefficients, under the risk neutral measure, given by the below SDE:

(3.1) $$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t$$

where r_t represents the interest rate, W_t is a standard Brownian motion which models the random market risk effect, i.e., the non-predictable part of the market. Figure 3.1 shows the evolution of a spot rate through time assuming that the spot rate follows the (3.1) SDE. More specifically, the Vasicek model has been simulated using the Euler's method, as explained above.

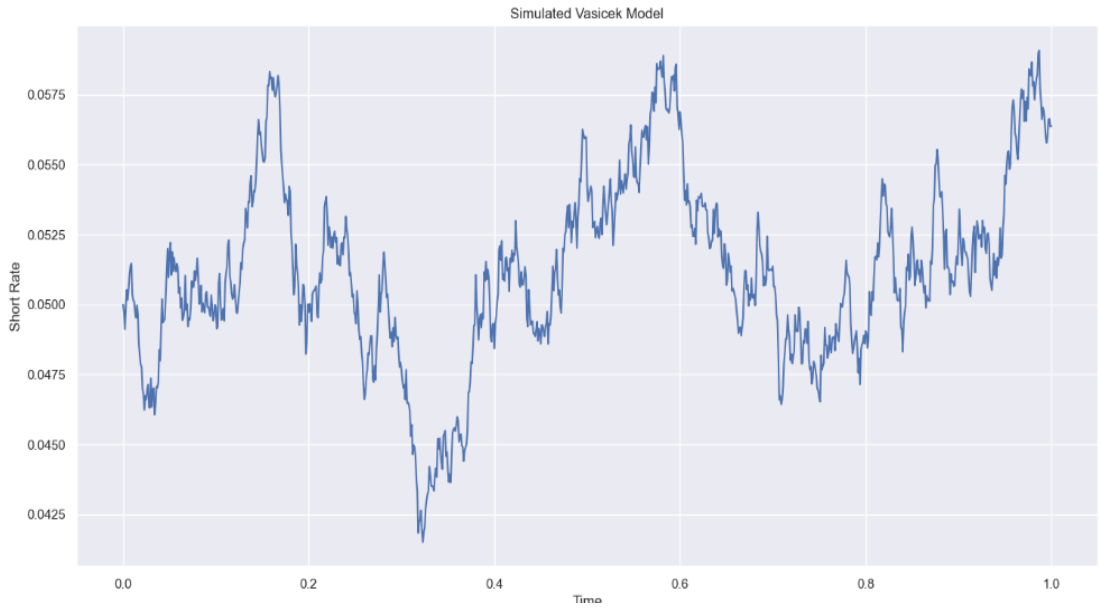


Figure 3.1 Simulation of Vasicek process-model implementing Eurel's method with assumed parameters (r0 = 0.05, k =0.1, θ = 0.07, σ =0.02).

The parameters κ, σ and θ characterize the dynamics of the model as follows:

- θ represents the value to which the interest rate r_t will converge in the long term; known as "long-term mean";

https://doi.org/10.26219/heal.aueb.5795

- κ characterizes the speed at which the interest rate r_t will converge to the long-term mean θ; higher values of κ mean that r_t will return very fast to the long-term mean; known as "mean-reversion speed";
- σ is the constant volatility of the process, the factor that controls the impact of randomness in the system; higher values of σ will result to higher randomness in the dynamics because the contribution of W_t will be greater.

Here θ, σ and κ are all positive constant. The Vasicek model exhibits mean reversion, as the instantaneous drift term $\kappa(\theta - r_t)$ describes the expected change in the interest rate and, also, the part of the model that indicates how quickly the interest rate reverts to the long-term mean level.

Three different trajectories, assuming the Vasicek model, are displayed in Figure 3.2, by implementing various scenarios of different long-term mean, mean-reversion speed and volatility.

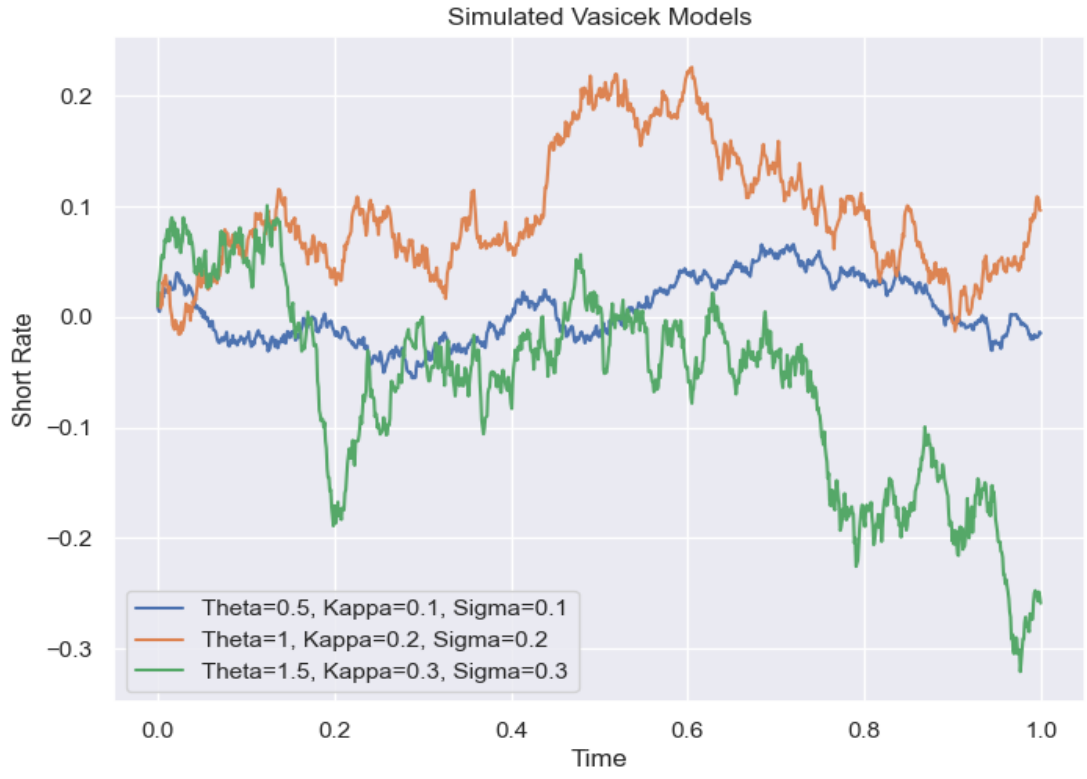


Figure 3.2 Simulation of different trajectories of Vasicek Model assuming various levels of mean reversion parameter k, long-term mean θ and volatility σ.

https://doi.org/10.26219/heal.aueb.5795

An attractive feature of this model is that a closed form solution for r_t is available, and it can be obtained by integrating the (4.1) SDE above. The solution of the model is using Ito's lemma, for each $s \leq t$

(3.2) $$r(t) = r(s)e^{-\kappa(t-s)} + \theta\left(1 - e^{-\kappa(t-s)}\right) + \sigma \int_s^t e^{-\kappa(t-u)} dW(u).$$

The interest rates assumed to be normally distributed, and the expectation and variance are given by:

(3.3) $$\mathbb{E}[r(t)] = r(s)e^{-\kappa(t-s)} + \theta\left(1 - e^{-\kappa(t-s)}\right)$$

(3.4) $$Var[r(t)] = \frac{\sigma^2}{2\kappa}\left[1 - e^{-2\kappa(t-s)}\right]$$

The drawback of the normality assumption is that it is possible for the interest rate to become negative. Nevertheless, as long $\kappa>0$ and $t\to\infty$, then the limit of expected rate and variance will converge to θ and $\frac{\sigma^2}{2\kappa}$ respectively.

Vasicek Yield Curve

We assume that the process of short rate r_t satisfies stochastic differential equation, with the solution given above by equation. To extract the yield curve the first step is to change the probability measure to risk neutral. This step is crucial since in bond pricing we wish to extinct the arbitrage possibilities, satisfying the non-arbitrage condition, which is only possible under the risk neutral probability. In this section we will obtain the Vasicek bond price formula, therefore the yield curve.

The Vasicek model possess an affine term structure having the following form

(3.5) $$p(t,T) = F(t, r_t; T) = e^{A(t,T) - r(t)B(t,T)}$$

(3.6) $$B(t,T) = \frac{1}{\kappa}(1 - e^{-\kappa(T-t)})$$

(3.7) $$A(t,T) = \frac{(B(t,T) - T + t)(\kappa\theta - \frac{1}{2}\sigma^2)}{\kappa^2} - \frac{\sigma^2 B^2(t,T)}{4\kappa}$$

https://doi.org/10.26219/heal.aueb.5795

where A and B are deterministic functions and t is the time value of the discount bond with maturity data T. For a model having an affine term structure, is an attractive characteristic from analytical point of view.

https://doi.org/10.26219/heal.aueb.5795

Chapter 4

4. Forecasting Financial Time-Series

In this chapter, methods and techniques will be briefly presented either from time-series analysis or from statistical-machine learning field applied to financial data. Financial data and their characteristics have been extensively studied in the past. These empirical studies have shown that they possess some properties known as 'stylized facts". Before we proceed further, we should first fix some notation. Let S_t denote the price of a financial asset, a stock price or an exchange rate for example, and $X_t = \ln S_t$ its logarithm. The log-return assuming a time scale Δt is defined as below:

(4.01) $$r(t, \Delta t) = X(t + \Delta t) - X_t \text{ or } r_t = \ln(\frac{S_t}{S_{t-1}})$$

Additionally, let realized or historical volatility σ of a financial asset defined as the standard deviation of the returns of this particular asset.

Figure 4.1 shows the time plots of the spot price of EUR/USD exchange rate during the period 2000-2022 (upper plot) and their log-returns (lower plot).

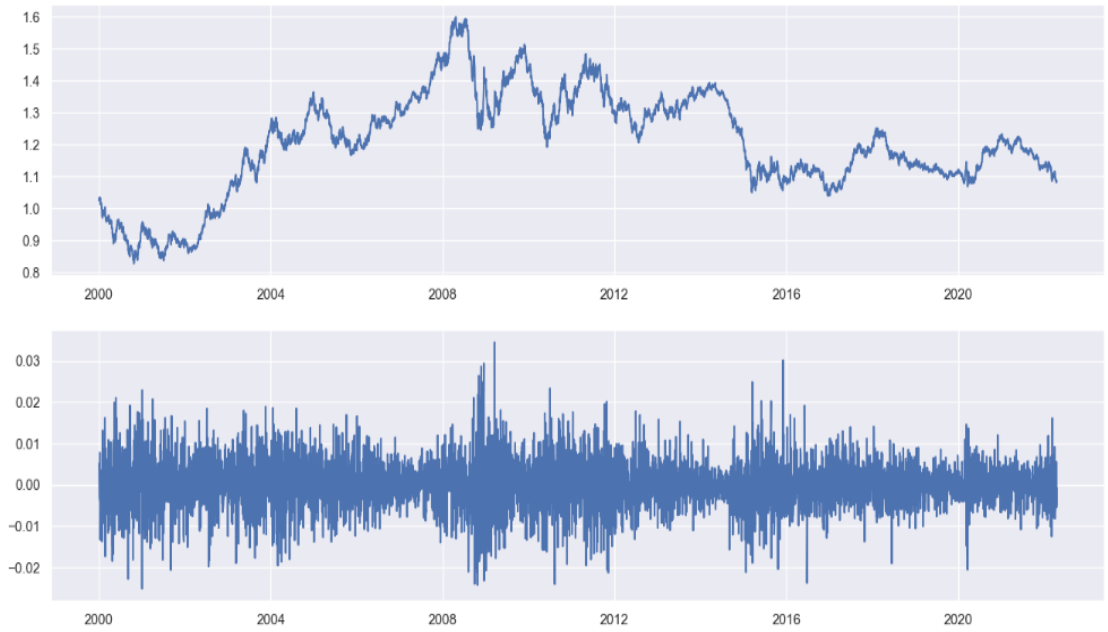


Figure 4.1 Time plots of daily EUR/USD exchange rate from January 2020 to April 2022. The upper panel is the spot price, and the lower panel is for log-returns.

https://doi.org/10.26219/heal.aueb.5795

First, the asset returns are characterized by the absence of autocorrelations. Most of the time, linear autocorrelations are insignificant. Figure 4.2 depicts the absence of autocorrelation in the daily log-returns of EUR/USD exchange rate. A second stylized fact is that financial time-series possess heavy tails, which can be observed either by plotting the kernel density of the distribution or by calculating the kurtosis. This characteristic is also related to the fact that the returns are non-normally distributed. Figure 4.3 below shows the probability density function of the normal distribution against the probability density function of the EUR/USD returns. It confirms the fact that returns have heavier tails than a standard normal distribution.

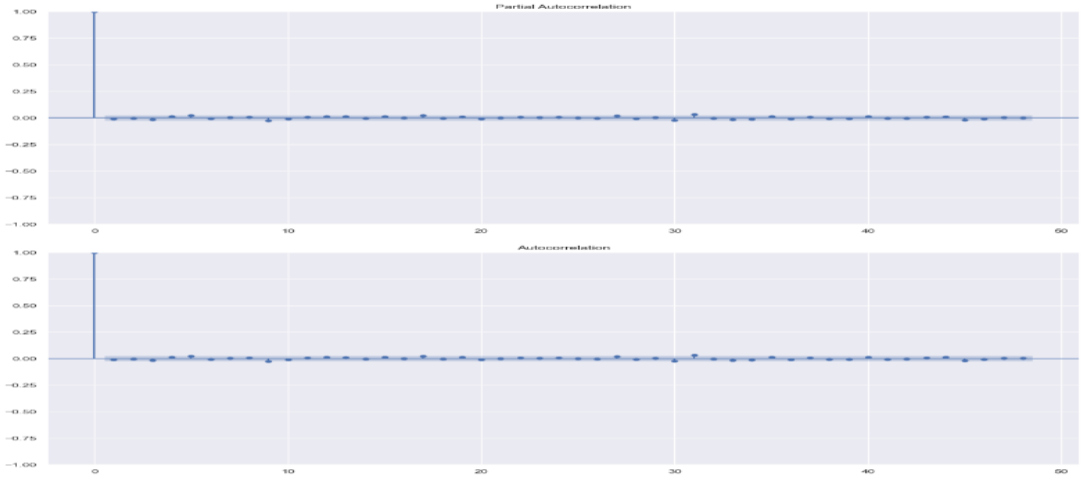


Figure 4.2 Upper panel partial autocorrelation function and lower panel autocorrelation function plot of log-returns of EUR/USD.

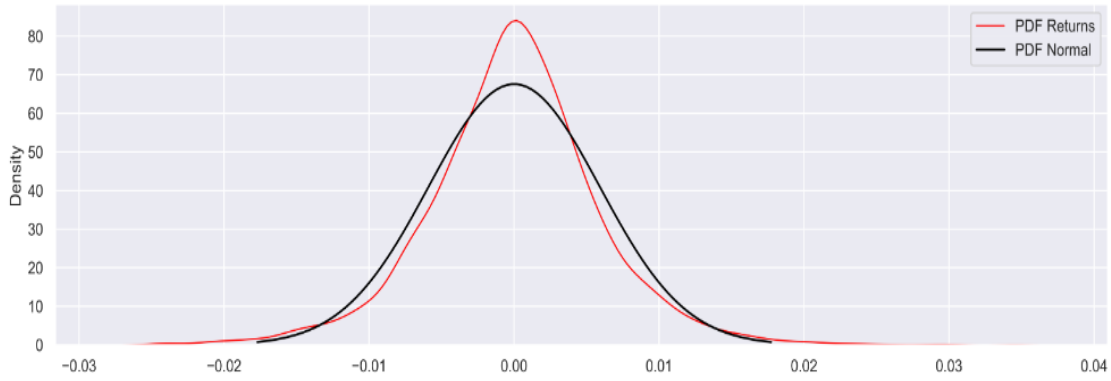


Figure 4.3 Comparison of empirical and normal densities for the daily log returns of EUR/USD.

https://doi.org/10.26219/heal.aueb.5795

Moving forward, a well-known stylized fact of financial data is volatility clustering. Per *Mandelbrot (1963)* "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes". High-volatility events tend to cluster in time and different measures of volatility exhibit a positive autocorrelation. An empirical way to test volatility clustering is to use the ACF plots of absolute returns or the ACF plots of squared returns. In Figure 4.4, we exhibit the ACF/PACF plots of absolute and squared returns of EUR/USD exchange rate. It is obvious that in this case the returns are serially dependent and that dependence volatility models aim to capture. Finally, another characteristic of the volatility is that it tends to react differently to high fluctuations in the financial asset, referred to the leverage effect. The leverage effect can also be interpreted as that volatility measures of an asset are negatively correlated with its returns.

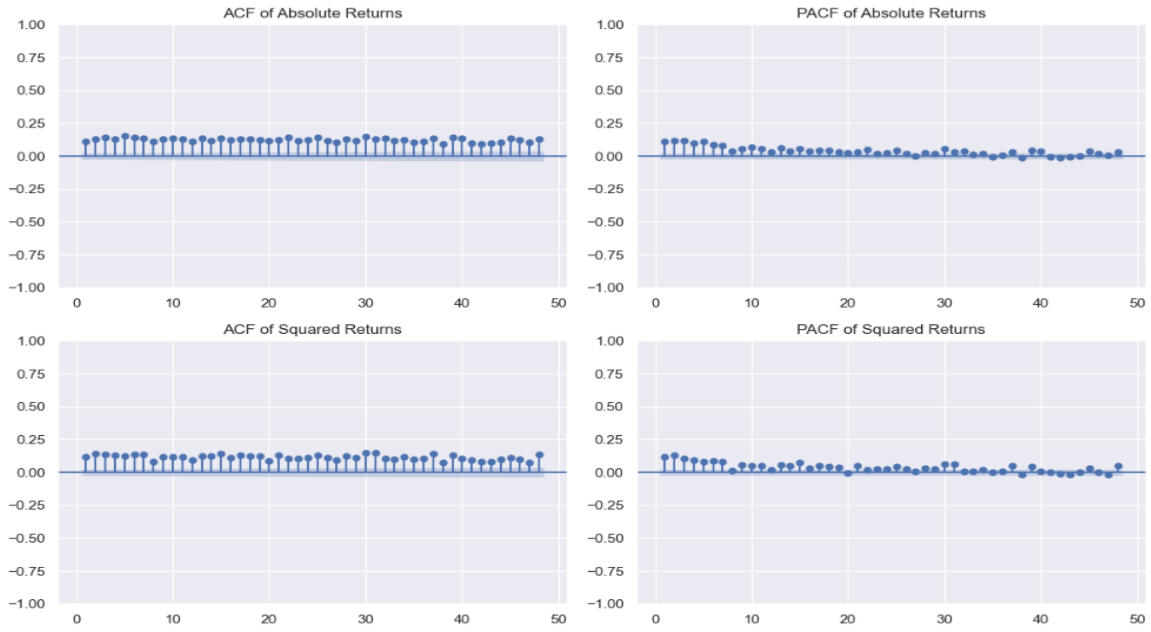


Figure 4.4 Autocorrelation and partial autocorrelation function of absolute and squared returns of EUR/USD.

Finally, before we proceed to the specification of the models used in this thesis, it is crucial to define first an integral part of time-series analysis, which is stationarity.

Let $\{X_t\}$ a time-series. $\{X_t\}$ is a strictly stationary series if the joint distribution of $X_{t1}, \ldots\ldots, X_{tk}$ is identical to the joint distribution of

https://doi.org/10.26219/heal.aueb.5795

$X_{t1+t}, \ldots\ldots, X_{tk+t}$ for all t, where k is an arbitrary positive integer. Strict stationarity requires that the joint distribution to be invariant under time shift. However, the strict stationarity condition is impossible to be verified in empirical application and that's why most of the times we assume a weaker version of stationarity. $\{X_t\}$ is a weakly stationary series if both the mean of the time-series and the covariance are time-invariant. More specifically, $\{X_t\}$ is a weakly stationary series if the below conditions stand

(4.0.2) $$\mathbb{E}[X_t] = \mu$$

(4.0.3) $$Cov(X_t, X_{t-k}) = \gamma_k$$

In other words, the expected value of the series is constant and the covariance between the series depends only on the lag k. In financial applications, we assume that the returns series is a weakly stationary time-series. In Figures 4.5-4.6 below, we can observe a non-stationary series, the spot price of EUR/USD exchange rate, and its rolling mean and a stationary series after taking the logarithm of the series and differentiate, building the returns of the EUR/USD exchange rate, along with the rolling mean of the returns series.

Figure 4.5 Evolution of spot price of EUR/USD, as a non-stationary series, along with its 25-day rolling mean.

https://doi.org/10.26219/heal.aueb.5795

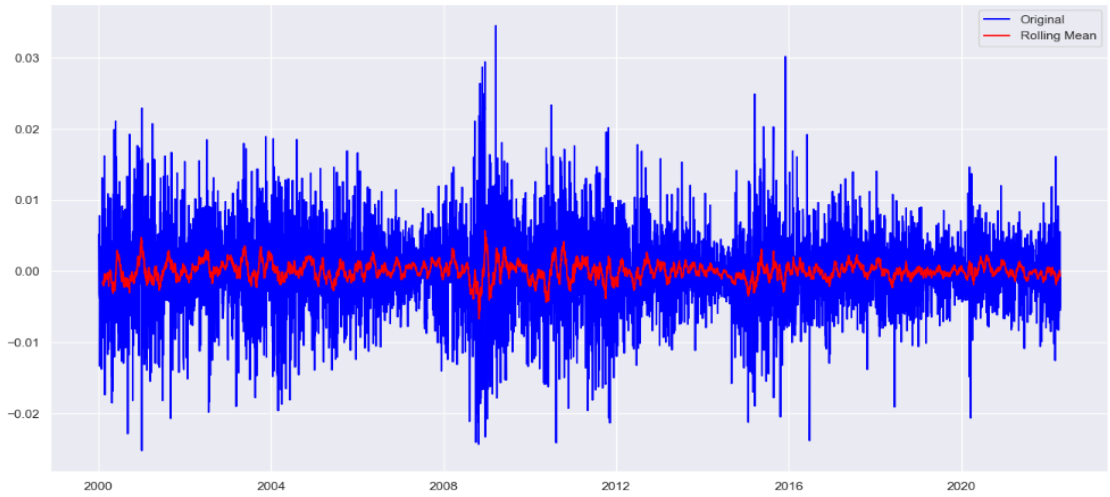


Figure 4.6 Log-returns of EUR/USD spot price, a stationary series, along with its rolling mean.

Moving forward, we will proceed to a brief description of time-series models such as auto-regressive models, auto-regressive moving-average models and conditional heteroscedastic models, such as ARCH-GARCH models, common models that describe the volatility of the asset return.

4.1 Time-Series Models

4.1.1 Simple Autoregressive Model

Let assume r_t is the log-return series up to time t.

(4.1.1) $$r_t = \varphi_0 + \varphi_1 r_{t-1} + \cdots + \varphi_p r_{t-p} + \varepsilon_t$$

where p is a non-negative integer and ε_t is a white noise series, meaning that the variables are independent and identically distributed with a mean equal to zero and variance equal to σ_ε^2. The interpretation of the AR(p) model above is that the past values given the past data jointly determine the conditional expectation of r_t. In our case r_t is a stationary process, so we can deduct its mean which is equal to

(4.1.2) $$\mathbb{E}[r_t] = \frac{\varphi_0}{1-\varphi_1-\cdots-\varphi_p}$$

provided that $(\varphi_1 + \cdots + \varphi_p) \neq 1$. Let's break down the above formula. Firstly, we assume the necessary condition of weakly stationarity, having that $\mathbb{E}[r_t] = \mu$, $Var(r_t) = \gamma_0$ and $Cov(r_t, r_{t-k}) = \gamma_k$. We can easily obtain the mean and

https://doi.org/10.26219/heal.aueb.5795

variance of the series as follows. Let's assume an AR (1) model $r_t = \varphi_0 + \varphi_1 r_{t-1} + \varepsilon_t$ and take the expectation of this model, we obtain

(4.1.3) $$\mathbb{E}[r_t] = \varphi_0 + \varphi_1 \mathbb{E}[r_{t-1}]$$

because $\mathbb{E}[\varepsilon_t] = 0$. Since the condition of weakly stationarity stands, $\mathbb{E}[r_t] = \mathbb{E}[r_{t-1}] = \mu$ and thus, $\mathbb{E}[r_t] = \mu = \frac{\varphi_0}{1-\varphi_1}$. If we apply the same steps and generalize this process for AR(p) model, we conclude to (4.1.1) formula above.

When it comes to the variance of an AR (1) model, we will use that $\varphi_0 = (1 - \varphi_1)\mu$ and in that way we can rewrite the AR (1) model as

(4.1.4) $$r_t - \mu = \varphi_1(r_{t-1} - \mu) + \varepsilon_t.$$

By repeated substitutions, the above formula can be represented by a series implementation,

(4.1.5) $$r_t - \mu = \sum_{i=1}^{\infty} \varphi_1^i \varepsilon_{t-1}.$$

The above equation implies that $r_t - \mu$ is a linear function of ε_{t-1} and since ε_t is an independent series we can conclude that $\mathbb{E}[(r_t - \mu)\varepsilon_{t+1}] = 0$. Assuming stationarity the covariance is dependent on the last lag, so it is equal to zero. The final step to obtain the variance of the AR (1) process is to take the square and after the expectation of the (4.1.4) equation, concluding to

(4.1.6) $$Var(r_t) = \varphi_1^2 Var(r_{t-1}) + \sigma_\varepsilon^2,$$

where σ_ε^2 is the variance of ε_t. Additionally, the covariance between the r_{t-1} and ε_t series is zero. Assuming the stationarity condition, then $Var(r_t) = Var(r_{t-1})$ which implies that:

(4.1.7) $$Var(r_t) = \frac{\sigma_\varepsilon^2}{1-\varphi_1^2}$$

In empirical applications, the main purpose is to determine the order p of an AR time-series model. In this analysis, two approaches will be the base for the order determination. The first one will be through the partial

https://doi.org/10.26219/heal.aueb.5795

autocorrelation function (PACF) and the second one will be through the Akaike Information Criterion (AIC).

Let assume AR models in consecutive orders as follows:

(4.1.8) $$r_t = \varphi_{0,1} + \varphi_{1,1} r_{t-1} + \varepsilon_{1t}$$

$$r_t = \varphi_{0,2} + \varphi_{1,2} r_{t-1} + \varphi_{2,2} r_{t-2} + \varepsilon_{2t}$$

$$r_t = \varphi_{0,3} + \varphi_{1,3} r_{t-1} + \varphi_{2,2} r_{t-2} + \varphi_{3,3} r_{t-3} + \varepsilon_{3t}$$

And so on, where $\varphi_{0,j}, \varphi_{i,j}\ and\ \varepsilon_{jt}$ are the constant term, the coefficient, and the error term of an AR (j) model. Assuming an estimation technique, such as least square method or MLE, the estimate $\hat{\varphi}_{1,1}, \hat{\varphi}_{2,2}, \hat{\varphi}_{3,3}$ are called the lag-1, lag-2 and lag-3 sample PACF of r_t, respectively and so on. Consequently, for an AR (p) model $\hat{\varphi}_{j,j}$ estimate should be close to zero for all $j > p$ and the lag-p sample PACF should not be zero. Using this property, it can be determined the order p of an AR model. In empirical applications, the order p of an AR model can be selected by plotting the PACF of the series and the lags fall out the standard-error limits, assuming a significance level, would be the order of the AR.

Furthermore, another approach is to determine the order p of AR process through information criteria. For this analysis, AIC, introduced by *Akaike (1973),* will be used. AIC is defined as

(4.1.9) $$AIC = 2k - 2\ln(\hat{L})$$

Where k is the number of parameters and $\ln(\hat{L})$ is the logarithm of maximized likelihood function. The first term of the AIC penalizes the model by the number of its parameters and it is called penalty function, while the second term measures the goodness of fit of the AR-model. The selection rule behind the use of AIC is that the model with the minimum AIC value will be selected as the optimal. AIC is a well-known information criterion for its predictive ability.

https://doi.org/10.26219/heal.aueb.5795

4.1.2 Simple Autoregressive-Moving Average Models

In most applications, AR models or Moving Average (MA) models may need higher orders with many parameters to fit the data well. *Box and Jenkins (1970)* introduced Autoregressive Moving Average (ARMA) models to overcome this difficulty. An ARMA model basically combines AR and MA models aiming to keep the number of parameters as small as possible. Additionally, ARMA models, that will also help to understand volatility modeling, will be introduced in the next section, while they are quite relevant to a GARCH model, which can be regarded as an ARMA process. Firstly, we will introduce very shortly the MA (q) process and afterwards we will proceed to the definition of the ARMA models.

Let r_t be the return series and ε_t the residual term with properties described above. An MA (q) model is

(4.1.10) $$r_t = c_o + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q},$$

where c_o is a constant. MA models are always weakly stationary, since they are linear combinations of a white noise process for which the first two moments are time-invariant. First, the constant term of an MA model is the mean of the process ($\mathbb{E}[r_t = c_o]$) and second the variance of the MA (q) model is

(4.1.11) $$Var(r_t) = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma_\varepsilon^2.$$

For further details, the reader may direct to *Ruey Tsay (2010)* Section 2.

After defining the MA (q) models and AR (p) models, the next step is to combine them and build the ARMA (p, q) model.

(4.1.12) $$r_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i r_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$$

where ε_t is the white noise (usually $\mathcal{N}(0, \sigma_\varepsilon^2)$). The main advantage of ARMA models is that they require fewer parameters than AR (p) and MA (q) models. The parameter estimation of the series is implemented by different methods, like least-square or maximum likelihood estimation (MLE). In Python, ARMA parameters have been estimated by using MLE, to find the values of

https://doi.org/10.26219/heal.aueb.5795

parameters, maximizing the probability of obtaining the data that we have observed. Similarly, to the AR-case explained above, the order of p and q is determined by PACF and in the case of Ma by ACF. Additionally, AIC is the information criterion used to determine the order of the ARMA models.

4.1.3 Conditional Heteroscedastic Models

Based on the stylized facts mentioned above, most of the returns series are not serially independent. This special characteristic is modeled by volatility models aiming to capture that dependence. The first model for volatility modeling was the ARCH model, build by *Engle (1982)*. Let z_t be the innovation of the return at time t and σ_t be the positive square root of σ_t^2. Assuming this notation an ARCH(p) model is defined as follows:

(4.1.13) $$z_t = \sigma_t \varepsilon_t,\ \sigma_t^2 = \omega + \alpha_1 z_{t-1}^2 + \ldots + \alpha_p z_{t-p}^2,$$

where $\{\varepsilon_t\}$ is the residual term, meaning an independent and identically distributed (iid) sequence of random variables with distribution $N(0,1)$ or in some cases a standardized t-student distribution. Additionally, α_i coefficients should meet some conditions to ensure that the variance of α_t is finite ($\omega > 0, and\ \alpha_i \geq 0$). To comprehend the ARCH models and its properties we will study the ARCH (1) model

(4.1.14) $$z_t = \sigma_t \varepsilon_t,\ \sigma_t^2 = \omega + \alpha_1 z_{t-1}^2,$$

where $\omega > 0, and\ \alpha_1 \geq 0$. The unconditional mean of z_t remains zero since $\mathbb{E}(z_t) = \mathbb{E}[\mathbb{E}(z_t|\mathcal{F}_{t-1})] = \mathbb{E}[\sigma_t \mathbb{E}(\varepsilon_t)] = 0$. The unconditional variance of z_t is calculated as follows $Var(z_t) = \mathbb{E}(z_t^2) = \mathbb{E}[\mathbb{E}(z_t^2|\mathcal{F}_{t-1})] = \mathbb{E}(\omega + a_1 z_{t-1}^2) = \omega + a_1 \mathbb{E}(z_{t-1}^2)$.

Since z_t is stationary process, it has constant mean $\mathbb{E}(z_t) = 0$ and its variance $Var(z_t)$ depends on the previous lag equal to $\mathbb{E}(z_{t-1}^2)$. Taking these into consideration we can conclude that the variance of a_t is equal to $\frac{\omega}{1-a_1}$. Furthermore, in financial applications we are interested in the tail behavior of the studied distributions or models. In that case, we are interested in higher order moments of z_t and we require its fourth moment to be finite. Provided

https://doi.org/10.26219/heal.aueb.5795

the unconditional moments of an ARCH (1) can be given by $\mathbb{E}(\varepsilon_t^2) = \frac{\omega}{1-a_1}$ and $\mathbb{E}(\varepsilon_t^4) = \frac{3\omega^2}{(1-a_1)^2}\frac{1-a_1^2}{1-3a_1^2}$, with $a_1 < 1\ and\ 3a_1^2 < 1$. By combining these moments, we can provide the kurtosis of the ARCH (1) model.

$$k = \frac{\mathbb{E}(\varepsilon_t^4)}{[\mathbb{E}(\varepsilon_t^2)]^2} = 3\frac{1-a_1^2}{1-3a_1^2}$$

In that case, the kurtosis is larger than 3 indicating that the tail distribution of z_t is heavier than the one of a normal distribution. These properties, of course, stand for ARCH (m) models. However, the computations of the higher order are much more complicated.

While the ARCH model was a breakthrough for volatility modelling, since it captures the volatility clustering stylized facts, it has several shortcomings. Firstly, it does not capture the leverage effect since the model assumes that positive and negative shocks affect the volatility of the returns similarly. Empirical applications have shown that this is far from the truth, while financial assets respond differently to positive and negative shocks. Additionally, the ARCH model has several limitations when it comes to capture excess kurtosis. From the $\mathbb{E}(\varepsilon_t^4)$ formula above, it can easily be derived that a_1^2 should be within $[0, \frac{1}{3}]$ interval in order the series to possess a finite fourth moment.

Because of its simplicity the ARCH model became popular, but due to its drawbacks, some alternative models should be sought. *Bollerslev (1986)* extended the ARCH model and proposed the generalized ARCH model, known as GARCH. Using the same notation as in the ARCH case, let z_t be the innovation at time t for a log-return time series. It can be said that the innovation follows a GARCH (p, q) model if

(4.1.15) $$z_t = \sigma_t\varepsilon_t, \sigma_t^2 = \omega + \sum_{i=1}^{p} a_i z_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

where again $\{\varepsilon_t\}$ is the residual term and $\omega > 0, \alpha_i \geq 0, \beta_j \geq 0, and\ \sum_{i=1}^{\max(p,q)}(a_i + \beta_j) < 1$. As before, $\{\varepsilon_t\}$ assumed to follow a standard normal or a Student t-distribution. To examine the properties of the GARCH

https://doi.org/10.26219/heal.aueb.5795

model, let assume the GARCH (1,1) model can be written a non-gaussian ARMA (1,1) model for the squared errors. Let $v_t = z_t^2 - \sigma_t^2$ and then:

(4.1.16) $$z_t^2 = \sigma_t^2 + (z_t^2 - \sigma_t^2) = \omega + a_1 z_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + v_t$$

$$= \omega + a_1 z_{t-1}^2 + \beta_1 (z_{t-1}^2 - v_{t-1}) + v_t = \omega + (a_1 + \beta_1) z_{t-1}^2 - \beta_1 v_{t-1} + v_t$$

From this representation, we can conclude that a GARCH (p, q) process can be regarded as an ARMA (p, q) process for the squared errors. Taking into consideration the moments of the GARCH (p, q) process, we can derive the unconditional mean assuming the ARMA representation,

$$\mathbb{E}[z_t^2] = \frac{\omega}{1 - \sum_{i=1}^{\max(p,q)} (a_i + \beta_i)},$$

under the constraint that $\sum_{i=1}^{\max(p,q)} (a_i + \beta_j) < 1$ in order the denominator to be positive. Provided the unconditional moments of a GARCH (1,1) process can be given by $\mathbb{E}[z_t^2] = \frac{\omega}{1 - a_1 - \beta_1}$ and $\mathbb{E}[z_t^4] = \frac{3\omega^2 (1 + a_1 + \beta_1)}{(1 - a_1 - \beta_1)(1 - 3a_1^2 - 2a_1\beta_1 - \beta_1^2)}$ then the kurtosis is given by

$$k = \frac{\mathbb{E}(\varepsilon_t^4)}{[\mathbb{E}(\varepsilon_t^2)]^2} = 3 + \frac{6a_1^2}{1 - 3a_1^2 - 2a_1\beta_1 - \beta_1^2}.$$

Thus, similarly to the ARCH model, GARCH has excess kurtosis larger than 3 and can capture the fat tail characteristics and the volatility clustering of financial data. However, it cannot capture the leverage effect. As a result, the GARCH model encounters the same problem as the ARCH model, even though it performs better in volatility modelling than the ARCH one.

Aiming to overcome the weakness of the GARCH model, *Nelson (1991)* proposed the exponential GARCH (E-GARCH) model. The E-GARCH model aims to overcome the lack of asymmetry in the response of shocks (leverage effect) encountered in GARCH model. Let's consider the same notation used in ARCH and GARCH models

(4.1.17) $$z_t = \sigma_t \varepsilon_t, \ln(\sigma_t^2) = \omega + \sum_{j=1}^{q} \beta_j \ln(\sigma_{t-j}^2) + \sum_{i=1}^{p} [\theta_i z_{t-i} + a_i (|z_{t-i}| - \mathbb{E}|z_{t-i}|)]$$

https://doi.org/10.26219/heal.aueb.5795

meaning that the logarithm of the conditional variance depends on lagged standardized errors, lagged absolute standardized errors and on lagged variances. The θ_i parameters signify the leverage effect of z_{t-i}. Of course, many models have been implemented as extensions of the GARCH models aiming to capture the stylized facts of financial time-series and aiming to model volatility in a more realistic and efficient way. In this chapter, the goal was to briefly present the volatility models and the basic theory behind them used in our analysis.

4.2 Machine Learning Approach

Machine Learning (ML) algorithms have been applied to forecasting financial time series, more frequently in recent years. The main goal of this section is to briefly describe the main concept of applied methods. In this thesis Recurrent Neural Network (RNN)-Long Short-Term Memory (LSTM) will be presented. Of course, there are many different efficient ML algorithms that may be used for the purpose of financial time-series analysis and forecasting.

4.2.1 Recurrent Neural Network - Long-Short Term Memory

Before we proceed to a brief introduction to Recurrent Neural Networks (RNN), let's first define some basic concepts of artificial neural networks (ANN). Firstly, the ANN has neurons that are linked to each other in various layers of the networks, which are known as nodes. The neurons are organized into multiple layers. The first layer is the input layer, the one that receives the data, while the final layer is the output layer that produces the result. In between these two basic layers, there are more hidden layers, in which the function applies weights to the inputs and directs them through an activation function as the output. Basically, an activation function is the function that delivers an output based on inputs. Some examples of activation function are the sigmoid $\sigma(x) = \frac{1}{1+e^{-x}}$ and hyperbolic tangent (tahn) $tahn(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$. Finally, the weights control the strength of the connection between two layers and mathematically they are matrices, while bias, which is a constant, guarantees that even when all inputs are zeros there will be an activation in the neuron.

https://doi.org/10.26219/heal.aueb.5795

A RNN is a type of artificial neural network, which uses sequential or time-series data. This type of neural networks is distinguished by their memory, as they extract information from prior inputs in order to influence the current input and output. Throughout network's life, the RNNs keep information that is relevant to various timesteps.

At each time step, the RNN takes an input vector, processed by the current neuron, and produces an output. The output of the precious time-step is fed back into the current neuron, providing it with a form of memory. In a RNN the information cycles through a loop, while it takes into consideration the current input and what it has learned from the inputs that received previously.

Let's consider an RNN with a hidden layer h_t, given an input layer x_t the hidden state at time t can be calculated as follows:

(4.2.1) $$h_t = f(W_x x_t + W_h h_{t-1} + b_h)$$

where W_x are the weights associated with the inputs, W_h are the weights associated with hidden units, b_h is the bias associated with the recurrent layer and a f is an activation function. The output $y_t$is computed as follows:

(4.2.2) $$y_t = f(W_y h_t + b_y)$$

where W_y are the weights associated with the outputs and b_y is the bias vector. These equations are applied recursively over the entire sequence, capturing dependences and patterns in the input data.

RNNs have the advantage of handling sequence data or time-series data and the ability to store historical information. However, it has some shortcomings. The main shortcoming is the vanishing gradient problem. The gradients are used to compute the weight update, and, in this procedure, they may get very close to zero, preventing the network from learning the new weights. Because of the vanishing gradient problem, RNNs lose memory in the long run. Long-Short Term Memory (LSTM) was designed by *Hochreiter and Schmidhuber* to handle the vanishing gradient problem in RNNs.

https://doi.org/10.26219/heal.aueb.5795

LSTM is a special kind of RNN, which adds memory storage (cell) and gates structure to the traditional RNN. Gates in a neural network act as a threshold for helping the network to distinguish when to use normal stacked layers versus an identity connection. Identity connection in neural networks is a type of connection that directly connects the output of one layer to the input of a subsequent layer, allowing to preserve the information and pass it through to deeper layers. A gate structure allows the network to selectively pass or block information based on a condition or a rule. A memory cell contains three gates, the forget gate, the input gate and output gate. These gates control the flow of information in the cells. More specifically, the forget gate $\{f_t\}$ determines the information discarded from the cell, input gate $\{i_t\}$ determines the new information added and output gate $\{o_t\}$ filters the output information.

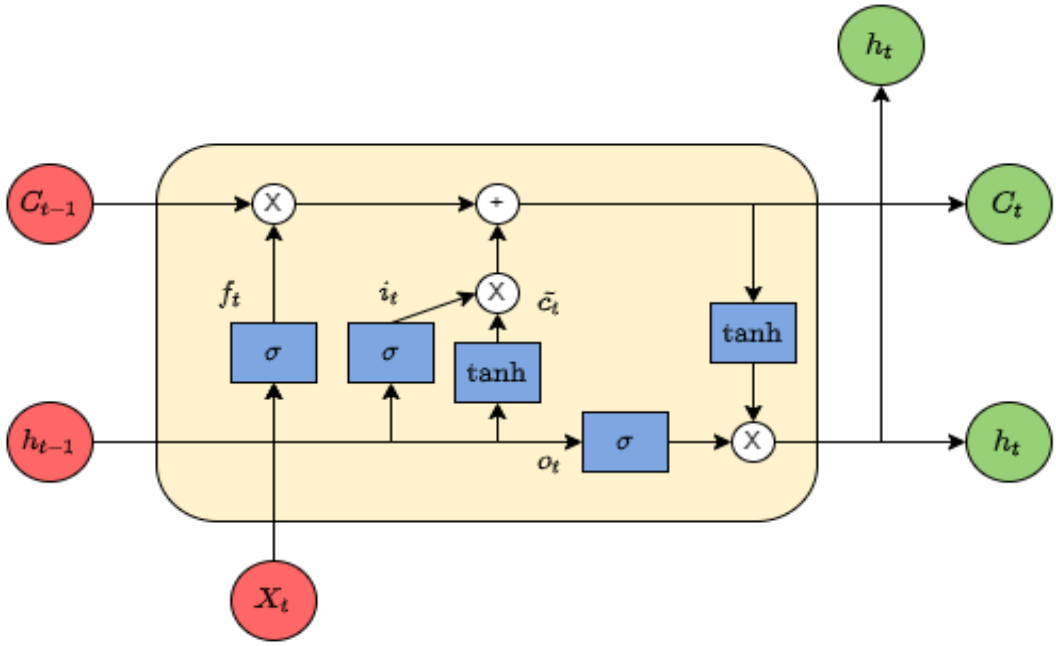


Figure 4.7 Structure of LSTM cell.

A LSTM cell, at each time step t, takes an input vector X_t and a hidden state vector h_{t-1} as inputs. By implementing the gated mechanisms explained above, it calculates the current hidden state h_t and the output y_t as follows:

(4.2.3) $$f_t = \sigma(W_f[h_{t-1}, X_t] + b_f$$

(4.2.4) $$i_t = \sigma(W_i[h_{t-1}, X_t] + b_i$$

https://doi.org/10.26219/heal.aueb.5795

(4.2.5) $$\tilde{C}_t = tanh(W_c[h_{t-1}, X_t] + b_c$$

(4.2.6) $$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t$$

(4.2.7) $$o_t = \sigma(W_o[h_{t-1}, X_t] + b_o$$

(4.2.8) $$h_t = o_t \times \tanh(C_t)$$

where W are weight matrices, h_{t-1} is the previous time step input, X_t the current time step input and b is the bias. Additionally, in the above notation the operator × denotes the elementwise product, known as Hadamard product. Equation 4.2.3 depicts how the forget gates operate and equation 4.2.4 demonstrates how the input gate determines which values are updated, both using a sigmoid activation function. The next step is the update gate (4.2.5) where a candidate memory cell is computed based on the input X_t and the previous output h_{t-1}, but implementing a tanh activation function, representing the new information of the memory cell. The actual cell state C_t is calculated based on the previous cell state and the candidate cell $\tilde{C}_t$ (4.2.6). Both forget and input gates determine how much information should be left out from the previous cell and how much should be retained from the candidate cell. Finally, based on the output gate o_t (4.2.7) and the current cell state, the output h_t is calculated using a than activation function.

https://doi.org/10.26219/heal.aueb.5795

Chapter 5

5.Black and Scholes Framework

In this chapter, it will be briefly presented the famous Black-Sholes model named after *Fischer Black* and *Myron Scholes (1973).* The Black-Scholes model (BS-model) aim to valuate European stock options, by giving a theoretical estimate of the price of a European option and showing that the option has a unique price assuming the volatility (risk) and the returns of the underlying asset. Before moving forward, let us clarify the basic terminology of financial derivatives and options. Firstly, financial derivatives are products that are based on the performance of an underlying asset. This underlying asset may be a stock, an interest rate, a foreign-exchange rate or a commodity price. An option contract is a financial contract that gives the right, but not the obligation, to buy or sell the underlying asset at a certain point of time in the future at a predetermined price. In the case of obtaining the option to buy, this is called a call option, while when one obtains the right to sell, it is called a put option. The most common option is the European style option. This type of option can only be executed on the day of maturity, meaning the last day of the predetermined contract. There are also American style options, which can be executed on or before the maturity. There are many types of options (Asian, lookback, barrier, digital, etc.) which are known as exotic option. In this chapter and in this thesis, we will deal with the European-style option, known as vanilla option as well. Throughout the thesis, the maturity time or expiry date will be denoted as T and the predetermined price, known as strike price, will be denoted as K.

Generally, the holder of a European option will exercise the option, when the asset price S_T is larger than the strike price (in case of call option) or is smaller than the strike price (in case of put option). The profit in the call option case is equal to $S_T - K$ while in the put option case is equal to $K - S_T$. The payoff function denoted as $V(T,S)$ will be equal to

$$V(T,S) := \begin{cases} \max(S_T - K, 0), if\ call\ option \\ \max(K - S_T, 0), if\ put\ option \end{cases} \tag{5.1}$$

https://doi.org/10.26219/heal.aueb.5795

In 1973 Black and Scholes in their paper assumed the pricing of European option with a riskless interest rate r_t while the underlying does not pay dividends. The assumptions we use to derive the Black-Scholes-Merton model can be find below:

a. The risk-free interest rate r is constant.
b. The underlying does not pay dividends.
c. The logarithmic returns of the assets follow a geometric Brownian motion (GBM) with constant drift and volatility.
d. Short selling is always possible.
e. There are no transactions costs.
f. There are no arbitrage opportunities.
g. The trading market is continuous.

The Black-Scholes framework assumes that the asset price follows a geometric Brownian motion. In a geometric Brownian motion, the logarithm of the quantity follows a Brownian motion. The Black-Scholes stochastic differential equation is given below:

(5.2) $$\frac{dS_t}{S} = \mu dt + \sigma dW_t$$

Where μ is the drift term of the asset and σ is the standard deviation of the asset, known as the volatility of the log-returns, while W_t is a Brownian motion. In the case of a European option, Black and Scholes derived an analytical solution to the Black and Scholes Equation (see below).

Let $V_C(T,S)$ denote the value of a call and $V_P(T,S)$ the respective value of a put option. The explicit analytical formulas for both contracts can be found below:

(5.3) $$V_C(T,S) = SN(d_1) - N(d_2)Ke^{-rT}$$

(5.4) $$V_P(T,S) = Ke^{-rT}N(d_2) - SN(-d_1)$$

where

(5.5) $$d_1 = \frac{\ln(S/K) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$$

https://doi.org/10.26219/heal.aueb.5795

(5.6) $$d_2 = \frac{\ln(S/K) + \left(r - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$

with cumulative standard distribution function (CDF) $N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz$.

The Black-Scholes equation mentioned above is a parabolic partial differential equation (PDE), which it has many solutions depending on the different derivatives products defined by S_T, the underlying variable. The particular financial derivative derived by the equation depends on the boundary conditions that are used. In our case of European-style options the boundary conditions are $V = \max(S_T - K, 0)$ in case of a call option and $V = \max(K - S_T, 0)$ in the case of put option. Equation (5.7) below is the famous Black-Scholes PDE:

(5.7) $$\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - rV = 0$$

For more specialized readers, the derivation of the Black-Scholes PDE using Ito's lemma can be found in *Hull, John (2008) Options, Futures and Other Derivatives*.

As mentioned above, the applications of the Black-Scholes model in financial derivatives are not limited to stocks, but also in other financial products. In this thesis, the empirical application will be in forex exchange rates (currencies). *Mark German* and *Steven Kohlhagen (1983)* provide an analytic valuation model for European options on currencies applying a similar approach used by *Merton* for European options on dividend-paying stocks. In Garman & Kohlhagen the list of assumption mentioned above in the Black & Scholes framework stand as well. The BS-formulas (5.3), (5.4) can be modified as follows:

(5.8) $$V_C(T, S) = Se^{-r_f T} N(d_1) - N(d_2) Ke^{-r_d T}$$

(5.9) $$V_P(T, S) = Ke^{-r_d T} N(d_2) - Se^{-r_f T} N(-d_1)$$

where

https://doi.org/10.26219/heal.aueb.5795

(5.10) $$d_1 = \frac{\ln(S/K)+\left(r_d - r_f + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$$

(5.11) $$d_2 = \frac{\ln(S/K)+\left(r_d - r_f - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$

with cumulative standard distribution function (CDF) $N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz$.

In this model, we have some different inputs compared to the standard Black-Scholes model. First, S is the FX spot rate and K is the FX strike rate, r_d is the domestic risk-free rate and $r_f$is the foreign risk-free rate. In Foreign Exchange Rates, known as FX, $S_t = FOR/DOM$ represents the number of units of domestic currency in exchange of one unit of foreign currency at time t. For example, EUR/USD = 1.067 can be interpreted as one Euro is worth 1.067 US Dollar. In this example, the foreign currency is EUR and USD is the domestic one.

In the previous chapter, we defined the realized/historical volatility as the standard deviation of log-returns of an asset. Volatility measures the magnitude of uncertainty regarding an asset. There is a distinction between two types of volatility, namely historical and implied volatility. Historical volatility results from the actual data (financial time-series), like the standard deviation of their returns. Implied volatility, on the other hand, is the volatility for which a financial model (in our case BS-model) returns the actual observed market price. Therefore, volatility is implied by the model. More specifically, in BS-formulas the only parameter that is not observable is the volatility of the asset. In this context, implied volatilities are used to monitor the market's opinion about the volatility. A basic distinction between historical and implied volatility is that the implied volatility is forward looking, while historical volatility is backward looking.

Merton Jump diffusion model is a result of *Rober Merton's 1979* paper *Option Pricing When Underlying Stock Returns Are Discontinuous*. The main concept of this model was to extend the Black-Scholes model to incorporate more realistic assumptions and to adjust to the fact that market returns do not follow a constant variance log-normal distribution. Merton Jump-Diffusion

https://doi.org/10.26219/heal.aueb.5795

(MJD) consists of a Brownian motion with a drift process (continuous diffusion part) and of an abnormal (discontinuous part, i.e., jump) component modeled by a compound Poisson Process. The MJD model of an asset price takes the SDE of the form:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t + d\left(\sum_{k=1}^{N(t)} J_k - 1\right) \tag{5.12}$$

where W_t is a standard Brownian motion, $N(t)$ is Poisson process with intensity rate λ and J_k is a sequence of independent identically distributed (i.i.d) non-negative random variables, while $\sum_{k=1}^{N(t)} J_k$ is a compound Poisson process with normal distributed jumps $N(\mu_j, \sigma_j^2)$. In the model, all stochastic processes are assumed to be independent. The solution of the above SDE gives the dynamics of the asset price:

$$S_t = S_0 e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t + \sum_{k=1}^{N_t} J_k} \tag{5.13}$$

From the solution of the MJD model (5.13) we take the logarithm of the asset price and derive the following:

$$\ln(S_t) = \ln\left(S_0 e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t + \sum_{k=1}^{N_t} J_k}\right)$$

$$= \ln(S_0) + \ln(e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma(W_{t+\Delta t} - W_t) + \sum_{k=1}^{N_t} J_k}$$

$$= \ln(S_0) + \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma(W_{t+\Delta t} - W_t) + \sum_{k=1}^{N_t} J_k$$

Then after the logarithmic returns of the asset price we can derive the MJD as shown above in 5.12. In Equation 5.12 μ is the drift or diffusion term. In BS model denotes the expected return of the asset. In MJD we substituted μ with $\mu - \lambda k$ and leading to the following SDE:

$$\frac{dS_t}{S_t} = (\mu - \lambda k)dt + \sigma dW_t + d\left(\sum_{k=1}^{N(t)} J_k - 1\right) \tag{5.14}$$

As in the above notation (see 5.12, 5.13) we can count on an explicit solution. In particular, we have

$$S_t = S_0 \exp\{\sigma W_t + (\mu - \lambda k - \tfrac{1}{2}\sigma^2)t\} \prod_{k=1}^{N_t} J_k \tag{5.15}$$

https://doi.org/10.26219/heal.aueb.5795

If we consider the log-return for the MJD model, conditional on the event $\{N_t = j\}$, we can write

(5.16) $$\log\left(\frac{S_t}{S_0}\right) = \sigma W_t + \left(\mu - \lambda k - \frac{1}{2}\sigma^2\right)t + \sum_{k=1}^{j} V_k$$

where $V_k = logJ_k \sim N(\mu_j, \sigma_j^2)$ and overall the $\log\left(\frac{S_t}{S_0}\right) \sim N((\mu - \lambda k - \frac{1}{2}\sigma^2 + j\mu_j, \sigma^2 t + j\sigma_j^2))$, expressed as a weighted sum of normal distributions.

Finally, for sake of completeness in equation 5.17 below, we present the MJD model as derived from Merton in his paper:

(5.17) $$\frac{dS_t}{S_t} = (\alpha - \lambda\kappa)dt + \sigma dW_t + (J_t - 1)dN_t$$

where α is the instantaneous expected return of the asset price, σ is the instantaneous volatility of the asset return, W_t is a standard Brownian motion and N_t is Poisson process with intensity λ.

https://doi.org/10.26219/heal.aueb.5795

Chapter 6

6. An Extension of Vasicek Model: Vasicek Jump-Diffusion Model

In chapter 3 we defined the Vasicek model and its application to yield curve. The Vasicek model is based on the following SDE:

(6.1) $$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t$$

In the same context as above the κ, θ and σ are constants and W_t is a Brownian motion. In this chapter we will extend the classical Vasicek model by adding a Jump-Diffusion process. The arrival of the jumps is driven by a Compound Poisson process N_t and the jump is random assuming a normally distributed size. In that case, the extended Vasicek Jump Diffusion model is explained by the following SDE:

(6.2) $$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t + JdN_t$$

Where θ is the long-term mean, κ is the mean-reversion coefficient and is positive, σ is the volatility of the short rate, W_t is the Brownian motion, J is the jump size and dN_t is the Compound Poisson process with intensity λ. Moreover, the Brownian motion and the Compound Poisson process are assumed to be independent. The jump size may be distributed normally $N(\mu, \sigma^2)$. Figure 6.1 demonstrates the Vasicek Jump-Diffusion Model assuming a normally distributed jump size with different levels of volatility and jump intensity, while the jump size follows a normal distribution with zero mean and different levels of jump standard deviation.

Applying the Euler scheme from section 2.4 we have the representation:

$$r_{t+1} = r_t + \kappa(\theta - r_t)\Delta t + \sigma\sqrt{\Delta t}W_t + \Delta t J_t$$

where Δt is this formula represents the time increments of the W_t and J_t respectively and $\Delta J_t = \sum_{i=N_t}^{N_{t+\Delta t}}(\xi_i - 1)$. $N_{t+\Delta t}$ is Poisson with arrival likelihood $\lambda\Delta t$ and for Vasicek Extended J is a $N(\mu, \gamma)$.

https://doi.org/10.26219/heal.aueb.5795

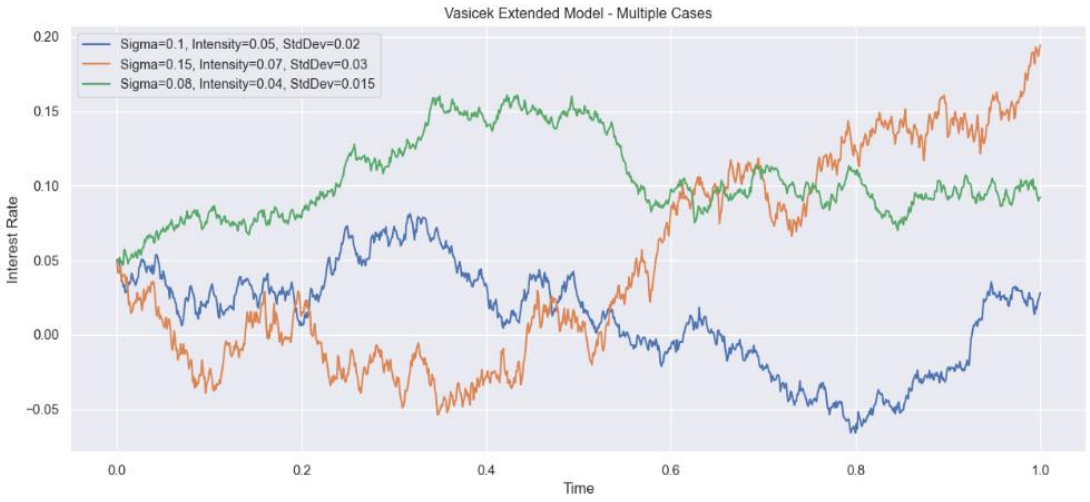


Figure 6.1 Simulation of Vasicek Jump Diffusion model, implementing Euler scheme and assuming a Compound Poisson process as the jump component with normally distributed jump size.

https://doi.org/10.26219/heal.aueb.5795

Chapter 7

7. Empirical Application: FX Market

This chapter conducts extensive empirical analysis of FX market and its applications in vanilla option pricing. Firstly, using the Black-Scholes formulas from section 5 we will calculate the at-the-money (ATM) option prices. Our goal is to calculate the theoretical option price after forecasting 1 step ahead the inputs of the model (volatility, spot exchange rate, foreign and domestic interest rate). In respect of volatility forecasting, we will use GARCH models, presented in chapter 4, and in respect of the spot price forecasting machine learning techniques, such as RNN-LSTM, will be implemented. In the case of the interest rates forecasting, techniques from time-series analysis/forecasting will be used and, more specifically, autoregressive (AR) and autoregressive moving average (ARMA) models. Based on various combinations of these predictions the forecasted call/put prices will be presented. Furthermore, option prices will be calculated based on the implied volatility, with the purpose of comparing the option prices after utilizing as input the realized against the implied volatility.

7.1 The Data

In pricing vanilla options and predicting the theoretical option price, we used the EUR/USD exchange rate and interest rates for the underlying currencies: the three-month (3M) Euribor and the three-month (3M) US Libor. Daily data were used ranging from January 2, 2019, to April 14, 2022, which contained 835 observations in the series. Additionally, we assumed maturity T of the ATM option prices to be 3-months. In the case of volatility, we have already mentioned the distinction between historical and implied volatility. Regarding implied volatility, the source of our data was *Bloomberg* and we extracted the 3M implied volatility. In case of historical volatility, we used the 25-day rolling window annualized standard deviation of the returns series. Figure 7.1.1 shows both the spot price of our dataset and their respective returns, while Figure 7.1.2 presents the historical and implied volatility. Additionally, Figure 7.1.3 demonstrates 3M US Libor interest rate (upper panel) and 3M Euribor rate (lower panel). An interesting point worth mentioning is that

https://doi.org/10.26219/heal.aueb.5795

Euribor rates during this time-period were negative, a policy introduced by ECB on April 21, 2015, with respect to 3M Euribor rate. Traditional pricing models (e.g., Black and Scholes) are based on a lognormal distribution, which does not allow negative rates, while implied volatility tends to infinity when forward rates approach zero.

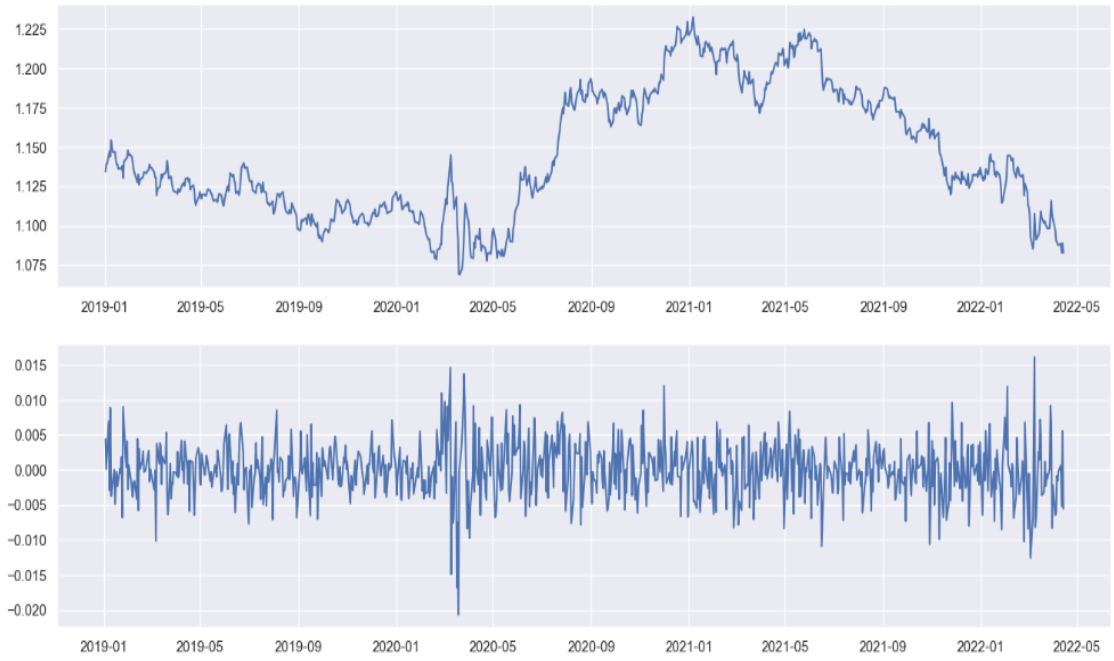


Figure 7.1.1 EUR/USD Spot Exchange rate during January 2019 and April 2022 (upper panel); log-returns of EUR/USD Spot Exchange rate at the same time-period.

Figure 7.1.2 Realized Volatility, as a 25-rolling standard deviation of returns against implied volatility 3M (Source: Bloomberg).

https://doi.org/10.26219/heal.aueb.5795

Figure 7.1.3 Evolution of 3m US Libor Interest rate during January 2019 until April 2022 (upper panel); 3m Euribor interest rate for the same time-period (lower panel).

To sum up, we have presented the input data of the BS-model that we will use in this empirical application. Our spot price would be the closing price of the EUR/USD rate; the strike price would be equal to the spot price since we are dealing with ATM options. The maturity of the option contract would be 3-months (0.25) and our interest rates would be 3M Euribor (foreign) and 3M US Libor (domestic). Finally, a 25-day rolling standard deviation of the returns would be assumed as the historical volatility and the 3M implied volatility, as extracted from Bloomberg.

7.2 FX European Option Pricing

In this part of the analysis, we will price European options with an underlying asset, the EUR/USD exchange rate. The inputs of the BS-formula have already been mentioned above in the data section. At the beginning, we will calculate the call and will put prices, assuming that the volatility is equal to the historical volatility of the returns. After that, we will calculate the call/put prices, assuming that the volatility is equal to the implied volatility and a

https://doi.org/10.26219/heal.aueb.5795

comparison will be made. The first step is to plug the inputs in the BS-formulas (5.8) and (5.9) to price the European call and put options.

$$V_C(T,S) = Se^{-r_f T}N(d_1) - N(d_2)Ke^{-r_d T}$$

$$V_P(T,S) = Ke^{-r_d T}N(d_2) - Se^{-r_f T}N(-d_1)$$

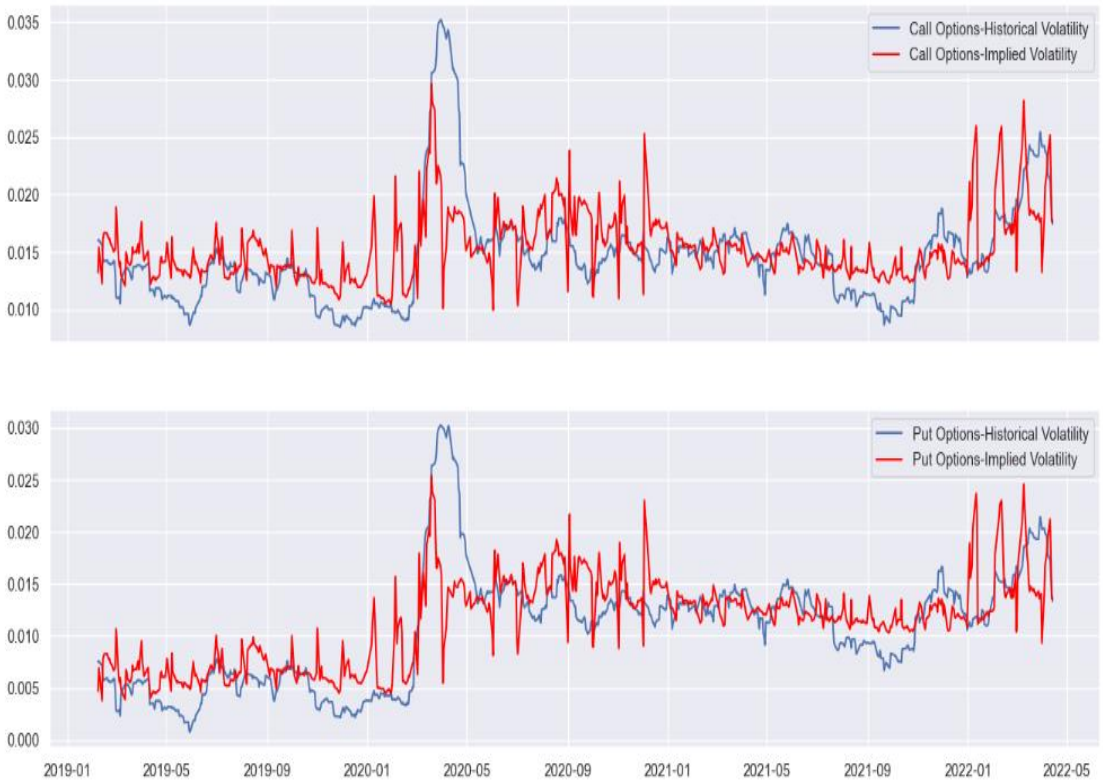
Implementing the above formulas to every data point, a new time series was created, with the calls and puts of FX European options. In Figure 7.2.1 we can observe the evolution of the call and put prices from February 7, 2019 until April 14, 2022.

Figure 7.2.1 ATM 3M Call and Put Options prices under Black-Scholes model, after assuming realized volatility.

Call and put prices have similar progress as the realized volatility (see Figure 7.1.1 above). This is quite reasonable, because it is the only input that is not observable in the market, and it is the most determinant factor in option pricing. In that first case, we assumed that historical volatility is adequate to interpret the fluctuations of the FX market. Moving forward, we will use as our uncertainty measure the implied volatility for the same period and the same option contracts. In Figure 7.2.2 call and put option prices are compared after assuming realized (blue lines) and implied volatility (red lines). In similar context, the option prices are determined based on the volatility we decided to use as an input in BS-formulas. An interesting point is that implied

https://doi.org/10.26219/heal.aueb.5795

volatility is a non-stationary time series, while the historical volatility is a stationary one. Intuitively this means that implied volatility does not depend on the previous quoted implied volatility. Furthermore, in Black-Scholes framework we have constrained ourselves with some assumptions, normality, and constant volatility. Figure 7.2.2 below is evidence that, while Black-Scholes model is a strong tool for quantitative analysts and traders, is not the model that captures the real characteristics of the real-world financial data. To be more precise, we can easily observe that the option prices coming from the BS-model assuming a historical volatility have large deviations from the reality (Call options with implied volatility). Black-Scholes model is a great model to derive the implied volatility, assuming you have the option data, but not a great candidate to calculate the option prices. In Figure 7.2.3, we can observe the probability density function of the implied volatility compared to the normal probability density. It is obvious that the normality assumption of the Black-Scholes model does not adequately capture the behavior of the implied volatility, since the behavior is asymmetric and skewed to the left with fat tails.



https://doi.org/10.26219/heal.aueb.5795

Figure 7.2.2 ATM 3m Call and Put Option prices under Black-Scholes model, after assuming both realized (blue trajectory) and implied volatility (red trajectory).

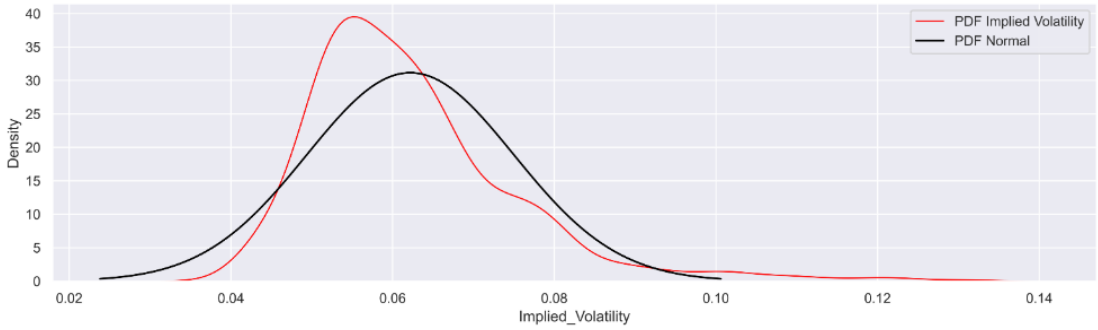


Figure 7.2.3 Comparison of empirical and normal densities for the implied volatility.

In option trading and analysis, Greek exposures or Greek letters provide a useful way to measure the sensitivity of an option's price to different quantifiable variables. Delta, Gamma, Vega, and Theta are the most-knows "Greeks".

(7.1) $$\Delta = \frac{\partial V}{\partial S}$$

where V is the option price and S is the spot rate. Call options always have a positive delta between 0 and 1, while puts always have a negative delta between 0 and -1. In Figure 7.2.3, deltas of call and put options are demonstrated either by using realized or implied volatility. Vega is the rate of change of the value of the option with respect to the volatility of the spot rate.

(7.2) $$Vega = \frac{\partial V}{\partial \sigma}$$

In case that Vega is highly positive/negative, then the option's value is very sensitive to changes in volatility. Figure 7.2.4 depicts Vega for call and put options, while assuming both realized and implied volatility. Since we are assuming Black and Scholes framework "Greek" has a closed form solution as follows:

(7.3.1) $$\Delta_c = e^{-r_f T} N(d_1)$$

(7.3.2) $$\Delta_p = e^{-r_f T}[-N(-d_1)]$$

https://doi.org/10.26219/heal.aueb.5795

where N is the normal cumulative distribution function while d_1 is equal to formula 5.10 above. Vega has the following analytical solution under the Black-Scholes framework:

(7.4) $$V_c = V_p = Se^{r_f T}\sqrt{T}\, N'(d_1)$$

where N' is the probability density function of normal distribution, d_1 equal as above and S is the spot price.

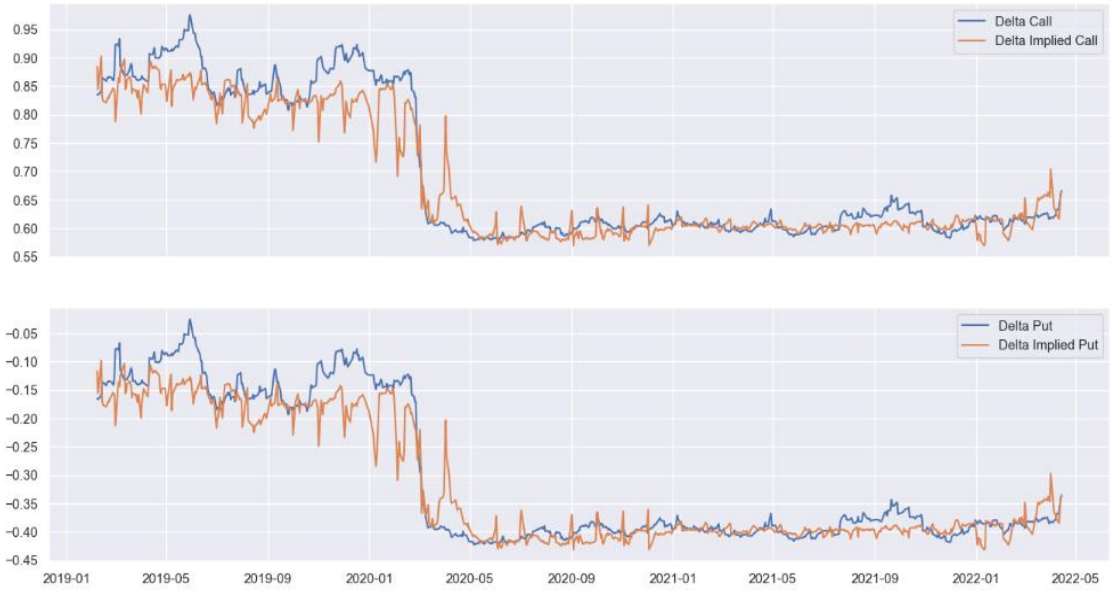


Figure 7.2.4 Evolution of Delta Call and Put after assuming realized (blue trajectory) and implied volatility (orange trajectory).

https://doi.org/10.26219/heal.aueb.5795

Figure 7.2.5 Evolution of Vega Call and Put after assuming realized (blue trajectory) and implied volatility (orange trajectory).

Both Delta and Vega sensitivities, based on realized and implied volatility, seem to have different values in period from 2019 to around April 2020, when both "Greeks" seem to have equal values when realized volatility is used as their input and implied volatility.

7.3 Forecasting Theoretical Option Prices

Moving forward in our analysis, in this section we will attempt to predict the next day value for every input of Black-Scholes formula implementing a variety of methods coming from time-series analysis and machine learning domains. Based on these predictions, the theoretical option prices will be calculated using different variations of the predictions. For example, in respect to volatility, the prediction model will be a GARCH model assuming both normal and t-student distributions. Additionally, a well-known variation of the GARCH, exponential GARCH (E-GARCH) will be presented again using both normal and t-student distribution. Furthermore, we assumed that interest rates could be predicted using either autoregressive models or autoregressive moving average models. Finally, the spot exchange rate was predicted after implementing a LSTM-Recurrent Neural Network.

7.3.1 Volatility Modelling-Forecasting using GARCH models.

Analyzing and predicting volatility is one of the major challenges for someone, who is dealing with financial products and their applications. Various techniques and models have been suggested in the literature, but in the context of our analysis we will be limited to GARCH models and its extension, known as EGARCH models. First, we will work through the return series in this application, since, as we mentioned above, the return series have desirable characteristics. Starting with the first and simplest model, we fitted a GARCH (1,1) model assuming normal distribution. Using the notation from chapter 4 the GARCH (1,1) has the following form:

$$r_t = \mu_t + z_t, z_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \omega + \alpha_1 z_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

https://doi.org/10.26219/heal.aueb.5795

As our estimation method, will use maximum likelihood and our target will be to minimize the negative log-likelihood. Moreover, we assumed that the mean model μ_t is equal to zero and we worked only with the error term of the return series z_t. Briefly, this choice was made since a first try was implemented by assuming first a constant mean and after an AR and ARMA models. In both cases the estimators were not statistically significant and that is why we proceed with the zero-mean assumption. In Table 7.3.1 below, the output of the GARCH (1,1) with normal distributed innovations is presented. Additionally, in Figure 7.3.1, in the upper panel, the standardized residuals of the innovations of the GARCH process are presented, while in the lower panel, we can observe the conditional volatility. Finally, in Figure 7.3.2 the reader may find how the estimated volatility captures the behavior of the volatility of the return series. Overall, the estimated volatility seems to fit the data quite well with some exceptions, especially during the start of the Covid pandemic, where the return series exhibited high volatility behavior with big jumps in their returns. From the output of the GARCH model we have that $a_1 + \beta_1 = 0.9668$, a value very close to one and omega very close to zero. These values are close to the integrated non-stationary bounds on the estimators. The sum of the alpha and beta measures the persistence in the variance and a value close to 1 can be interpreted as the estimated conditional volatility is truly persistent. Additionally, our estimates are statistically significant, assuming that a 5% threshold and that the AIC value of the model is equal to 753.8. AIC will be used in case we wish to compare the model and its predictive ability with other conditional heteroscedastic models or with different error distribution.

https://doi.org/10.26219/heal.aueb.5795

Zero Mean - GARCH Model Results

Dep. Variable:	EURUSD	R-squared:	0.000
Mean Model:	Zero Mean	Adj. R-squared:	0.001
Vol Model:	GARCH	Log-Likelihood:	-373.899
Distribution:	Normal	AIC:	753.799
Method:	Maximum Likelihood	BIC:	768.059
		No. Observations:	857
Date:	Tue, Jul 25 2023	Df Residuals:	857
Time:	13:35:55	Df Model:	0

Volatility Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
omega	4.9609e-03	2.726e-03	1.820	6.882e-02	[-3.827e-04,1.030e-02]
alpha[1]	0.0659	1.858e-02	3.547	3.895e-04	[2.949e-02, 0.102]
beta[1]	0.9009	3.095e-02	29.110	2.644e-186	[0.840, 0.962]

Table 7.3.1 GARCH (1,1) model's output assuming normal errors.

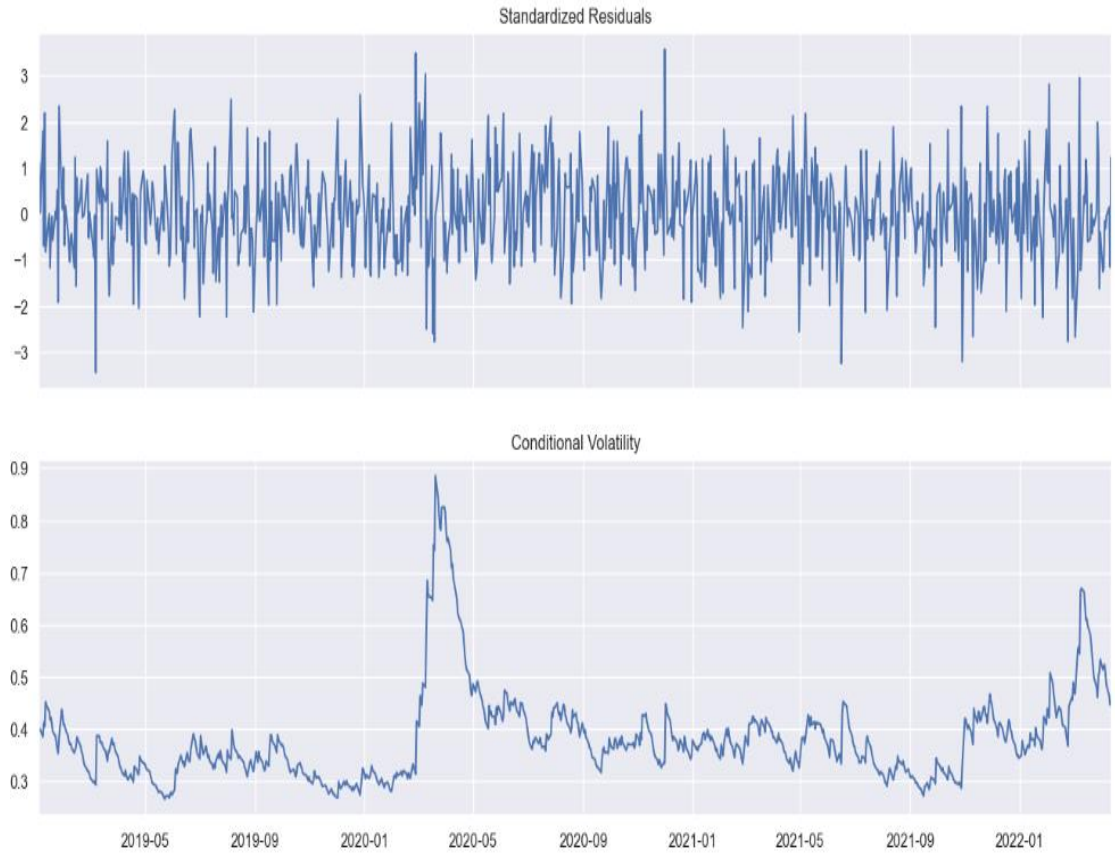


Figure 7.3.1 Standardized residuals of GARCH (1,1) model with normal errors (upper panel); conditional volatility process of GARCH (1,1) model with normal errors.

https://doi.org/10.26219/heal.aueb.5795

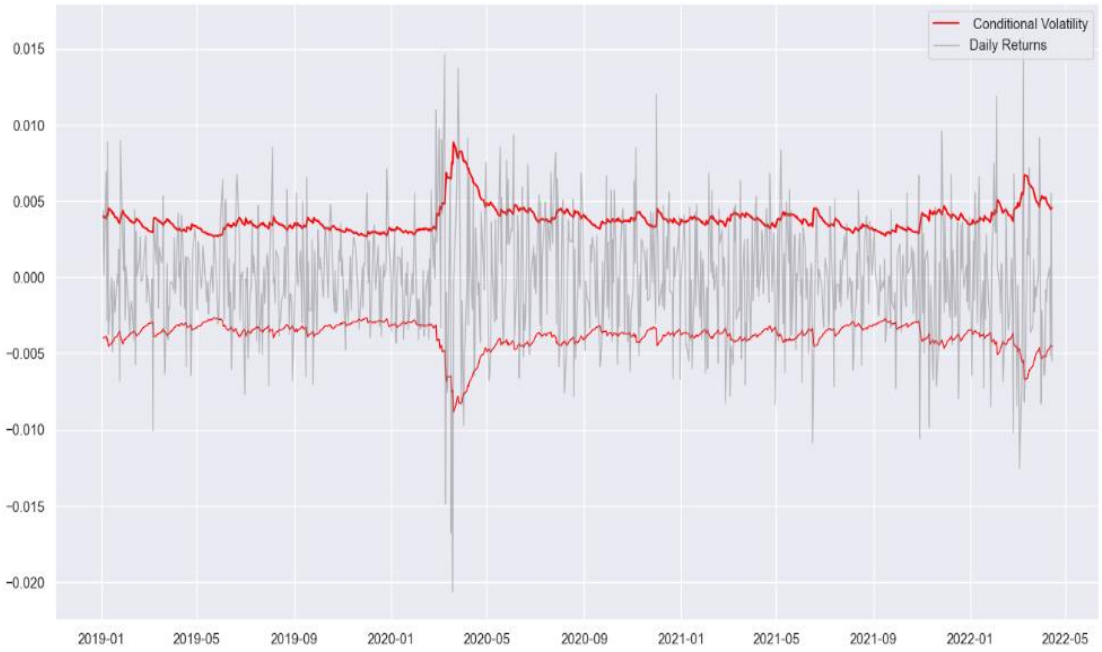


Figure 7.3.2 Conditional volatility estimation against the daily log-returns of EUR/USD spot exchange rate (GARCH (1,1) normal errors).

apart from normal density. In that context, we will fit a GARCH (1,1) model, but in that case, we will select different distribution assumption for the innovations of the model. T-Student distribution is a good candidate for the distribution assumption when it comes to financial log-returns.

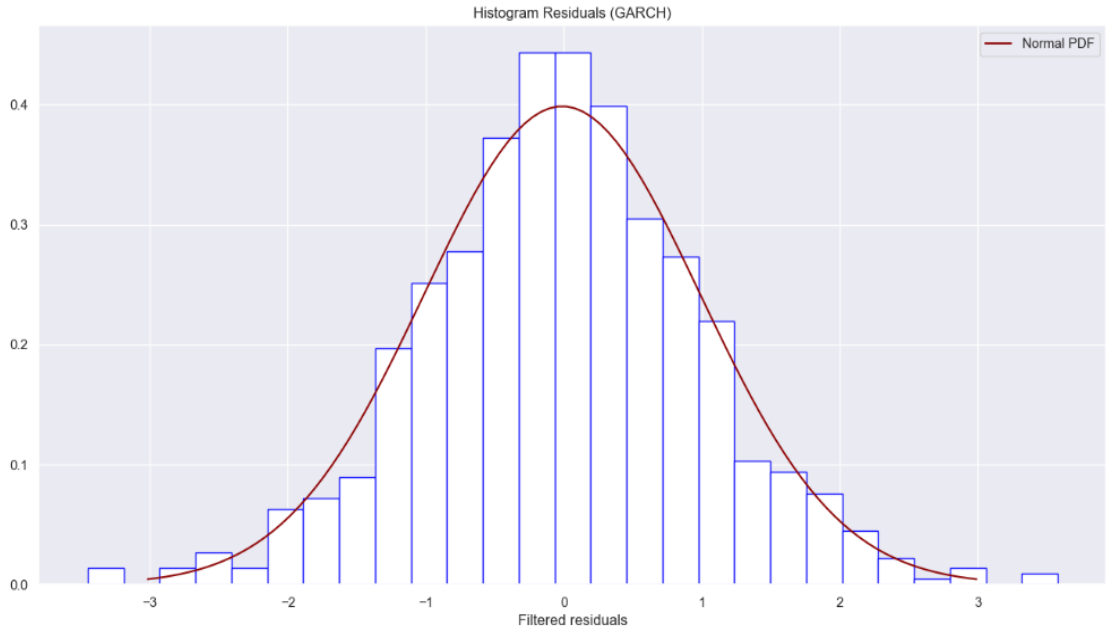


Figure 7.3.3 Histogram of fitted residuals from the GARCH (1,1) model's estimation of EUR/USD log-returns compared to the normal probability density function (red line).

https://doi.org/10.26219/heal.aueb.5795

The second model presented will be a GARCH (1,1), but in this case Student-T distribution will be assumed. A Student's T distribution is symmetric and bell shaped to a normal distribution, but it has fatter tails and this is the reason that makes it a good candidate. Using the above notation, the model is quite similar with the main distinction that $z_t \sim T(\nu)$. The ν parameter indicates the shape of the distribution and larger the ν the more peaked the curve becomes. Table 7.3.2 presents the output of the second suggested model. In the same context as above, we have that $a_1 + \beta_1 = 0.9756$ a value very close to 1 and omega almost very close to zero. These values are again close to the integrated non-stationary bounds on the estimators. In addition, the degrees of freedom ν are approximately equal to 12, meaning that the shape of the distribution has larger peaks compared to the normal distribution. In this case, the estimated parameters are not all statistically significant, assuming a 5% threshold, since the omega produces a p-value equal to 0.109. The AIC value of the model is equal to 747.87 approximately. In Figure 7.3.4, in the upper panel again, they are presented the standardized residuals of the model, while in the lower panel the estimated conditional volatility. Figure 7.3.5 illustrates the returns series and how the estimated volatility fits the data, in the same way as above.

Zero Mean - GARCH Model Results

Dep. Variable:	EURUSD	R-squared:	0.000
Mean Model:	Zero Mean	Adj. R-squared:	0.001
Vol Model:	GARCH	Log-Likelihood:	-369.934
Distribution:	Standardized Student's t	AIC:	747.868
Method:	Maximum Likelihood	BIC:	766.882
		No. Observations:	857
Date:	Tue, Jul 25 2023	Df Residuals:	857
Time:	13:43:41	Df Model:	0

Volatility Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
omega	3.7789e-03	2.358e-03	1.603	0.109	[-8.420e-04,8.400e-03]
alpha[1]	0.0638	1.870e-02	3.411	6.472e-04	[2.713e-02, 0.100]
beta[1]	0.9118	2.892e-02	31.529	3.466e-218	[0.855, 0.969]

Distribution

	coef	std err	t	P>|t|	95.0% Conf. Int.
nu	12.2542	4.349	2.818	4.832e-03	[3.731, 20.777]

Table 7.3.1 GARCH (1,1) model's output assuming T-student errors.

https://doi.org/10.26219/heal.aueb.5795

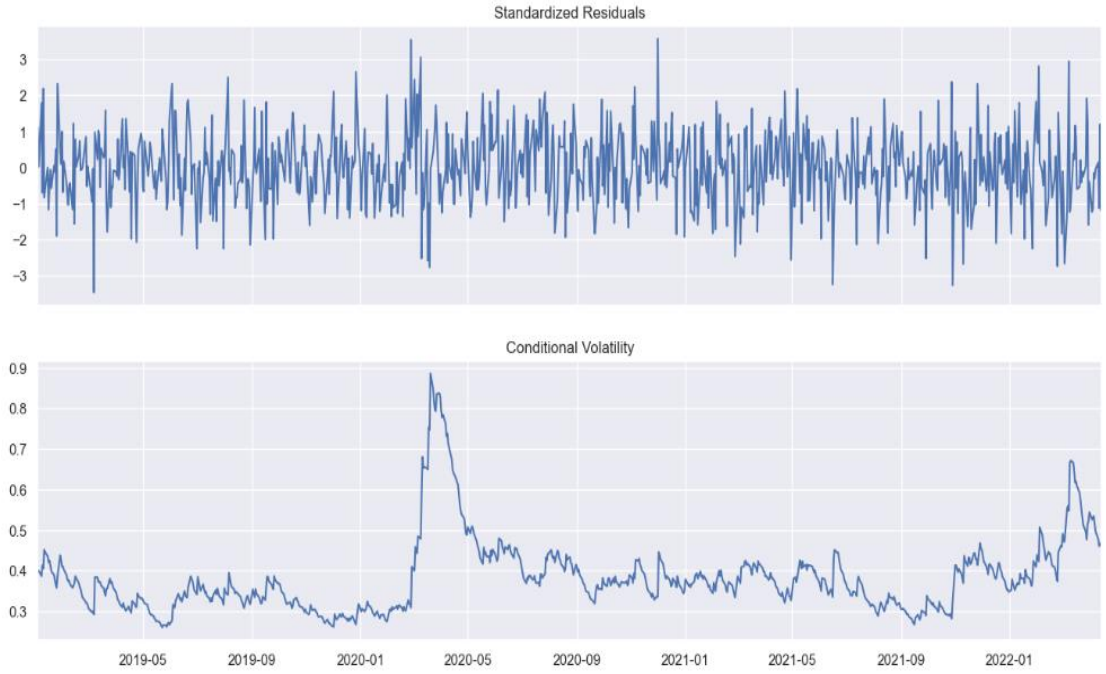


Figure 7.3.4 Standardized residuals of GARCH (1,1) model with T-student errors (upper panel); conditional volatility process of GARCH (1,1) model with T-student errors.

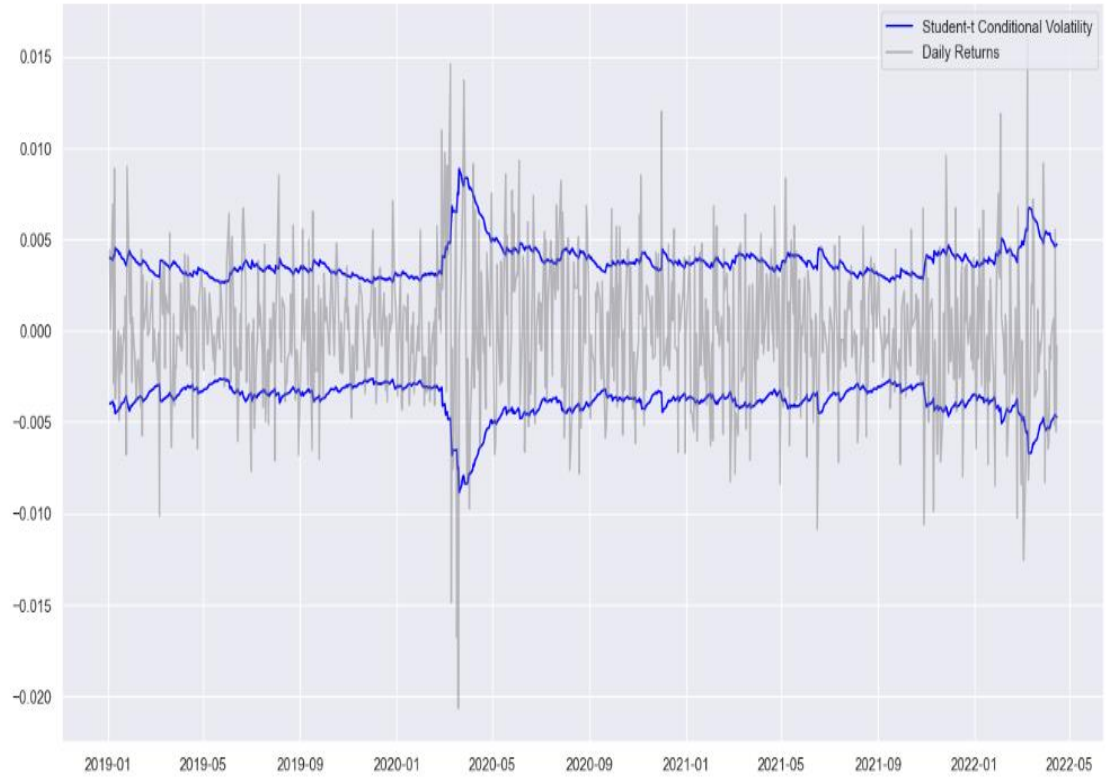


Figure 7.3.5 Conditional volatility estimation against the daily log-returns of EUR/USD spot exchange rate (GARCH (1,1) T-Student errors).

https://doi.org/10.26219/heal.aueb.5795

Based on these GARCH models, we will forecast 1-step ahead the variance of the series and through variance we will calculate the annualized volatility. For starters, the annualized volatility is given from the below formula:

$$\sigma_Y = \sqrt{252} * \sqrt{\sigma_d^2}$$

For models such as GARCH, analytical forecasts are available based on the relationship $\mathbb{E}[z_{t+1}^2] = \sigma_{t+1}^2$, which recursively compute forecasts. In terms of conditional variance, the forecast is built as follows:

$$\sigma_{t+1}^2 = \omega + \alpha_1 z_t^2 + \beta_1 \sigma_t^2$$

$$\sigma_{t+h}^2 = \omega + \alpha_1 \mathbb{E}_t[z_{t+h-1}^2] + \beta_1 \mathbb{E}_t[\sigma_{t+h-1}^2]$$

$$\sigma_{t+h}^2 = \omega + (\alpha_1 + \beta_1)\mathbb{E}_t[\sigma_{t+h-1}^2]$$

Based on the above formulas, the forecast of the GARCH (1,1), $z_t \sim N(0,1)$ and GARCH (1,1), $z_t \sim T(\nu)$ are given in the Table 7.3.3 below:

1-step Ahead Forecast σ_Y	**GARCH (1,1) - Normal**	**GARCH (1,1) – Student T**
	0.070223	0.072302

Table 7.3.3 1-day ahead annualized volatility forecast of GARCH models.

Moving forward, in the volatility modelling we will implement the same analysis as above, but assuming that the variance process is explained by an EGARCH model both for normal and Student-T innovations. In both following models again, we will work only with the residuals assuming a zero-mean equation. As we have shown above in Chapter 4 the EGARCH model has the following notation:

$$z_t = \sigma_t \varepsilon_t$$

$$\ln(\sigma_t^2) = \omega + \beta_1 \ln(\sigma_{t-1}^2) + \theta_1 z_{t-1} + \alpha_1(|z_{t-1}| - \mathbb{E}|z_{t-1}|)$$

In the Appendix, there are available the outputs of the models along with the figures of standardized residuals and the estimated conditional volatility of the EGARCH models, in the same way presented for the GARCH models

https://doi.org/10.26219/heal.aueb.5795

above. As we mentioned above, the EGARCH model is an extension of GARCH with the aim to include in the analysis of volatility the leverage effect. The leverage effect in the EGARCH output is based on the estimation of the γ_1 parameter. This notation is equivalent to the θ_1 of our notation in EGARCH model. In EGARCH model, there is a leverage effect if γ_1 estimator that is less than zero. The condition for volatility persistency is the β_1 coefficient close to 1. Starting with the EGARCH (1,1) and having errors normally distributed, the β_1 coefficient is equal to 0.9620, which is close to 1, meaning that the estimated volatility is persistent. The γ_1 estimator is equal to 0.0152, which is a positive value, meaning that, according to the model, there is no leverage effect. In addition, all estimators are statistically significant, except from the γ_1 coefficient. As a result, the extra complexity we added in the GARCH model did not enhance the model and its estimators. Moreover, the EGARCH (1,1) with normal innovations had an AIC value equal to 762.306, which is the largest among the three models so far.

Finally, the last model implemented, with respect to volatility modeling, would be an EGARCH (1,1) model with Student-T errors. Based on the output, on the Appendix section the β_1 coefficient is equal to 0.9690, which is close to 1, meaning that the estimated volatility is persistent in this model as well. The γ_1 estimator is equal to 0.0176, which is a positive value, meaning that, according to the model. there is no leverage effect once again. Additionally, the ν coefficient, which controls the shape of the student-T distribution, is equal to approximately 11. All estimators are statistically significant, except from the γ_1 coefficient. The AIC value for this model is equal to 754.56. Slightly better than the previous EGARCH model, but in general, the addition of the leverage effect component did not add any value to our models. In table 7.3.4, the reader may find the forecasts produced by the EGARCH models, the forecasts of the GARCH models along with the AIC values of the respective models.

	1-step Ahead Forecast σ_Y	**AIC Value**
GARCH (1,1) -	0.070223	753.799

https://doi.org/10.26219/heal.aueb.5795

Normal		
GARCH (1,1) – Student T	0.072302	747.868
EGARCH (1,1) - Normal	0.067745	762.306
EGARCH (1,1) – Student T	0.068877	754.562

Table 7.3.2 1-day ahead annualized volatility forecast and AIC value of GARCH and EGARCH models.

For illustration purposes and only, we will use all the forecasts as input in the BS-formula for the volatility to produce a variety of call/put prices. Nevertheless, the forecast value of the GARCH (1,1) with Student-T distribution assumption seems to be the better choice based on the AIC value. Additionally, the implied volatility for the forecasted date is equal approximately to 7.5%. Of course, there is deviance from the implied volatility for that specific date, but the model achieved the smallest deviance among the other models implemented.

7.3.2 Interest Rate Prediction-Time Series Forecasting Approach

A vital part of calculating the option prices based on BS framework is to first choose the interest rates correctly and in terms of calculating the theoretical option prices 1-step-ahead is to select models that interpret the movements of the rates in the most efficient way. The models implemented in this section would be ARIMA (p,d,q) models. These models require stationary series. The ACF and PACF plots in Figure 7.3.6 identify both interest rates series as non-stationary. Additionally, it is obvious that there is a linear trend of slowly decreasing to low values with the improvement of lags. To remove this trend and make the series stationary, we will take the first difference (d=1), creating the returns series of the interest rates. In the case of Euribor, because of the presence of negative values, the arithmetic returns have been selected, while in the case of Libor have been selected the log-returns. Figure 7.3.7 demonstrates the ACF and PACF plots of the rates' returns. The trend has been removed successfully and the plots now present the remaining correlations between the data points. From the ACF plot, we determined the

https://doi.org/10.26219/heal.aueb.5795

MA (q) process order and from the PACF plot, we determined the AR (p) process order. The term 'd' was calculated by differentiating the trend and would be equal to 1. To determine the optimal hyperparameters for the models, we used the grid search framework which, to find the best model, it optimizes for a given information criterion. In this analysis, we select as our information criterion the AIC. The grid search returns the ARIMA model, which minimizes the AIC value. An important notice here is that we decided to set maximum p and q order equal to 5 in order not to result in very complex models. In this section, the simple ARIMA approach will be presented, meaning that only with the Euribor/Libor US interest rate, with no covariates. Finally, apart from the minimum AIC value, we will consider the statistical significance of the AR and MA components of each model and the best 2 candidates' models for each interest rate will be presented.

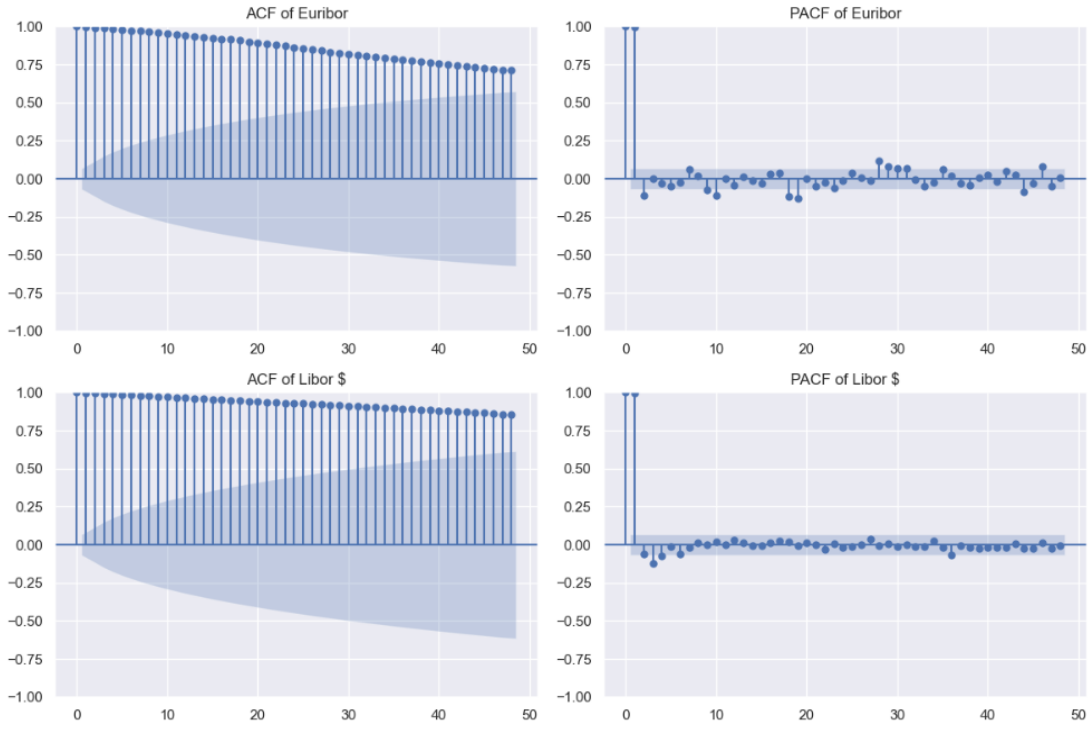


Figure 7.3.6 Autocorrelation function and partial autocorrelation function plots of 3m Euribor (upper panels) and 3m US Libor (lower panels) interest rates.

https://doi.org/10.26219/heal.aueb.5795

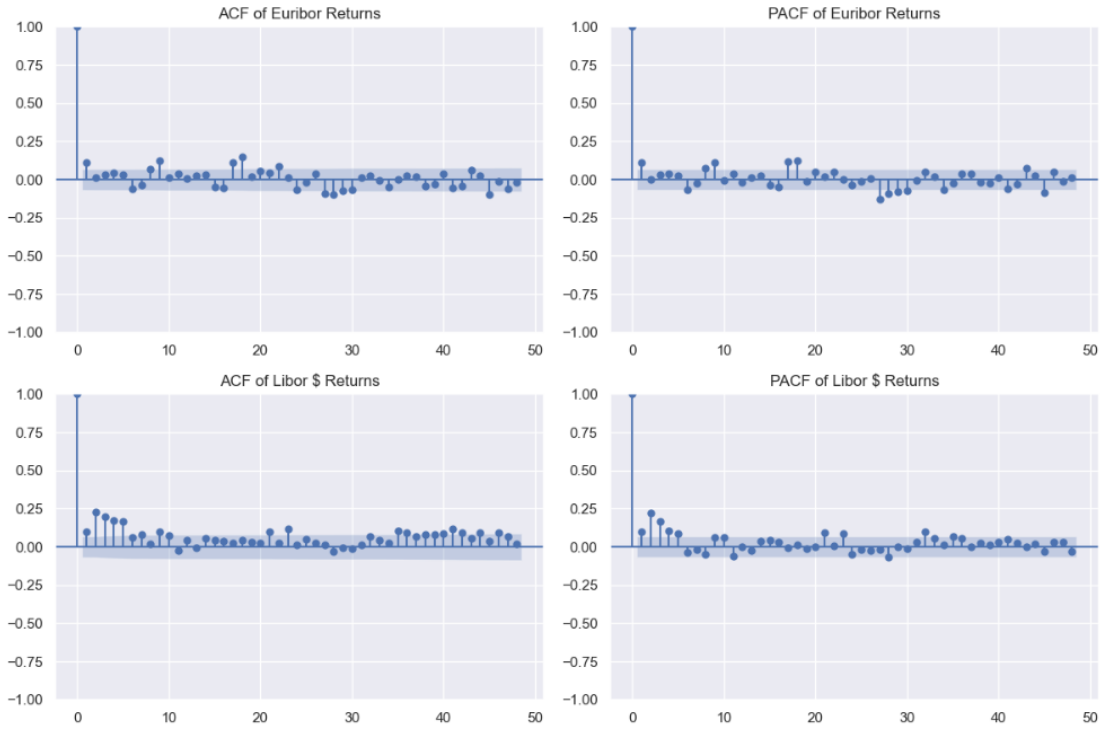


Figure 7.3.7 Autocorrelation function and partial autocorrelation function plots after first order linear differentiation for 3m Euribor (upper panels) and first order log differentiation for 3m US Libor (lower panels).

Starting with the 3M Euribor interest rate, the automated grid search process suggested an ARIMA (1,1,0) model, achieving -13618.09 AIC value. This result is, also, in line with the PACF plot of Euribor returns in Figure 7.3.7, since it cuts off after lag 1. However, there is an indication from ACF plot that there should be included and a MA component in the model. In that context, we fitted as a 2nd candidate model an ARIMA (1,1,2), which achieved a -13616.199 AIC value. In both models, all the ARIMA coefficients are statistically significant assuming a 5% threshold. In Table 7.3.5 the reader may find the output of the first model ARIMA (1,1,0), basically a simple AR (1) on the returns of the 3M Euribor. Intuitively, an AR (1) model may be interpreted as the return is only related to the previous days' return. In the lowest part of the table, various residuals tests are presented. Ljung-Box test has a p-value equal to .98, thus we fail to reject the null hypothesis of independently distributed data. There is no significant remaining correlation in the residuals of the model. This means that the model is adequate to capture the correlation information in our times series. On the other hand, we

https://doi.org/10.26219/heal.aueb.5795

can reject the normality assumption, through the Jarque-Bera test (*0.00*) and the heteroscedasticity assumption of the residuals (*0.00*).

Dep. Variable:	Euribor	**No. Observations:**	835
Model:	ARIMA(1, 1, 0)	**Log Likelihood**	6811.045
Date:	Thu, 27 Jul 2023	**AIC**	-13618.090
Time:	20:00:26	**BIC**	-13608.638
Sample:	0	**HQIC**	-13614.466
	- 835		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.1097	3.59e-20	3.06e+18	0.000	0.110	0.110
sigma2	4.555e-09	8.51e-11	53.499	0.000	4.39e-09	4.72e-09

Ljung-Box (L1) (Q):	0.00	**Jarque-Bera (JB):**	7828.64
Prob(Q):	0.98	**Prob(JB):**	0.00
Heteroskedasticity (H):	1.67	**Skew:**	1.33
Prob(H) (two-sided):	0.00	**Kurtosis:**	17.77

Table 7.3.5 ARIMA (1,1,0) model's output of 3m Euribor Interest Rate.

In Appendix, Table 7.3.13 depicts the output of the second model regarding 3M Euribor interest rate, an ARIMA (1,1,2) model. In the same context, Ljung-Box test indicates that the model was successful in respect of removing the correlation between the residuals, but the non-normality and the heteroscedasticity issues are again present. The same procedure will be implemented regarding the US 3M Libor interest rate. The automated grid search suggested an ARIMA (4,1,0) model, achieving -11872.354 AIC. Considering the PACF and ACF plots, we were excepting order p around 4 or 5, since after the 5th spike the correlation cuts off. However, it is unexpected that there is no MA component since based on the ACF there are significant spikes. Table 7.3.6 demonstrates the output of the proposed model. Once again, all the AR-coefficients are statistically significant and. based on statistical tests. we observed that there is no serial correlation (*Ljung-Box*), there is non-normality (*Jarque-Bera*) and heteroscedasticity. The second candidate model is an ARIMA (4,1,1) model. This model achieved an AIC equal to -11870.377, slightly worse performance than the first model. Adding an extra parameter in the model increases the AIC, since it penalizes this

https://doi.org/10.26219/heal.aueb.5795

extra parameter. In Appendix, Table 7.3.14 shows the output of the ARIMA (4,1,1) model.

Dep. Variable:	Libor $	**No. Observations:**	835
Model:	ARIMA(4, 1, 0)	**Log Likelihood**	5941.177
Date:	Fri, 28 Jul 2023	**AIC**	-11872.354
Time:	17:13:46	**BIC**	-11848.723
Sample:	0	**HQIC**	-11863.294
	- 835		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.0399	0.009	4.490	0.000	0.022	0.057
ar.L2	0.4417	0.007	65.230	0.000	0.428	0.455
ar.L3	0.2562	0.004	62.894	0.000	0.248	0.264
ar.L4	-0.0939	0.002	-46.335	0.000	-0.098	-0.090
sigma2	3.788e-08	5.62e-10	67.393	0.000	3.68e-08	3.9e-08

Ljung-Box (L1) (Q):	0.01	**Jarque-Bera (JB):**	30497.87
Prob(Q):	0.92	**Prob(JB):**	0.00
Heteroskedasticity (H):	0.75	**Skew:**	-1.63
Prob(H) (two-sided):	0.02	**Kurtosis:**	32.45

Table 7.3.3 ARIMA (4,1,0) model's output for 3m US Libor Interest Rate.

After defining the models, the next step, which is the main purpose of this analysis, is to proceed to the 1-step ahead forecast. First, we will split the data (80/20) into a train sub-sample and a test sub-sample. The train dataset will be used to run the models and the test data will be used to evaluate the performance of the techniques we used based on an evaluation metric. The metrics used in our analysis would be the Mean Squared Error (MSE) and Mean Absolute error. MSE represents the squared difference between actual and predicted values and is calculated as follows:

$$MSE = \frac{1}{n}\sum(y - \hat{y})^2$$

MSE is a well-known and most used evaluation metric in statistics and machine learning. MAE measures the average absolute difference between the predicted values and the actual values and is used to assess the effectiveness of a model. MAE is based on the following formula:

https://doi.org/10.26219/heal.aueb.5795

$$MAE = \frac{1}{n}\sum |y - \hat{y}|$$

After running the models and evaluating them, we will proceed to the 1-step ahead out-of-sample forecast for both interest rates. Table 7.3.7 below demonstrates the evaluation metrics of the models along with their respective AIC value. The blue color indicates the better model based on each evaluation metric between the two models for each interest rate. According to all the metrics a simple AR (1) to the returns of the Euribor is adequate to capture the behavior of the interest rate and forecast with high accuracy. In respect of Libor, both MSE and MAE indicate that ARIMA (4,1,1) model has better forecasting accuracy, but based on AIC an AR (4) model to the returns of the Libor is a better model. Moreover, Figures 7.3.8 and 7.3.9 illustrate the observed values and the predicted values in test dataset for both models and interest rates.

	MSE	MAE	AIC Value
ARIMA (1,1,0) - Euribor	0.0000000039008	0.0000448205	-13618.09
ARIMA (1,1,2) – Euribor	0.0000000039208	0.0000448485	-13616.199
ARIMA (4,1,0) - Libor	0.000000024281	0.0000848506	-11872.354
ARIMA (4,1,1) – Libor	0.000000024082	0.0000843141	-11870.377

Table 4.3.7 Comparison of evaluation metrics (MSE, MAE and AIC) of implemented time-series models (blue color indicates the lowest value).

As a final thought, Euribor is explained with a simple model without much complexity indicating that looking 1-day back in the data is adequate to explain the behavior of the interest rate. On the other hand, Libor seems to

https://doi.org/10.26219/heal.aueb.5795

need a more complicated model with strongest relations with the past data, since the model indicates relations between the returns of Libor beyond the previous day but further in the past. Additionally, Libor model seem to fit the data smoothly without significant fluctuations, even after the constant increase of the rates after February 2022. Euribor, on the contrary, seems to be a more challenging case, because there is higher variability and more 'jumps' in its behavior.

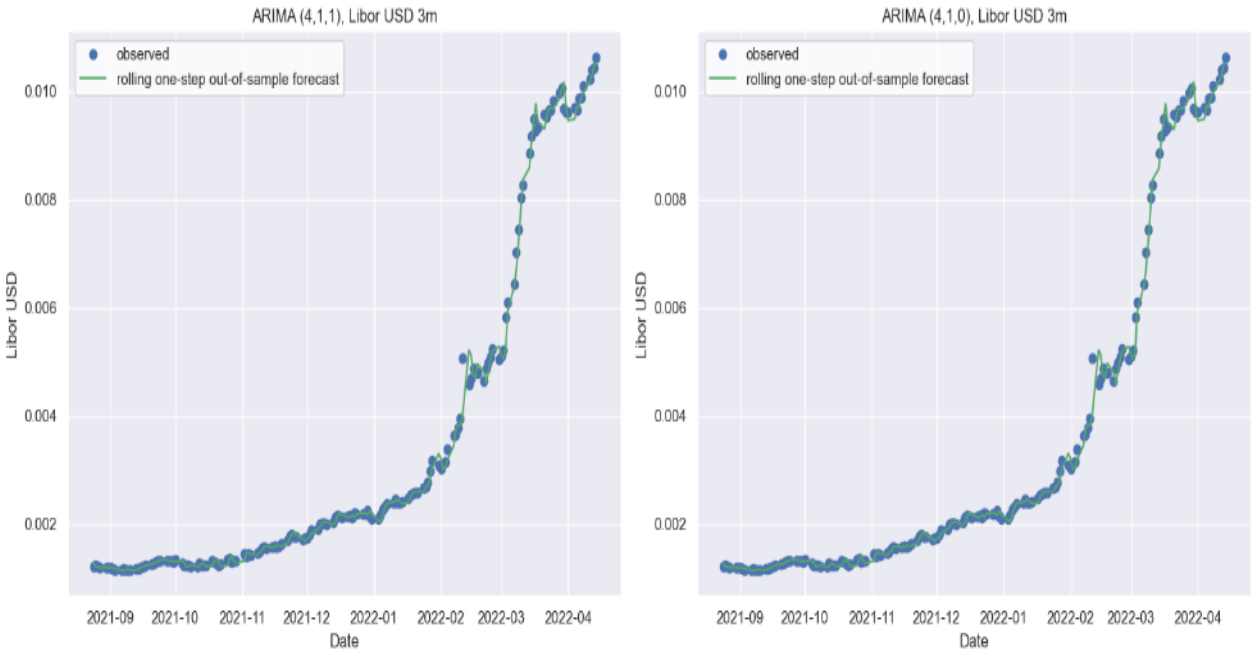


Figure 7.3.8 Observed and predicted values of ARIMA (4,1,1) and ARIMA (4,1,0) of 3m US Libor Interest Rate.

https://doi.org/10.26219/heal.aueb.5795

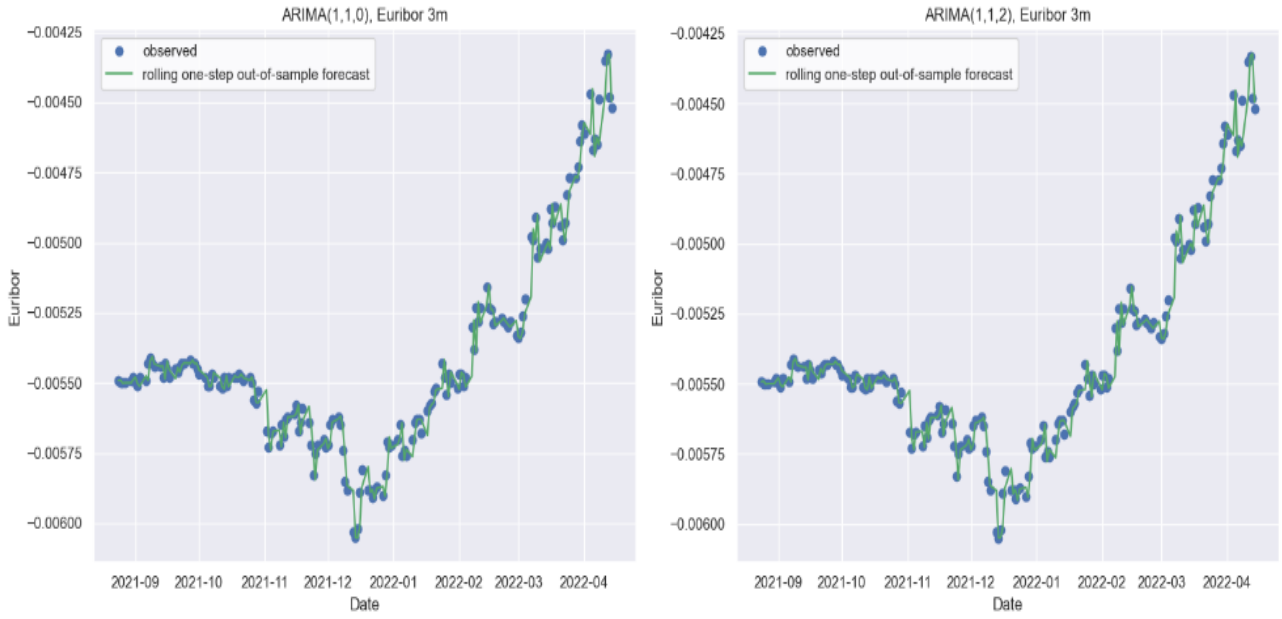


Figure 7.3.9 Observed and predicted values of ARIMA (1,1,0) and ARIMA (1,1,2) of 3m Euribor Interest Rate.

The final step of the analysis of the interest rates is to produce the 1-day out-of-sample forecast based on the implemented models. We test and evaluate the models having as aim of this procedure the prediction of the next-day theoretical interest rates. The predictions will be presented along with the lower and upper limit of a 95% confidence interval. Table 7.3.8 sums up the forecasts for each model along with the respective confidence intervals. In the parenthesis it is exhibited the standard error of the respective forecast. An interesting part is that both ARIMA models with respect to Libor have the same forecast, with slightly altered lower and upper bounds in their confidence intervals.

	1d Forecast	Lower CI	Upper CI
ARIMA (1,1,0) -	-0.004524	-0.04658	-0.004391

https://doi.org/10.26219/heal.aueb.5795

Euribor	(0.000068)		
ARIMA (1,1,2) – Euribor	-0.004521 (0.000068)	-0.004654	-0.004387
ARIMA (4,1,0) - Libor	0.010694 (0.000195)	0.010312	0.011075
ARIMA (4,1,1) – Libor	0.010694 (0.000195)	0.010313	0.011076

Table 7.3.5 1-day Interest Rates Forecasts of best models along with 95% confidence intervals.

In that way, we concluded the prediction of the inputs of the BS-formula. In the next short section, we will combine all the forecasts and present results based on the predicted models implemented in these last three sections.

7.3.3 Spot Exchange Rate Prediction-Machine Learning Approach

In this section we will describe the aspect of the LSTM-RNN specification. The model was generated in Python using among others Keras API in Tensoflow, which is an open-source library for machine learning (*Purkait 2019*). As a first step in LSTM-RNN specification, a min-max scaler should be used to normalize the input values. In that way, all values will be transformed within the range [0,1]. After forecasting results, the data would be rescaled using the inverse procedure. The min-max normalization is as follows:

$$y_{norm} = \frac{y_i - \min(y)}{\max(y) - \min(y)}$$

while the inverse transformation used is the following:

$$y_{scaled} = \hat{y}_i(\max(y) - \min(y)) + \min(y)$$

Moving forward, another preprocessing step is necessary for the implementation of time-series forecasting using LSTM-RNN. We used the sliding window method to target look-back steps of normalized data. This approach reconstructs time-series datasets into supervised learning problems by using the previous time step values to predict the next time step value. Additionally, in this setup we split the dataset into train and test sets,

https://doi.org/10.26219/heal.aueb.5795

assuming an 80/20 split and we select a window size equal to 21 (21 trading days is the standard 1-month period in financial applications). Table 7.3.9, below, demonstrates the hyperparameters used for training the LSTM model. Our simple LSTM model consists of an LSTM layer and a dense layer. The LSTM layer performs the calculations, shown in equations 4.2.3 to 4.2.7 above, and the dense layer performs the simple linear combination equations, shown in 4.2.1 and 4.2.2 ones. We selected 200 neurons in respect of LSTM layer and 1 in dense layer. When training a LSTM model, it is crucial to select a loss measure and to minimize the loss. In this application, we select MSE as our loss function, which is widely used in time series forecasting (*Makridakis and Hibon, 1991*). The next hyperparameters are the number of epochs and batch size. A LSTM RNN is trained multiple times with a batch size equal to 32. To minimize the training loss of the model, 200 epochs are used. Finally, an optimizer is selected to optimize the loss function. In our application Adaptive Moment Estimation (Adam) is selected. Adam optimizer involves a combination of two gradient descent methodologies. This algorithm considers the 'exponential weighted average' of the gradients and uses averages. Using averages makes the algorithm converge at a faster pace.

Hyperparameters	**Values**
Internal Units (Neurons)	200,1
Epochs	200
Optimizer	Adam
Loss Function	MSE
Batch Size	32

Table 7.3.6 Hyperparameters used for training the LSTM model.

After fitting the model and rescaling the dataset, we present the results of the test dataset in the next Figure 7.3.10, where the actual data are represented by the green line and the predicted values by the red line. In general terms, our model seems to fit well with the data having small deviances from the real values. The model also achieved MSE equal to $2.657e^{-05}$. Moreover, through

https://doi.org/10.26219/heal.aueb.5795

the optimization procedure, explained above, the model achieved a MAE equal to 0.00389. The final part would be to investigate the test data set and, based on this result, to produce a forecasted value for the next day ahead. After reshaping the data based on the sliding window method and taking the inverse transform, as we explained above, the forecasted value of our LSTM-RNN is equal to 1.08259.

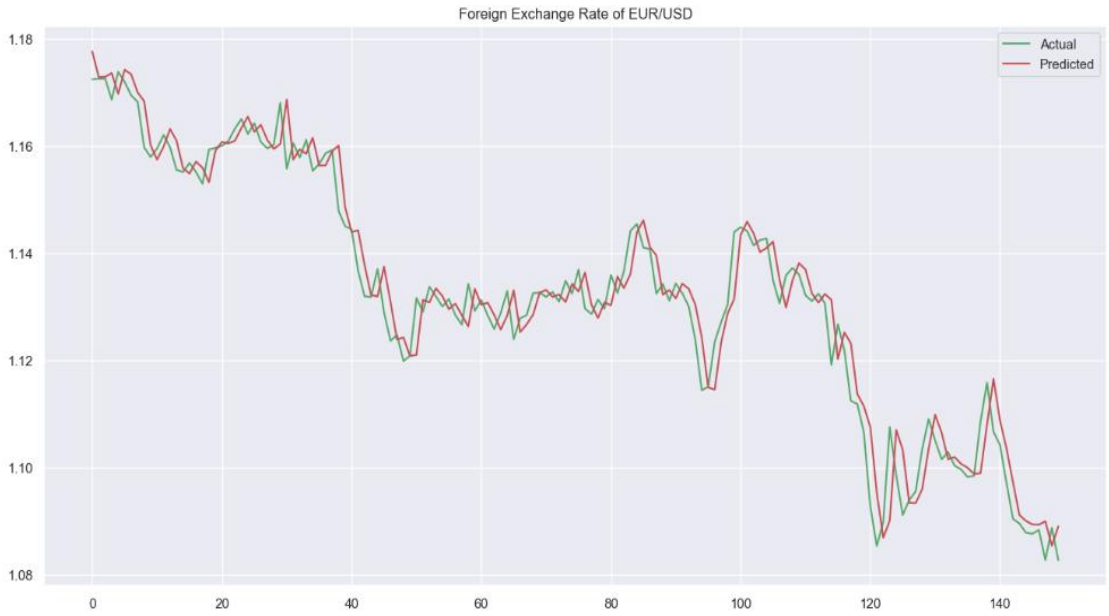


Figure 7.3.10 Actual against predicted values of EUR/USD Exchange rate of LSTM model.

The Foreign exchange rate of EUR/USD according to our model would be 1.08259. Based on the real data available in different platforms, the respective real value of the EUR/USD on April 15, 2022, is 1.0821, which is quite near to our predicted price. Overall, the model can be considered as a successful attempt to capture the behavior of the EUR/USD exchange rate. The model achieved a MAE of 0.00389, which is far larger than the deviation we have in our predicted exchange rate in comparison to the one observed in the market. In this application, we build a simple LSTM neural network, which can be improved in many ways. For instance, we could add more hidden layers for starters and, in addition, to increase the neurons or the epochs of our algorithm. However, these improvements would have a cost computationally wise.

https://doi.org/10.26219/heal.aueb.5795

The final part is to piece all together the predictions made in this analysis and, based on these predictions, to calculate the theoretical option prices of an ATM 3M call and put option. To sum up, the proposed models from the overall analysis suggest that we should use in respect of volatility a GARCH (1,1) model, assuming T-student innovations, in respect of interest rates ARIMA (1,1,0) for Euribor and ARIMA (4,1,1) for Libor and a LSTM-RNN with the aforementioned metrics and hyperparameters. For illustration and benchmarking purposes, we will present the theoretical call/put options in the first case assuming a GARCH (1,1) T-student model and the best time-series models in respect of interest rates. In the second case, we will assume the second-best model in respect of volatility, a GARCH (1,1) normal model, and the second-best models in respect of interest rates, as we have shown in section 7.3.2 above. These combinations are presented in Table 7.3.10 below. As a first comment, the Option prices, both calls and puts, which assumed a GARCH with t-student innovations, have higher theoretical prices, than the ones that assumed a GARCH model with normal innovations. In this case, again, it is shown that the major factor that affects the option prices is volatility. If we compare the theoretical call/put prices with the ones with same volatility modelling approach, but with different interest rate modelling approach, we observe that there is no deviation or a slight deviation in the case of the put option price. This observation boosts the argument mentioned before that the volatility modelling approach, along with the assumptions about it, is crucial when it comes to option pricing and it is the most influential input of the Black-Scholes model.

https://doi.org/10.26219/heal.aueb.5795

Models' Assumptions	Call Option Price (ATM)	Put Option Price (ATM)
GARCH (1,1) Normal; ARIMA (1,1,0); ARIMA (4,1,0)	0.01654	0.012425
GARCH (1,1) T-Student; ARIMA (1,1,0); ARIMA (4,1,0)	0.017006	0.012891
GARCH (1,1) Normal; ARIMA (1,1,2); ARIMA (4,1,1)	0.01654	0.012426
GARCH (1,1) T-Student; ARIMA (1,1,2); ARIMA (4,1,1)	0.017006	0.012891

Table 7.3.7 Call and Put Options' Prices under different Models' assumptions (GARCH and ARIMA variations).

7.4.1 Merton Jump Diffusion Model against Black-Scholes Model

In this final section of FX option pricing application, we will illustrate the call option prices, both calculated assuming realized and implied volatility and both assuming standard Black-Scholes model and Merton Jump Diffusion model. The Aim of this analysis is to compare the results of these different approaches and to comment any significant results after adding an extra component of randomness in the well-known BS formula. The approach followed on this application is based on the closed form solution within the jump-diffusion model in terms of Black-Scholes prices. Let's denote $V_{BS}(S, K, \sigma, r_f, r_d, T)$ as the BS call price. The corresponding price within the jump diffusion framework can be written as follows:

https://doi.org/10.26219/heal.aueb.5795

$$V_{MJD}(S,K,\sigma,r_f,r_d,T,\lambda,\mu_j,\sigma_j) = \sum_{k=0}^{\infty} \frac{e^{(\lambda T e^{\mu_j+\frac{1}{2}\sigma_j^2})}(\lambda T e^{\mu_j+\frac{1}{2}\sigma_j^2})^k}{k!} V_{BS}(S,K,\sigma_k,r_k,T)$$

Where $\sigma_k = \sqrt{\sigma^2 + k\sigma_j{}^2/T}$ and $r_k = r_d - r_f - \lambda\left(e^{\mu_j+\frac{1}{2}\sigma_j^2} - 1\right) + \frac{k(\mu_j+\frac{1}{2}\sigma_j^2)}{T}$. This approach is based on Merton's derivation using a PDE approach. The reader may consult *Joshi 2003* for further details in the above formulation of European option prices using MJD. The summation, above, assumes that the BS price conditional on knowing exactly how many jumps will occur and weights these values by their corresponding probability under the Poisson distribution. In the Python code, we will truncate this by using a large k number of jumps. Technically speaking we have a semi-closed form solution involving an infinite series. In practice, we will cut off after a finite number of terms. It can be shown that the price of a vanilla call is a convex function of spot in the Black-Scholes world. Now, in the case of the presence of jumps, the price always increases. This increase intuitively can be interpreted as compensation for the extra risk taken due to the presence of jumps. For more details regarding this matter the reader may address to *Joshi, 2003 section 15.5*.

After the short formulation and introduction of MJD, we will proceed to the empirical application in the same manner as Section 7.2. The distinction between realized and implied volatility has already been covered extensively in section 7.2. In the same context, the FX European option prices were calculated from early February 2019 (02/07/2019) to mid-April 2022 (14/04/2022), assuming, in the first case, realized volatility as the 25-rolling standard deviation of the returns of the spot Eurodollar and implied volatility in the second case. While in section 7.2 our goal was to compare the option prices after assuming these different volatility measures, in this section our goal is to compare the option prices after implementing Black-Scholes standard model and MJD model, assuming both volatility measures. First, we must set the new model parameters for the jump part. In this analysis, we assumed the mean of the jump to be equal to 0.01 and the volatility parameter of the jump to be equal to 0.03. The intensity of the jump, meaning the

https://doi.org/10.26219/heal.aueb.5795

number of the jumps per annum, was set to be equal to 1. This set of parameters normally should be calibrated with real option data. In this thesis, we select these parameters intuitively with the purpose of confirming and illustrating the fact that the option prices should be increased due to the presence of the jump component. After setting the parameters and assuming the same inputs of the BS-model as in section 7.2, we will calculate the MJD option prices with both realized and implied volatility. Figures 7.4.1 and 7.4.2 demonstrate the call and put option prices after implementing typical Black-Scholes framework and MJD framework with realized volatility and with implied volatility respectively. In both cases, as expected, option prices under MJD model have been priced higher than the ones under the BS model. An interesting point is that during the start of the covid crisis around early 2020, we observe that the option prices under both models have the lowest deviance across the time-period we examined. Of course, this is also due to the selection of the specific parameters in this particular experimental setup. In the case of selecting higher mean jump and volatility jump parameters, the deviance would be greater.

Figure 7.4.1 Call and Put Option assuming realized volatility under Black-Scholes and MJD models (blue line: Black-Scholes, red line: Merton Jump Diffusion).

https://doi.org/10.26219/heal.aueb.5795

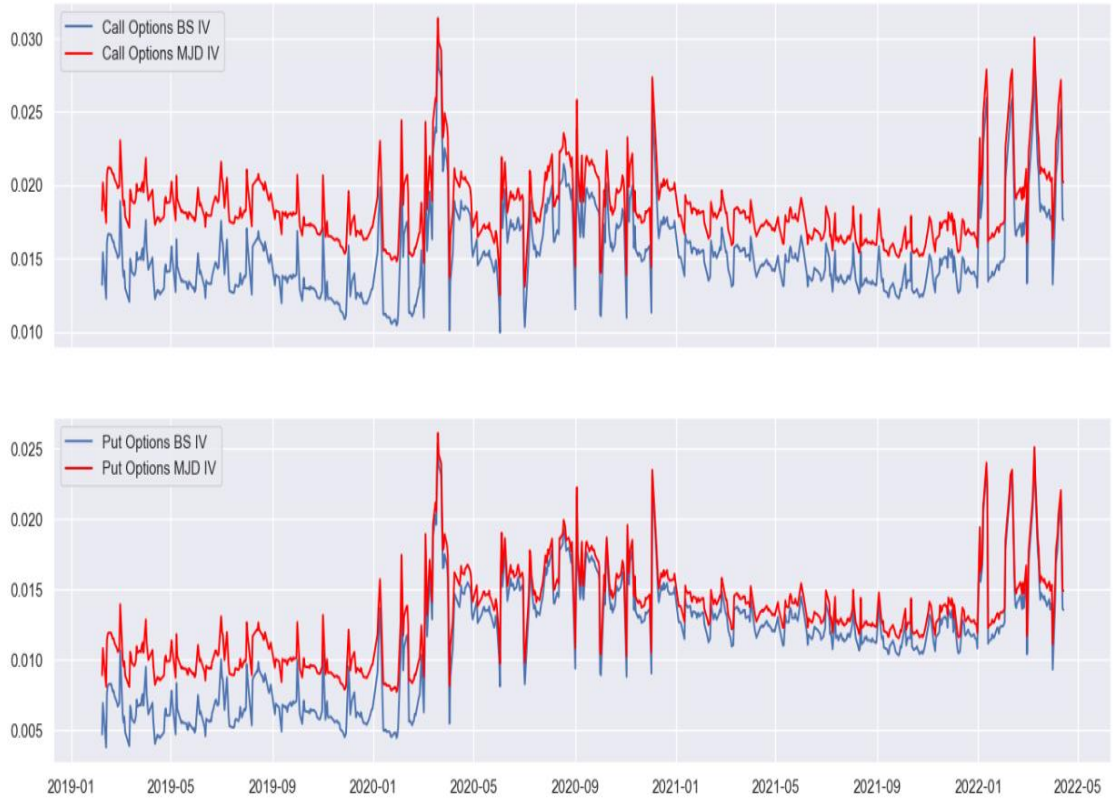


Figure 7.4.2 Call and Put Option assuming implied volatility under Black-Scholes and MJD models (blue line: Black-Scholes, red line: Merton Jump Diffusion).

In Appendix in Figures 7.4.3 and 7.4.4, they are presented different levels of μ_j and σ_j and different λ parameters. The purpose of altering these parameters is to demonstrate the importance of choosing the right levels through calibration procedures in option pricing. In this experimental setup, in the first case (7.4.3) we select $\mu_j = 0.03$, $\sigma_j = 0.06$ and $\lambda = 1.5$ while in the above setup the respective values were 0.01, 0.03 and 1. In the second setup (Figure 7.4.4), we decreased the number of the assumed jumps to 0.5 per annum, but we increased the mean of the jump to 0.1 and the volatility of the jump to 0.15. In both setups the option prices produced by the MJD model were much higher than the BS model and, at the same time, much higher than the first setup illustrated in the main body of the empirical application (see Figure 7.4.1).

https://doi.org/10.26219/heal.aueb.5795

Chapter 8

8. Conclusion

In this work we have presented the theoretical aspect of various well-known and most-used stochastic processes in financial literature. For starters, we set the theoretical foundation of Lévy processes and Ornstein Uhlenbeck processes and a combination of them, the affine jump-diffusion processes. The simulation of these processes was illustrated after implementation of Euler scheme. Moving forward, the Vasicek model was presented and simulated with various parameters and the connection between the yield curve was also illustrated. Financial data is strongly related to time and how it altered within time. In that context, when it comes to financial data analysis it is important to implement techniques coming from time-series analysis and in many occasions from machine learning, which is a fast-growing domain with many applications, one of which is the financial time-series analysis. Techniques such as autoregressive models (AR), autoregressive integrated moving average models (ARIMA) and generalized autoregressive conditional heteroscedastic models (GARCH) from time-series analysis and recurrent neural networks (RNN) and more specific long-short term memory (LSTM) were introduced and presented theoretically in the context of financial time-series.

The Black-Scholes (BS) framework and its jump-diffusion extension, known as Merton Jump-Diffusion (MJD) model, which is obtained from the BS one by adding a compound Poisson process to the main stochastic differential equation, were introduced by presenting their closed form solution, in one case (BS), and the semi-closed form solution in the other (MJD). In addition, as a final theoretical concept of this dissertation was the Vasicek Jump Diffusion model. We extended the Vasicek model by adding a compound Poisson process, assuming a normally distributed jump size. Moreover, we simulated this new model assuming different parameters each time and the produced trajectories were presented after implementing Eurel scheme. This concludes the theoretical part of this thesis.

https://doi.org/10.26219/heal.aueb.5795

The main aim of the empirical application was to first calculate the FX European option prices under BS model assuming both realized and implied volatility and after to compare the results. The option prices, both calls and puts, have significant deviance when we assume realized and implied volatility. Implied volatility is a way for traders to capture the state of the options market at a given instant. The assumption of constant volatility and the assumption of normality, both basic assumptions of BS model, do not stand in a real-world application. Moreover, the same comparison was made with respect to the Greeks sensitivities Delta and Vega. An interesting point, here, is that, while at first the Delta sensitivity behaved quite differently comparing realized and implied volatility, around 2021 and after the evolution of Delta was quite similar, with slight exceptions in some small time-periods. Additionally, Vega sensitivity had large deviations when measured assuming realized and implied volatility but after mid 2020 the behavior of Vega was quite similar assuming both volatility measures.

The BS model is based on several inputs, such as the volatility of the asset, the risk-free rate and, here the interest rates of the Eurodollar and the evolution of the asset price, here the evolution of the spot exchange rate of Eurodollar likewise. In this application, we aimed to forecast the individual price of each input implementing various techniques from financial time-series analysis and machine learning and combine these forecasts to calculate a theoretical option price based on these predictions. The evaluation metrics, assumed in this analysis, were the AIC value, the MSE and MAE of the model. For volatility, various GARCH models were fitted and based on the metrics, the best model was a GARCH (1,1) with T-student distribution for the residuals. In the case of the interest rates (3m Euribor and 3m US Libor), a simple ARIMA (1,1,0) for the Euribor and ARIMA (4,1,1) for the Libor were the best models. An interesting point here is that the ARIMA (4,1,1) and an ARIMA (4,1,0) both for Libor interest rate had the same prediction, despite the addition of a moving average component. Finally, with respect to the prediction of the spot exchange rate, the method selected was an RNN-LSTM. The model achieved quite small MSE and MAE metrics and the in-sample forecast was quite impressive (see Figure 7.3.14). The predicted value

https://doi.org/10.26219/heal.aueb.5795

of the EUR/USD was slightly overpriced compared to the real value of the Eurodollar available on any platform for the day ahead of our time-period.

The final part of our FX application was to compare the European option prices after assuming BS model and MJD model. The option prices (both calls and puts) were overpriced after implementing MJD model, compared to BS model, since the jump component (compound Poisson process) add some extra randomness in the model. A point worth highlighting here is that during the start of covid, the MJD model did not diverge from the option prices under the BS model. This is quite weird, since the addition of the jumps has the purpose of capturing fat tails events and to model better unpredicted events. However, in our application we did not calibrate the extra parameters of the MJD model, because of the lack of real option data. We select the parameters intuitively.

To conclude, the purpose of this thesis was twofold. The one part focused on the theoretical aspects of various stochastic processes mostly used in financial mathematics domain, such as Lévy, Ornstein Uhlenbeck and Jump Diffusion processes. Additionally, we presented well-known models such as Vasicek, Black-Scholes and Merton Jump diffusion model and we, also, extend some of them (Vasicek Jump Diffusion Model) on the one hand and on the other hand various techniques for handling financial time-series data. In our empirical application, the second part, we implemented and combined most of these models and techniques in FX Markets (Eurodollar exchange rate) with purpose of understanding volatility, forecasting the inputs of the BS formula, and comparing BS world and jump diffusion world.

8.1 Future Work

This work focused on comparing realized and implied volatility under the Black-Scholes framework in FX markets and, more specifically, in European Options on Eurodollar. The extension of Black-Scholes model, the Merton Jump Diffusion model was also implemented with the aim of comparing the option prices between the models. Nevertheless, in this application we have several limitations mostly because of lack of data. A first extension of this

https://doi.org/10.26219/heal.aueb.5795

work could be to calibrate the parameters of the MJD model and not to select them intuitively, as we did in our application.

Moreover, another idea is to implement more sophisticated models, such as stochastic volatility models, to capture in a better way the volatility surface of the options. In that context, some extensions of well-known models such as *Heston (1993)* can be used to option pricing and *Bates (1996)* model. In these models, we may implement more sophisticated techniques, which may use the characteristic function of the models and Fast Fournier Transform (FFT).

Finally, in this application we limited ourselves to investigating the Eurodollar, one of the most conventional currency exchange rates. A great idea and extension of this work is to investigate different exchange rates, both conventional ones and others coming from the emerging markets. The focus of this future study would be to compare conventional models such as Black-Scholes in conventional currency exchange rates to more complicated models, such as various jump-diffusion models in emerging markets exchanges rates.

https://doi.org/10.26219/heal.aueb.5795

Appendix A

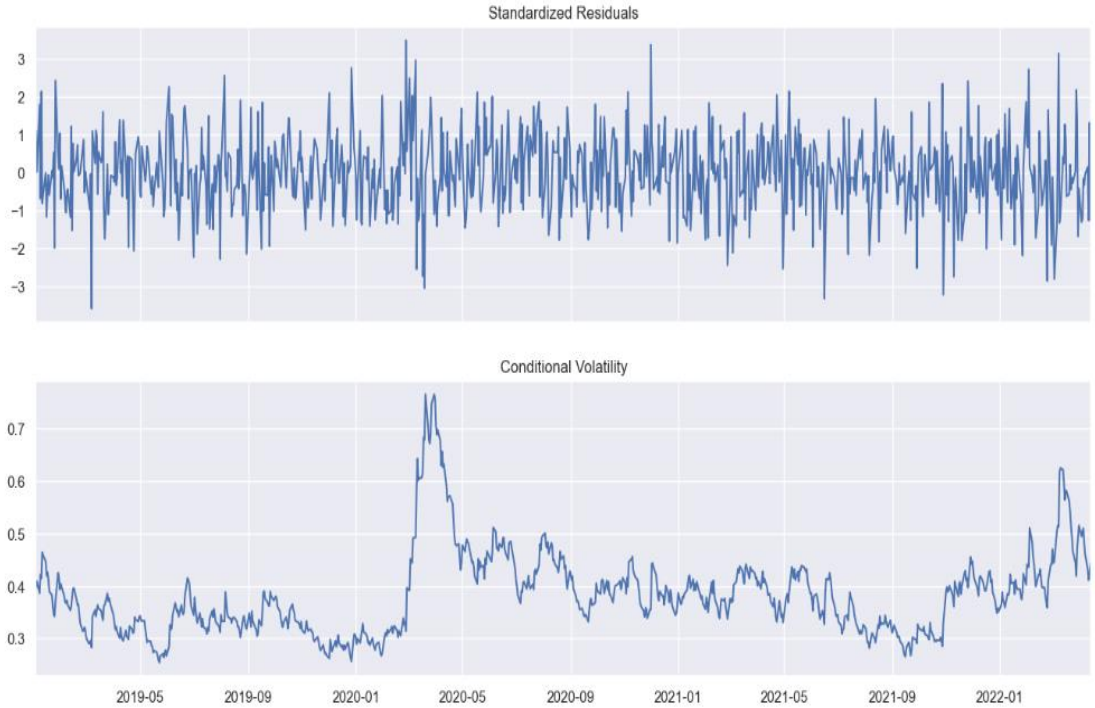
Zero Mean - EGARCH Model Results

Dep. Variable:	EURUSD	R-squared:	0.000
Mean Model:	Zero Mean	Adj. R-squared:	0.001
Vol Model:	EGARCH	Log-Likelihood:	-377.153
Distribution:	Normal	AIC:	762.306
Method:	Maximum Likelihood	BIC:	781.320
		No. Observations:	857
Date:	Tue, Jul 25 2023	Df Residuals:	857
Time:	13:43:59	Df Model:	0

Volatility Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
omega	-0.0707	2.666e-02	-2.651	8.026e-03	[-0.123,-1.842e-02]
alpha[1]	0.1463	3.312e-02	4.417	1.002e-05	[8.138e-02, 0.211]
gamma[1]	0.0152	2.150e-02	0.707	0.479	[-2.693e-02,5.734e-02]
beta[1]	0.9620	1.379e-02	69.742	0.000	[0.935, 0.989]

Table 7.3.8 EGARCH (1,1) model's output assuming normal errors.



https://doi.org/10.26219/heal.aueb.5795

Figure 7.3.11 Standardized residuals of EGARCH (1,1) model with normal errors (upper panel); conditional volatility process of EGARCH (1,1) model with normal errors.

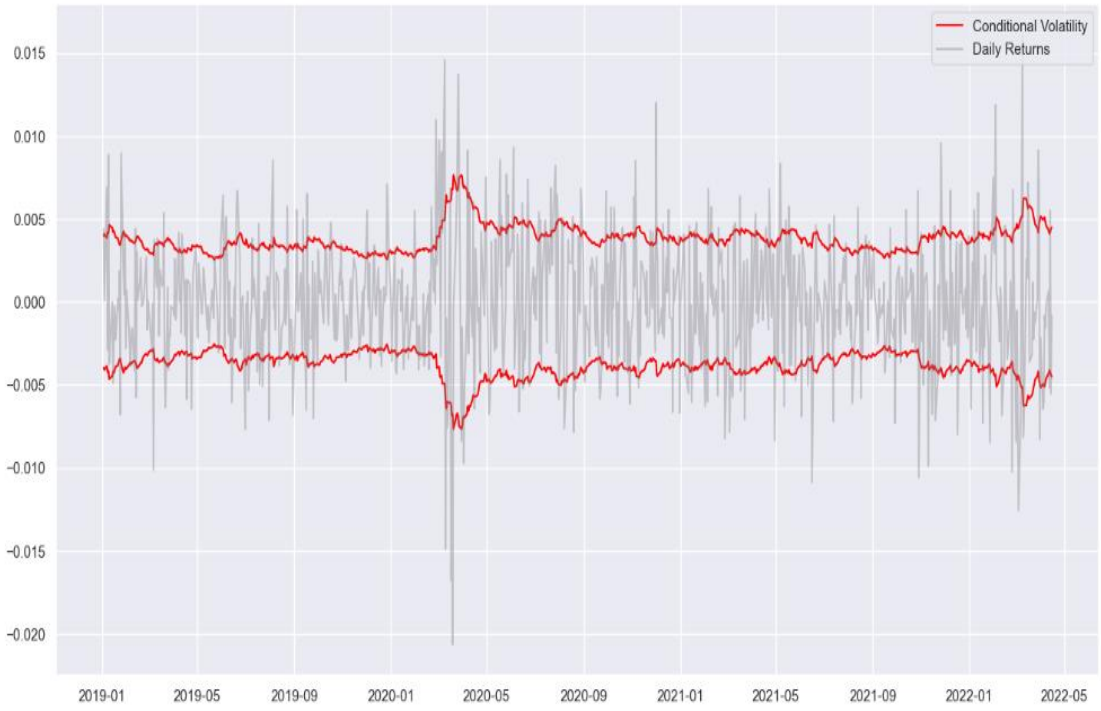


Figure 7.3.12 Conditional volatility estimation against the daily log-returns of EUR/USD spot exchange rate (EGARCH (1,1) Normal errors).

https://doi.org/10.26219/heal.aueb.5795

Zero Mean - EGARCH Model Results

Dep. Variable:	EURUSD	R-squared:	0.000
Mean Model:	Zero Mean	Adj. R-squared:	0.001
Vol Model:	EGARCH	Log-Likelihood:	-372.281
Distribution:	Standardized Student's t	AIC:	754.562
Method:	Maximum Likelihood	BIC:	778.329
		No. Observations:	857
Date:	Tue, Jul 25 2023	Df Residuals:	857
Time:	13:44:23	Df Model:	0

Volatility Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
omega	-0.0568	2.463e-02	-2.307	2.104e-02	[-0.105,-8.553e-03]
alpha[1]	0.1436	3.324e-02	4.320	1.560e-05	[7.844e-02, 0.209]
gamma[1]	0.0176	2.124e-02	0.830	0.406	[-2.399e-02,5.927e-02]
beta[1]	0.9690	1.270e-02	76.307	0.000	[0.944, 0.994]

Distribution

	coef	std err	t	P>|t|	95.0% Conf. Int.
nu	11.1813	3.585	3.119	1.817e-03	[4.154, 18.209]

Table 7.3.9 EGARCH (1,1) model's output assuming T-Student errors.

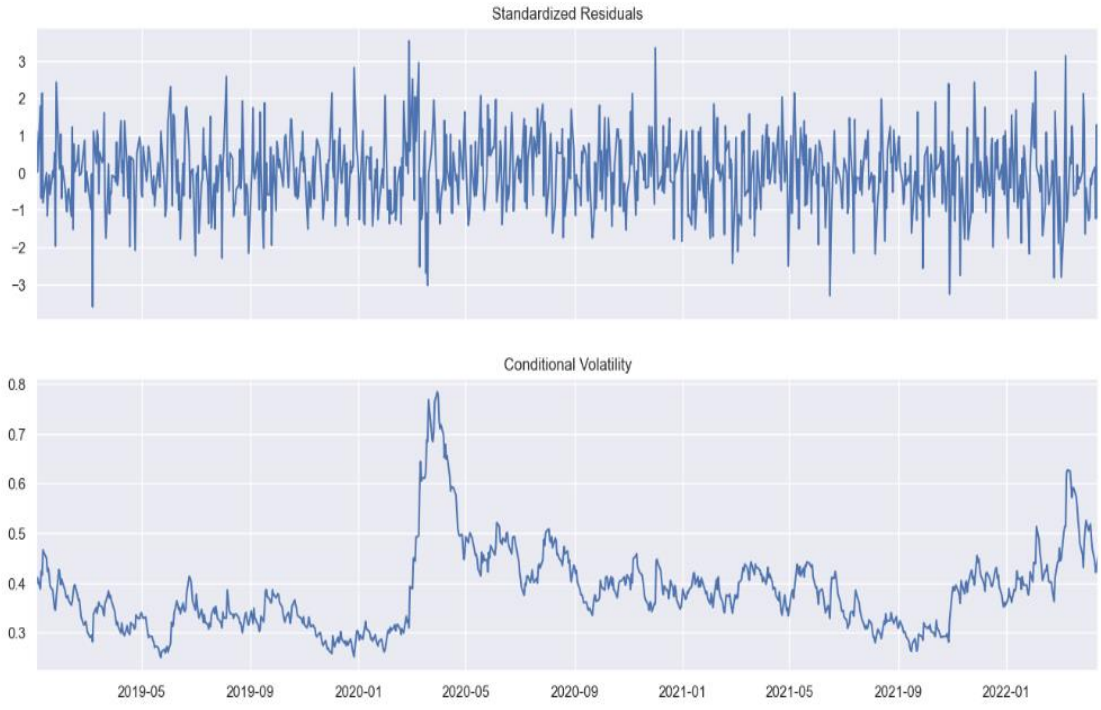


Figure 7.3.13 Standardized residuals of EGARCH (1,1) model with T-student errors (upper panel); conditional volatility process of EGARCH (1,1) model with T-student errors.

https://doi.org/10.26219/heal.aueb.5795

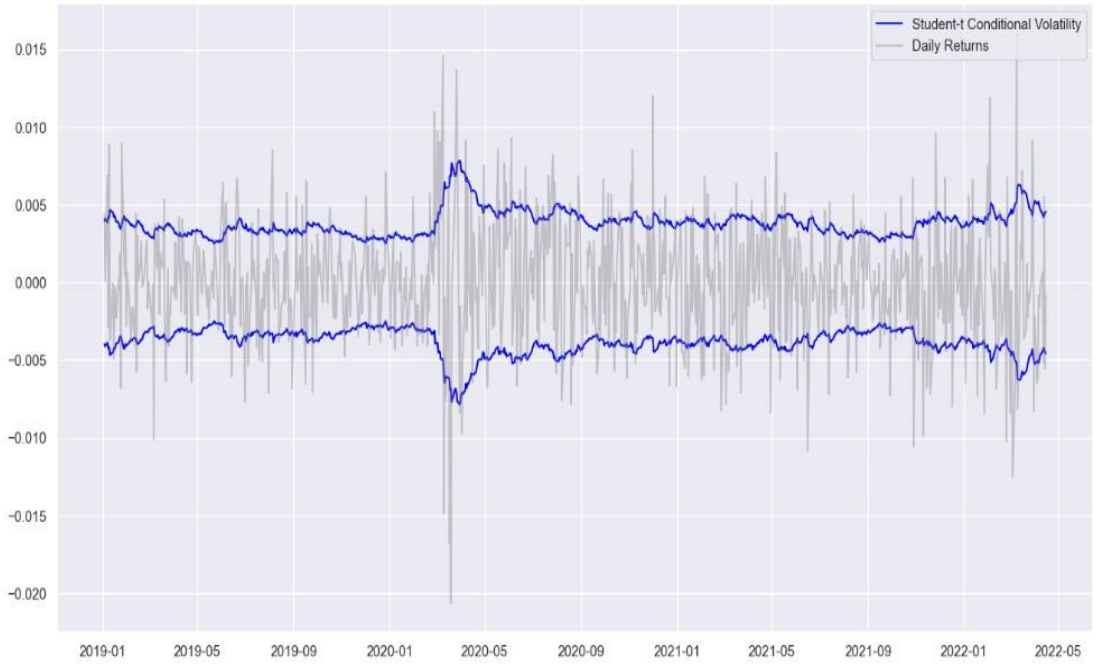


Figure 7.3.14 Conditional volatility estimation against the daily log-returns of EUR/USD spot exchange rate (EGARCH (1,1) T-Student errors).

Dep. Variable:	Euribor	**No. Observations:**	835
Model:	ARIMA(1, 1, 2)	**Log Likelihood**	6812.100
Date:	Fri, 28 Jul 2023	**AIC**	-13616.199
Time:	14:14:13	**BIC**	-13597.294
Sample:	0	**HQIC**	-13608.951
	- 835		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.9431	0.001	704.228	0.000	0.940	0.946
ma.L1	-0.8352	0.001	-718.286	0.000	-0.838	-0.833
ma.L2	-0.0943	0.001	-88.249	0.000	-0.096	-0.092
sigma2	4.757e-09	8.5e-11	55.970	0.000	4.59e-09	4.92e-09

Ljung-Box (L1) (Q):	0.01	**Jarque-Bera (JB):**	8037.92
Prob(Q):	0.90	**Prob(JB):**	0.00
Heteroskedasticity (H):	1.67	**Skew:**	1.28
Prob(H) (two-sided):	0.00	**Kurtosis:**	17.99

Table 7.3.10 ARIMA (1,1,2) model's output for 3m Euribor Interest Rate.

https://doi.org/10.26219/heal.aueb.5795

Dep. Variable:	Libor $	No. Observations:	835
Model:	ARIMA(4, 1, 1)	Log Likelihood	5941.189
Date:	Fri, 28 Jul 2023	AIC	-11870.377
Time:	17:13:47	BIC	-11842.020
Sample:	0	HQIC	-11859.505
	- 835		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.0978	0.005	21.303	0.000	0.089	0.107
ar.L2	0.4344	0.006	72.710	0.000	0.423	0.446
ar.L3	0.2309	0.004	61.954	0.000	0.224	0.238
ar.L4	-0.0939	0.002	-45.818	0.000	-0.098	-0.090
ma.L1	-0.0580	0.005	-12.293	0.000	-0.067	-0.049
sigma2	3.763e-08	5.44e-10	69.139	0.000	3.66e-08	3.87e-08

Ljung-Box (L1) (Q):	0.01	Jarque-Bera (JB):	30956.43
Prob(Q):	0.91	Prob(JB):	0.00
Heteroskedasticity (H):	0.75	Skew:	-1.63
Prob(H) (two-sided):	0.01	Kurtosis:	32.67

Table 7.3.11 ARIMA (4,1,1) model's output for 3m US Libor Interest Rate.

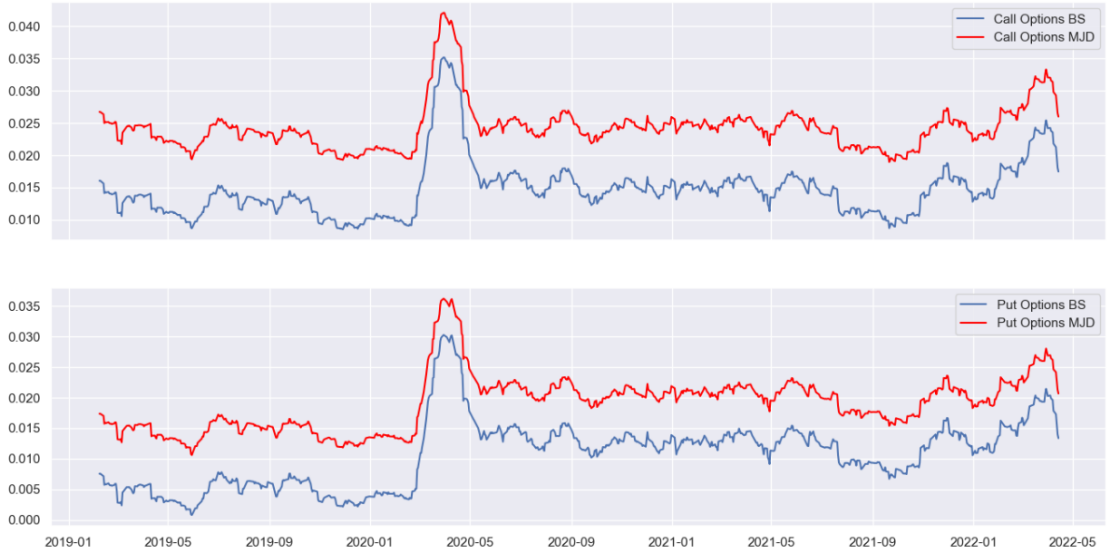


Figure 7.4.3 Call and Put Option assuming realized volatility under Black-Scholes and MJD models (blue line: Black-Scholes, red line: Merton Jump Diffusion) with different jump diffusion parameters (μj =0.03, σj = 0.06 and λ =1.5).

https://doi.org/10.26219/heal.aueb.5795

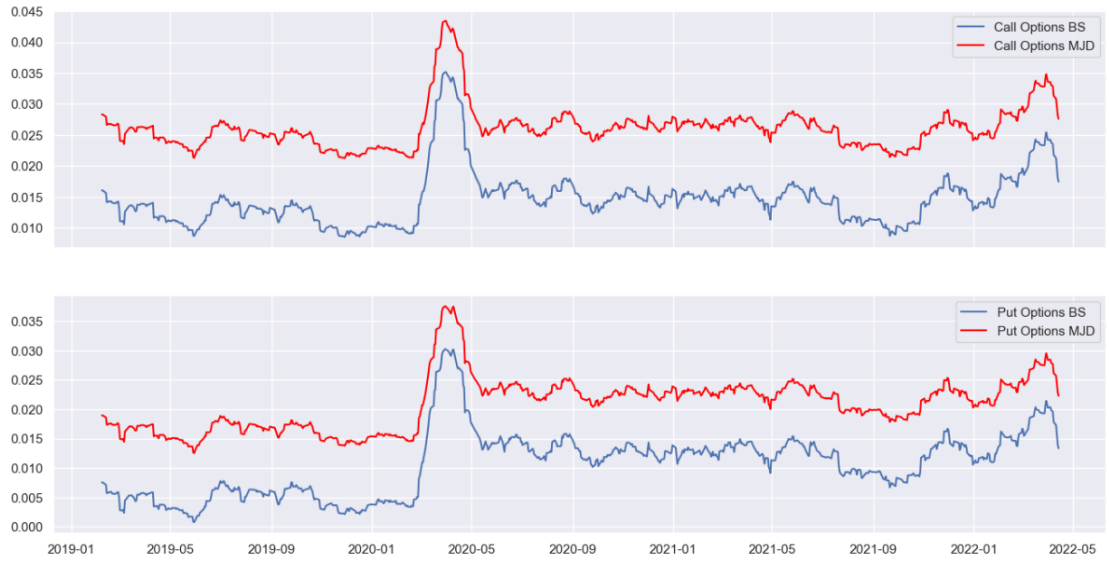


Figure 7.4.4 Call and Put Option assuming implied volatility under Black-Scholes and MJD models (blue line: Black-Scholes, red line: Merton Jump Diffusion) with different jump diffusion parameters (μj =0.1, σj = 0.15 and λ =0.5).

https://doi.org/10.26219/heal.aueb.5795

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrox, B.N., Caski, F. (Eds.), Proceedings of Second International Symposium on Information Theory. AkademiaiKiado, Budapest, pp. 67–281.

B. Oksendal (2009). Applied stochastic control of jump diffusions.

Black F. and Scholes M., The pricing of options and corporate liabilities, The Journal of Political Economy, 81, No. 3 (1973), 637-654, 10.1086/260062.

Bollerslev, T., 1986. A generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Box, G. and Jenkins, G. (1970) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

Cont, R. and Tankov, P., Financial Modelling with Jump Processes, Chapman & Hall / CRC Press, 2004.

D. Duffie, D. Filipović, W. Schachermayer "Affine processes and applications in finance," The Annals of Applied Probability, Ann. Appl. Probab. 13(3), 984-1053, (August 2003).

David S. Bates, "Jumps and Stochastic volatility: Exchange Rate Processes Implicity in Deutsche Mark Options", The Review of Financial Studies, volume 9, number 1, 1996, pages 69–107.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. Econometrica 50, 987–1008.

Garman, M. and Kohlhagen, S. (1983) Foreign Currency Options Values. Journal of International Money and Finance, 2, 231-237.

G. M. Ljung; G. E. P. Box (1978). "On a Measure of a Lack of Fit in Time Series Models". Biometrika. **65** (2): 297–303.

Heston, Steven L. (1993). "A closed-form solution for options with stochastic volatility with applications to bond and currency options". *Review of Financial Studies*. **6** (2): 327–343

Hochreiter, S., & Schmidhuber, Jurgen. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

Hull, J.C, Options, Futures, And Other Derivatives, Pearson, USA (2012).

https://doi.org/10.26219/heal.aueb.5795

Jarque, Carlos M.; Bera, Anil K. (1980). "Efficient tests for normality, homoscedasticity and serial independence of regression residuals". *Economics Letters*. **6** (3): 255–259.

Makridakis, Spyros & Hibon, Michele, 1991. "Exponential smoothing: The effect of initial values and loss functions on post-sample forecasting accuracy," International Journal of Forecasting, Elsevier, vol. 7(3), pages 317-330, November.

Mandelbrot B 1963. The variation of certain speculative prices J. Business XXXVI 392–417.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. Journal of financial economics, 3(1-2), 125-144.

Nelson, D.B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica: Journal of the Econometric Society, 59, 347-370.

Papapantoleon, Antonis (2008): An introduction to Lévy processes with applications in finance.

Thomas P. Trappenberg. Neural networks and Keras. Oxford University Press, 2019. ISBN:978-0-19-188387-3.

Tsay, R.S. (2010) Analysis of Financial Time Series. 3rd Edition, John Wiley & Sons, Hoboken.

Uhlenbeck, G. E.; Ornstein, L. S. (1930). "On the theory of Brownian Motion". Phys. Rev. 36 (5): 823–841.

Vasicek O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics, Vol. 5, 177-186.

https://doi.org/10.26219/heal.aueb.5795