UNIVERSIDADE D
COIMBRA

José Francisco Simões Silva

MARKET SENTIMENT ANALYSIS USING HERMITE POLYNOMIALS

AN EMPIRICAL APPROACH

Dissertação no âmbito do Mestrado em Métodos Quantitativos em Finanças, orientada pela Professora Doutora Ana Margarida Monteiro e pelo Professor Doutor Rui Pascoal e apresentada ao Departamento de Matemática da Faculdade de Ciências e Tecnologia e à Faculdade de Economia.

Março de 2023

Market sentiment analysis using Hermite polynomials

José Francisco Simões Silva

UNIVERSIDADE D
COIMBRA

Master in Quantitative Methods in Finance

Mestrado em Métodos Quantitativos em Finanças

MSc Dissertation | Dissertação de Mestrado

February 2023

Acknowledgements

My path during this master has truly been a fulfillment experience. With good and bad moments throughout these years, lived with amazing friends and family.

First of all, I would like to express my deep gratitude to my advisers for all the patience throughout the last year. Your guidance as crucial for the success of this research and for my academic growth.

A thank you to all the causes I fought for. They taught me the sense of mission, delivery, duty and work for my peers. Particularly, a big thank you to the organisations that allowed me to grow personally and professionally: the Erasmus Student Network, the European Campus of City Universities and the University of Coimbra.

A thank you to all the people who provided this moment. Everyone who crossed my path, acquaintances and friends, and those who will always remain. Those with whom I shared concerns and laughs, but especially the good moments.

A special thank you to my sister, Isabel, and my parents, José and Isabel, who will always be by my side. With longing love for my father, whom I lost along the way.

It could be goodbye, but it will just be a see you soon. Once Coimbra, forever *saudade*. The knowledge I acquired throughout the last years, academic and personal, has changed me and made me grow, teaching me to never give up on my goals and on the problems society faces, always perceiving the world with a curious eye.

And the future, it will be seen.

The work presented in this paper was partially carried out in the scope of the RiskBigData project: Complex Risk Management in the Big Data Regime (PTDC/MAT-APL/1286/2021), financed the Portuguese Foundation for Science and Technology (FCT).

Cofinanciado por:

Abstract

This dissertation aims to ascertain the robustness of Hermite polynomials in estimating risk-neutral densities (RND) with simulated data from the Black-Scholes-Merton (BSM) model and market data from S&P 500 (SPX) index, Arch (ARCH) Ressources and Cassava (SAVA) companies. Hermite polynomials are an expansion method, within the family of semi-nonparametric approaches for the estimation of risk-neutral densities, introduced by Madan and Milne (1994). Through comparative analysis we were able to analyse the deviation of estimated risk-neutral densities from the theoretical ones.

Furthermore, in order to extract important information regarding market sentiment, we retrieved skewness and kurtosis for the estimated risk-neutral density functions obtained from the Black-Scholes-Merton simulated data and market data. With this information we concluded that as skewness increases, kurtosis decreases; and, since we obtained leptokurtic distributions we may expect higher risk.

We observed that for simulated data from the BSM model the obtained estimates, when a noise condition is introduced, only deviates from the theoretical densities for longer maturities. Also, when maturity increases, apparently the quality of the estimation decreases, as expected. In addition, when the number of strikes is small, the estimation process is more difficult. Higher open interest, associated with more relevant option contracts, is a possible criteria for strike selection.

Finally, Hermite polynomials seem to be effective in obtaining proper RND estimates. Investor seem to be more pessimist regarding the S&P 500 index, and more confident about SAVA and ARCH companies.

Key-words: Black-Scholes-Merton model, Expansion methods, Hermite polynomials, Risk-neutral density, Semi-nonparametric methods.

Resumo

Esta dissertação visa verificar a robustez dos polinómios de Hermite na estimação de densidades neutras ao risco (RND), com dados simulados do modelo Black-Scholes-Merton (BSM) e dados de mercado do índice S&P 500 (SPX), e das empresas Arch Ressources (ARCH) e Cassava. Os polinómios de Hermite são um método de expansão, dentro da família de abordagens semi-não paramétricas para a estimação de densidades neutras ao risco, introduzido por Madan e Milne (1994). Através de análise comparativa pudemos analisar o desvio das referidas densidades neutras ao risco relativamente à densidade teórica considerada.

Além disso, a fim de extrair informações importantes sobre o sentimento do mercado, recuperamos a assimetria e o achatamento para as funções de densidade neutra ao risco estimadas, obtidas a partir de dados simulados do modelo de Black-Scholes-Merton e de dados de mercado. Com essa informação concluímos que, à medida que a assimetria aumenta, a curtose diminui; e, como obtivemos distribuições leptocúrticas, podemos esperar maior risco.

Observamos que, para os dados simulados do modelo BSM, as estimativas obtidas, quando uma condição de ruído é introduzida, se desviam das densidades teóricas perante maturidades mais longas. Além disso, quando a maturidade aumenta, aparentemente a qualidade da estimação diminui, como esperado. Quando o número de Strikes é pequeno, o processo de estimação é mais difícil. Um maior open interest está associado a uma maior relevância do contrato de opção e é um possível critério para a seleção de Strikes.

Finalmente, os polinómios de Hermite parecem ser eficazes na obtenção de estimações da RND adequadas. Os investidores parecem estar pessimistas relativamente ao índice S&P 500 e mais confiantes relativamente às empresas ARCH e SAVA.

Palavras-Chave: Modelo de Black-Scholes-Merton, Métodos de Expansão, Polinómios de Hermite, Densidade Neutra ao Risco, Métodos Semi-não Paramétricos.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Over the last decades, there has been a substantial development of financial markets. An important example of a financial market is the derivatives market. Options, swaps, forwards and futures represent an opportunity for investors: to insure their positions against price movements (hedging); to increase exposure to price movements for speculation; or, eventually, getting access to arbitrage opportunities.

A derivative is a contract between two parties that specifies conditions - in particular, dates and prices for the underlying asset - under which payments are to be made between the parties (Rubinstein, 1999) [27]. The underlying asset, on which derivative payoff depends, comprehends, for instance, commodities, stocks, bonds, interest rates or currencies. As a consequence of the 2008 subprime crisis, derivative markets are now under strict regulation to prevent such situations.

Risk is the central element that influences financial behavior. Measuring that influence and analyzing ways of controlling and allocating it allows investors to manage their equity, enhancing their investment preferences.

Options are financial instruments that reveal investors market expectations, which are of great interest in the world of finance. A major breakthrough in option pricing theory was achieved by Black-Scholes (1973) [28] and Merton (1973) [22] by presenting an option pricing formula. This was the cornerstone for developments in option pricing, formally known as the Black–Scholes–Merton (BSM) model. The model has had a huge influence on the way that traders price and hedge derivatives.

The BSM model assumes that the underlying asset price follows a geometric Brownian motion, with constant volatility. Implied volatility is the value of volatility obtained from the inverted BSM option pricing formula, by considering the remaining parameters as constants. Previous studies have shown that implied volatility seems to be non-constant across exercise prices and option's maturities (Jondeau et al., 2007) [17]. That implies the BSM assumption of a lognormal stock price distribution does not appear to occur in the markets. This evidence suggests that it is necessary to find new option pricing models more general than BSM model.

In literature, there are several methods to estimate the risk-neutral density (RND), amongst which there are semi-nonparametric methods, where expansion methods such as Hermite polynomials and Edgeworth expansions can be included. Semi-nonparametric models propose an approximation for the RND. Madan and Milne (1994) [21] exhibit how the RND can be obtained using Hermite polynomials. Jarrow and Rudd (1982) [16] use an Edgeworth expansion around a lognormal density.

Hermite polynomials have been explored by researchers over the years. This semi-nonparametric approach was further used to express Gram-Charlier expansions, as shown by Corrado and Su (1996, a) [4].

Nowadays, there are various applications for the RND, amongst the most usual are: risk management, option pricing of exotic options, and option pricing of new options in the market.

In this dissertation it is analyzed how Hermite polynomials perform in the estimation of the RND from option prices, using Madan and Milne (1994) [21] approach. To examine the reliability of this method to recover the RND, simulated data obtained from the BSM model was used. Market option prices were also considered on the S&P500 index, and two of the most shorted companies during April 2022: Arch Resources (ARCH) - coal mining and processing company; and, Cassava Sciences (SAVA) - pharmaceutical company.

This dissertation is organized in six chapters. Chapter 2 introduces the basic concepts related to options and the BSM option pricing model. Also, it presents a short summary on the volatility smile curve. Chapter 3 presents a summary of the wide variety of approaches in literature to estimate the RND, using option prices, focusing on the approaches related to expansion methods. Chapter 4 outlines how to deduce option valuation formulas and RND's with Hermite polynomials, and Gram-Charlier expansions. Chapter 5 depicts the analysis of the empirical results. Briefly, by applying Madan and Milne (1994) method, we start by showing the performed analysis for simulated option data from the BSM model, and thereafter to market option prices. Finally, in Chapter 6, we comment on the results obtained throughout the dissertation.

Chapter 2

Basic Concepts

2.1 Options

Options are financial contracts that depend on the price of the underlying asset. This agreement considers a buyer and a seller, i.e., two *counterparts*; whereby the *buyer* acquires the right to buy - *call option* - or sell - *put option* - a specific quantity of a certain asset or financial instrument (the *underlying asset*) at a pre-established price - the *exercise price* or *strike price* -, at a future date - *European options* -, called the *expiration date*, or during the period that elapses until then - *American options* -, paying the seller a given price - *premium*. The option seller assumes the obligation to buy or sell the underlying asset, under the pre-established conditions, if the buyer decides to exercise his right; the seller of the option does not have the possibility of refusing the exercise of the option. Trader who buy an option are said to have a *long position*; symmetrically, if they sell an option, they are said to have a *short position*. The underlying asset may always get traded, i.e., purchased or sold, in a market called the *spot market*. The time until an option expires is called *time to maturity* or *tenor*.

Option contracts allow the price of a transaction that takes place in the future to be fixed in the present; however, the terms of the contract are asymmetrical in the sense that the seller is subject to the buyer's will to exercise or not the option. As a result of this asymmetry, the seller only accepts to enter into the contract if he is monetarily compensated by the buyer, thus receiving the value of that right from the buyer, that is, the premium. The buyer, on the other hand, limits the amount of the loss to the premium paid for the option. In fact, the buyer has the possibility to buy or sell the underlying asset in the spot market, in which financial instruments are traded for immediate delivery, at a more favorable price, not exercising the option, and the loss results exclusively from the premium paid to the seller at the time of transaction of the option. If the exercise price is favorable, then the buyer exercises the option, and the gain is discounted from the premium previously paid.

Options can be traded in the derivatives market, such as over-the-counter markets or exchanges. In the first case, the terms of the contracts are completely shaped according to the wishes of the parties; whilst, in the second case, the options are completely standardized, leaving investors to determine their price through multilateral negotiations.The standardization of options contracts requires the establishment, by the exchange, of all the parameters of the contract, including the exercise price.

Let S_t denote the price of the underlying asset at the time t, traded in the spot market. We denote by T the expiration date of the option, and $\tau = T - t$ the tenor of the contract. Also, let K denote the strike price. The payoffs on the expiration date of a long position and a short position in call and put options are given by the following table:

	Long	Short
Call	$max(S_T - K, 0)$	$-max(S_T - K, 0) = \min(K - S_T, 0)$
Put	$max(K - S_T, 0)$	$-max(K - S_T, 0) = \min(S_T - K, 0)$

Table 2.1 Call and put option payoffs

As a result of the asymmetry of counterpart's rights-obligations, option payoffs are not linear. In this case, it is up to the buyer to decide on their exercise. The decision is made considering the relationship between the price of the underlying asset in the spot market on the exercise date, S_T, and the exercise price, K. The buyer of a call option exercises the option only when $S_T > K$, obtaining a gain of $S_T - K$. If $S_T \leq K$, exercising the call would mean that the investor would pay at least as much as the underlying asset is actually worth in the spot market, so the option is not exercised and the gain is null. Once the premium previously paid by the buyer to the seller is taken into account (without, however, taking into account the time value of money) those profits are reduced by the value of that premium, C_0, being respectively $S_T - K - C_0$ and $-C_0$.In turn, the buyer of a put option exercises the option only when $S_T < K$, assuming that the seller received a premium of P_0. The following table summarizes the profits from positions on call and put options:

Type	Position	$S_T < K$	$S_T \geq K$
Call	Long (buyer)	$-C_0$	$S_T - K - C_0$
	Short (seller)	C_0	$-(S_T - K - C_0)$
Put	Long (buyer)	$K - S_T - P_0$	$-P_0$
	Short (seller)	$-(K - S_T - P_0)$	P_0

Table 2.2 Call and put option profits

An option for which the payoff, at time T, is null is said to expire *out-of-the-money*. An option for which the payoff is positive is said to expire *in-the-money*. It is also possible to extend these notions to a moment where the option has not yet expired. For an european call (put) option, the ratio S_t/K is called *moneyness*, where r is the risk free interest rate and the exponential term accounts for the discount factor. If this ratio is larger (smaller) than one, the option is said to be in-the-money. For the situation where it is smaller (larger) than one, the option is said to be out-of-the-money. Option traders also often consider a option to be *at-the-money* if the ratio is one.

Figure 2.1 shows the gains from long and short positions in call and put options. Considering the option premium, the buyer's payoff is moving downward, and the seller's payoff is moving upward, by an amount equal to the premium.

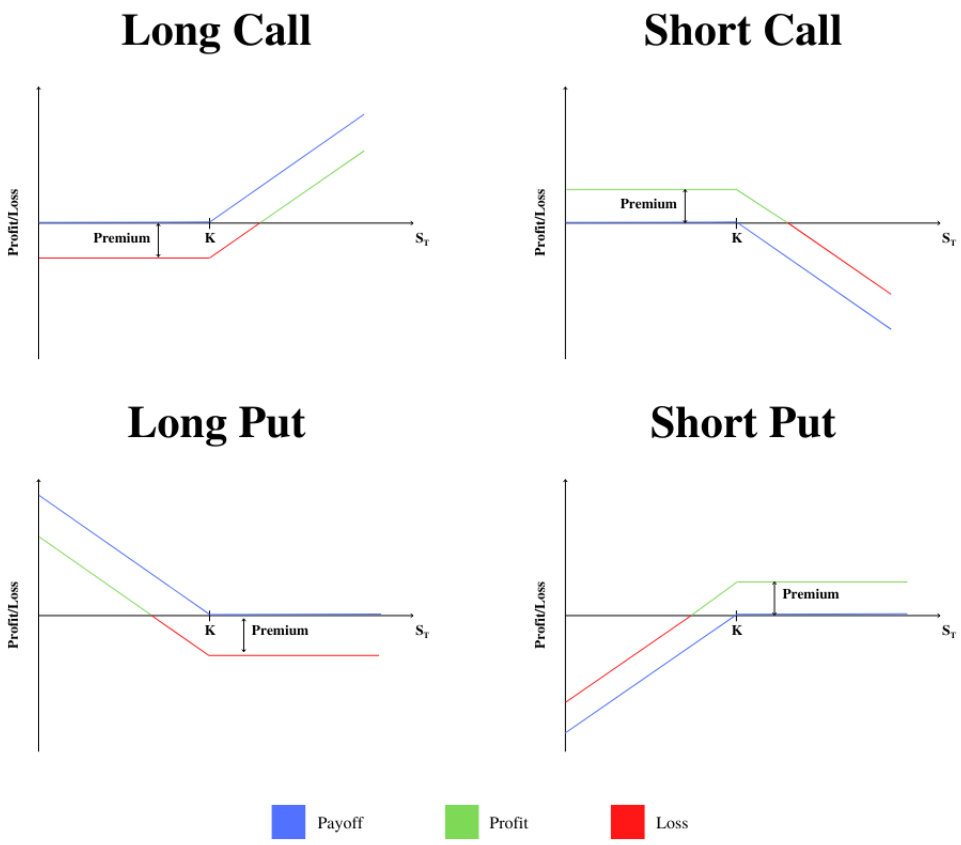

Figure 2.1 Option payoff and profit

Finally, it should be perceived that for European options, the put and call prices, denoted by C_t and P_t, where both options have the same strike and tenor, and where the underlying asset does not pay dividends, are related by the so-called *put-call parity* as follows,

$$C_t - P_t = S_t - Ke^{-r\tau}. \tag{2.1}$$

The position $C_t - P_t$ corresponds to a purchase of a call and a sale of a put. The payoff is $max(S_T - K, 0) - max(0, K - S_T)$. This equals $S_T - K$. On the right-hand side of (2.1), S_t represents the purchase of the underlying asset which will yield a value of S_T at T. In addition, $Ke^{-r\tau}$ is the discounted value of K, i.e., a credit. At time T, this portfolio will be worth $S_T - K$. Given that the payoff on the left-hand side corresponds to the one on the right-hand side, it follows that the value of the assets yielding both payoffs are also equal. Otherwise, an arbitrage opportunity would exist.

2.2 Black-Scholes-Merton Model

Black–Scholes-Merton model assumes that the market consists of at least one asset with risk, usually called the stock or the underlying asset, and one riskless asset, usually called the money market, cash, or bond.

Firstly, assumptions about the *market* are considered. There are no arbitrage opportunities (i.e., it is not possible to make a riskless profit). Investors have the ability to borrow and lend any amount, even fractional, of cash at the riskless rate; and they have the ability to buy and sell any amount, even fractional, of the stock (this includes short selling). The transactions do not incur any fees or costs (i.e., frictionless market).

Secondly, assumptions are made about the *assets*. The rate of return on the riskless asset is constant and thus called the risk free interest rate, r. The stock price, S_t, follows a geometric Brownian motion with stochastic differential equation (SDE)

$$dS_t = S_t \mu dt + S_t \sigma dW_t, \tag{2.2}$$

where dS_t denotes the instantaneous price change, μ is the expected return, σ is the volatility of the price process, and dW_t is a standard Brownian motion. The parameters μ and σ are assumed to be constant over time. The stock does not pay a dividend.

Several of these assumptions of the original model have been removed in subsequent extensions of the model. Later versions account for dynamic interest rates, transaction costs and taxes, and dividend payouts (Merton, 1973) [22], (Merton, 1974) [23], (Ingersoll, 1976) [14].

Considering these assumptions, suppose there is a derivative security (an option) also trading in this market. It is specified that this security will have a certain payoff at a specified date in the future, depending on the values taken by the stock up to that date. Even though the path the stock price will take in the future is unknown, the derivative's price can be determined at the current time.

With the pricing dynamic for the underlying asset in hand, (2.2), one must now infer the corresponding pricing dynamic for the derivative asset. We denote by $f(S_t, t)$ the price of a derivative asset and we introduce the notations $f_{SS} = \frac{\partial^2 f}{\partial S^2}$, $f_S = \frac{\partial f}{\partial S}$, $f_t = \frac{\partial f}{\partial t}$. To obtain the pricing dynamic of f, we apply Itô's lemma to (2.2), yielding

$$df = \left(\frac{1}{2}\sigma^2 S_t^2 f_{SS} + \mu S_t f_S + f_t\right) dt + \sigma S_t f_S dW_t. \tag{2.3}$$

Next, we create a portfolio consisting of one unit of the derivative asset and a short position of f_S units in the underlying asset. The portfolio value is $V_t = f - f_S S_t$, with pricing dynamic $dV_t = df - f_S dS_t$, now substituting df from (2.3) and dS_t from (2.2) we obtain

$$\begin{aligned} dV_t &= \left(\frac{1}{2}\sigma^2 S_t^2 f_{SS} + \mu S_t f_S + f_t - \mu S_t f_S\right) dt + \sigma S_t f_S dW_t - \sigma S_t f_S dW_t \\ &= \left(\frac{1}{2}\sigma^2 S_t^2 f_{SS} + f_t\right) dt + 0 dW_t. \end{aligned} \tag{2.4}$$

The dynamic of this portfolio is without risk because the term dW_t has a coefficient of 0. To avoid arbitrage, the instantaneous return of this portfolio must be the same as the risk-free rate of interest.

Hence, the change in portfolio value is

$$dV_t = rV_t dt = r\left(f - f_S S_t\right) dt. \tag{2.5}$$

From (2.4) and (2.5), we get

$$\frac{1}{2}\sigma^2 S_t^2 f_{SS} + rS_t f_S + f_t = rf.$$

This is the Black-Scholes-Merton fundamental partial differential equation (FPDE). It governs the price of all derivatives written on S_t that have a pricing dynamic given by (2.2).

The explicit solution of this fundamental partial differential equation depends on the boundary conditions related to the terms of the derivative contract. For instance, considering a European call option with strike price K and time to maturity T. The terminal payoff of this option is $max\left(S_T - K, 0\right)$. The terminal payoff defines the boundary condition for solving the FPDE. Let $C\left(S_t, t\right)$ be the price of the call option. We can then write the FPDE for the European call as

$$\frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 C}{\partial S^2} + rS_t \frac{\partial C}{\partial S} + \frac{\partial C}{\partial t} = rC, \tag{2.6}$$

with terminal constraint $C\left(S_T, T\right) = max\left(S_T - K, 0\right)$.

Given the stock price dynamics in (2.2) and considering $f(S_t, t) = log(S_t)$. By Itô's lemma we have

$$dlog\left(S_t\right) = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW_t. \tag{2.7}$$

Integrating (2.7) from t to T gives

$$log\left(S_T\right) = log\left(S_t\right) + \left(\mu - \frac{1}{2}\sigma^2\right)\left(T - t\right) + \sigma\left(W_T - W_t\right).$$

S_T has a *log-normal distribution*, so

$$log\left(S_T\right) \sim N\left(log\left(S_t\right) + \left(\mu - \frac{1}{2}\sigma^2\right)\tau, \sigma^2\tau\right) \tag{2.8}$$

where $\tau = T - t$.

The solution of (2.6), by adding proper constraints, leads to the Black–Scholes–Merton formulas for the prices of European call and put options:

$$C\left(S_t, t\right) = S_t \Phi\left(d_1\right) - Ke^{-r\tau}\Phi\left(d_2\right) \tag{2.9}$$

and

$$P\left(S_t, t\right) = Ke^{-r\tau}\Phi\left(-d_2\right) - S_t\Phi\left(-d_1\right) \tag{2.10}$$

with

$$d_1 = \frac{log\left(S_t/K\right) + \left(r + \frac{1}{2}\sigma^2\right)\tau}{\sigma\sqrt{\tau}}, \tag{2.11}$$

$$d_2 = \frac{log\left(S_t/K\right) + \left(r - \frac{1}{2}\sigma^2\right)\tau}{\sigma\sqrt{\tau}} = d_1 - \sigma\sqrt{\tau}. \tag{2.12}$$

2.3 Volatility smile curve

Considering a plot of the implied volatilities of options with a certain maturity as a function of its strike prices, its shape is known as volatility smile. Plotting implied volatility against strike prices for a given maturity, produces a skewed smile instead of the expected flat surface, in the context of the BSM model. The pattern differs across various markets. It is believed that investor reassessments of the probabilities of fat-tail leads to higher prices for in-the-money and out-of-the-money options. This anomaly implies deficiencies in the standard BSM option pricing model which assumes constant volatility and log-normal distributions of underlying asset returns. Empirical asset returns distributions, however, tends to exhibit fat-tails (leptokurtosis) and skewness. Modelling the volatility smile is an active area of research in quantitative finance, and better pricing models such as the stochastic volatility models, partially address this issue. The volatility smile for European calls with a certain maturity is the same as that for European puts with the same maturity.

Chapter 3

Literature Review

Black, Scholes and Merton revolutionized the option pricing theory with their approach. Initially, Black and Scholes (1973) [28] started off by introducing a valuation formula for options, while Merton (1973) [22] was also in parallel reaching similar conclusions within a different context, namely, extending the option pricing formula in order to include dividends and stochastic interest rates. Clearly, Black-Scholes-Merton (BSM) initial model assumes that: the rate of return of the asset is constant and thus risk-free; the asset price follows a geometric Brownian motion, and it is assumed that its drift and volatility are constant; there are no arbitrage opportunities nor dividends; finally, the market is frictionless.

However, constant volatility is not observed in the markets. BSM model considers that all variables are observale, except for the volatility. Nevertheless, by inverting the option valuation formula one is able to obtain the implied volatility. Hentschel (2003) [11] noted that implied volatilities suffer from biases if option prices are observed with errors, such as finite quote precision, bid-ask spreads, or nonsynchronous prices.

Since the option pricing formula is derived from a no arbitrage relation between the option and its underlying stock, in which market risk has been hedged away, the model and any parameters implied from it are not affected by investors' risk preferences (Figlewski, 2018) [9].

The determination of the risk-neutral density (RND) will be the main subject of analysis within the scope of this dissertation. Some initial steps were given by Cox and Ross (1976) [8] while searching for the relation between the choice of the process and the solution to option valuation problems. They state that the current option price is the discounted expected value of future payoffs under the RND.

Thereafter, Breeden and Litzenberger (1978) [3] proposed a relation between the risk-neutral density and call option prices, deriving the first from the latter's second partial derivative. Macbeth and Merville (1980) [20] suggested that the model of Cox and Ross (1976) [8] exhibited better option pricing performance than the BSM model.

Rubinstein (1985) [26] proposed an extensive work finding alternative non-parametric tests for option pricing models, which concluded that the level of market prices, the level of stock market volatility and the level of interest rates may be relevant macroeconomic variables that should be considered in future models.

Hull and White (1987) [12] followed the previous work of Merton (1973) [22] and Cox and Ross (1976) [8] to find that Black-Scholes price overvalues at-the-money options and undervalues deep in and out-of-the-money options, and that mispricing increases with the time to maturity.

Hull and White (1988) [13] extend their research to conclude that the value of a contingent claim is independent of the drift rate and of the market price of risk for a derivative which is a traded security.

The estimation of the risk-neutral density can follow various approaches. In this regard, we follow the categorisation proposed by Jondeau, Poon and Rockinger (2007) [17]. Models illustrated in the literature can be divided in two broad approaches: *structural* and *non-structural*. A model is qualified as structural if it proposes a full description for the stock price dynamics and/or for the volatility process. Whereas, a model is qualified as non-structural if it yields a description of the RND without completely describing the dynamic of the price.

Non-structural approaches can subsequently be subdivided into three categories: *parametric*, *semi-nonparametric* and *non-parametric* models.

Parametric methods involve a small number of parameters on which the RND function depends, without referring to any price dynamic. As example, there is the mixture of two lognormal distributions, with known mean and volatility, on which the BSM model is based. We have seen that a single lognormal distribution is not sufficiently flexible to fit observed option prices.

Non-parametric methods involve a big number of parameters, and does not define a specific form for the density. The number of parameters that this approaches comprise is much larger than in the parametric case.

Semi-nonparametric inference techniques have become increasingly important tools for solving statistical estimation. Semi-nonparametric models usually propose a number of parameters smaller when compared to non-parametric models and bigger when compared to parametric ones. It may appear at first that semi-nonparametric models include non-parametric models, however, a semi-nonparametric model is considered to have less parameters than a completely non-parametric model.

An important family of semi-nonparametric methods are the expansion methods, to which Edgeworth expansions, Hermite polynomials and Gram-Charlier expansions belong. Expansion methods use a straightforward probability distribution (often normal or lognormal) and then add correction terms to it. For the purpose of this dissertation, the main focus will be expansion methods. In the following paragraphs, we will present some research on Edgeworth expansions, Hermite polynomials and Gram-Charlier expansions.

Jarrow and Rudd (1982) [16] present how a given probability distribution can be approximated by an arbitrary distribution in terms of a series expansion involving second and higher moments - Edgeworth expansion. The resulting option price is expressed as the sum of a BSM price plus adjustment terms which depend on the second and higher moments of the underlying asset stochastic process. This approach allows the evaluation of the impact of skewness and kurtosis of the underlying stock's distribution, on the option price.

Madan and Milne (1994) [21] started off by presenting a family of Hermite polynomials. This was further expanded by Abken, Madan and Ramamurtie (1996) [1] that used the previous findings to price options on Eurodollar futures by using restrictions on the prices of Hermite polynomial risk for contingent claims, with different times to maturity.

Corrado and Su (1996, a) [4] use Gram-Charlier expansions for the RND to adjust skewness and kurtosis for the BSM formula. In the same year, Corrado and Su (1996, b) [5] present further research to extend the widely used procedure of obtaining implied standard deviations to also include procedures for simultaneously obtaining implied skewness and implied kurtosis coefficients.

Aït-Sahalia and Lo (1998) [2] and Jackwerth (2000) [15] show that there exists a theoretical relationship between the RND, the Subjective Density (SD) (also referred to as observed or physical density), and the Risk Aversion Function (RAF). Coutant (1999) [6] deduces the RND and SD using Hermite polynomials' expansions to obtain a time-varying estimator of the investors' RAF.

Guasoni (2004) [10] adapts the Madan and Milne model to a multiple expiration setting and considers an estimation method for the RND at a moving horizon of fixed length, which exploits the prices of options with different maturities.

Rompolis and Tzavalis (2007) [24] employ a method to retrieve the RND of future asset prices, or their implied log-returns, based on an exponential form of a Gram–Charlier series expansion. This type of expansion guarantees that the values of the RND will always be positive, and it can account for strong deviations of the stock price distributions from the Gaussian one. They conclude that this type of Gram–Charlier expansion can sufficiently approximate the true RND. Later, the same authors tackle the estimation of the RND for option pricing models, whose risk-neutral density form is not given [25].

Chapter 4

Expansion Methods

In this chapter, we outline some expansion methods available to retrieve the RND. Firstly, we present the fundamentals of Hermite polynomials, as it is the method we employ in Chapter 5. Thereafter, in Section 4.2, we present the fundamentals for Gram-Charlier expansions, as they are defined using Hermite polynomials (Corrado and Su, 1996, a) [4].

4.1 Hermite polynomials

Hermite polynomials technique is a semi-nonparametric approach to estimate RND for the underlying asset, based on the Madan and Milne (1994) [21] approximation of the Gaussian density, and implemented by Abken, Madan and Ramamurtie (1996) [1], and Coutant, Jondeau and Rockinger (2001) [7], on pricing call options. The main objective within the scope of this dissertation is to estimate the RND, based on Hermite polynomials.

The Gaussian density of a normal distribution with parameters (μ, σ^2) is given by

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{4.1}$$

In order to present the Hermite polynomials approach we will simplify (4.1) and instead of considering the entire expression, we will consider e^{-x^2}. Therefore, its n-th derivative, $\frac{d^n(e^{-x^2})}{dx^n}$, is evaluated, as follows:

$$\begin{aligned}
\frac{d(e^{-x^2})}{dx} &= -2xe^{-x^2} \\
\frac{d^2(e^{-x^2})}{dx^2} &= -2e^{-x^2} + 4x^2e^{-x^2} = (4x^2-2)e^{-x^2} \\
\frac{d^3(e^{-x^2})}{dx^3} &= 8xe^{-x^2} + (-2x)(4x^2-2)e^{-x^2} = -(8x^3-12x)e^{-x^2}.
\end{aligned}$$

Proceeding iteratively, we obtain the following expression for the n-th order derivative,

$$\frac{d^n(e^{-x^2})}{dx^n} = P_n(x)e^{-x^2},$$

where $P_n(x)$ is a polynomial that can be further written as $(-1)^n H_n(x)$, that is

$$\frac{d^n(e^{-x^2})}{dx^n} = (-1)^n H_n(x)e^{-x^2}. \tag{4.2}$$

Thus, we can present the Hermite polynomials as the $H_n(x)$ polynomials. The term $(-1)^n$ will allow for the highest order coefficient of $H_n(x)$ to always be positive.

From (4.2) we have

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n(e^{-x^2})}{dx^n}. \tag{4.3}$$

Generally, if we re-introduce the Gaussian density (4.1) instead of e^{-x^2}, then Hermite polynomials could be defined as

$$H_n(x) = (-1)^n \frac{1}{\phi(x)} \frac{d^n(\phi(x))}{dx^n}.$$

4.1.1 Recurrence Relations

The first derivative of (4.3) is given by

$$H_n^{'}(x) = (-1)^n \left[2xe^{x^2} \frac{d^n(e^{-x^2})}{dx^n} + e^{x^2} \frac{d^{n+1}(e^{-x^2})}{dx^{n+1}} \right] = 2xH_n(x) - H_{n+1}(x). \tag{4.4}$$

On the other hand, let us consider the general Leibniz rule that generalizes the product derivative rule and states that, if f and g are n-times differentiable functions, then the product fg is also n-times differentiable and its n-th derivative is given by

$$(fg)^{(n)} = \sum_{k=0}^{n} \binom{n}{k} f^{(n-k)} g^{(k)}.$$

Hence, we can rewrite the derivative of the Hermite polynomial as

$$\begin{aligned} H_n^{'}(x) &= (-1)^n \left[2xe^{x^2} \frac{d^n(e^{-x^2})}{dx^n} + e^{x^2} \frac{d^n\left(-2xe^{-x^2}\right)}{dx^n} \right] \\ &= (-1)^n \left[2xe^{x^2} \frac{d^n(e^{-x^2})}{dx^n} - 2e^{x^2} \sum_{k=0}^{n} \binom{n}{k} x^{(k)} \left(e^{-x^2}\right)^{(n-k)} \right]. \end{aligned}$$

In the sum above, since we are considering derivatives of x, only two terms are going to remain, that is, we are only considering the 0-th derivative of x and the first derivative, because every other term is going to be equal to 0. Thus,

$$\begin{aligned} H_n^{'}(x) &= (-1)^n \left[2xe^{x^2} \frac{d^n(e^{-x^2})}{dx^n} - 2e^{x^2} \left[x\frac{d^n(e^{-x^2})}{dx^n} + n\frac{d^{n-1}(e^{-x^2})}{dx^{n-1}} \right] \right] \\ &= (-1)^{n-1} \left[2ne^{x^2} \frac{d^{n-1}(e^{-x^2})}{dx^{n-1}} \right] \\ &= 2nH_{n-1}(x). \end{aligned} \tag{4.5}$$

Finally, from (4.4) and (4.5) we can obtain $H_{n+1}(x)$ from $H_n(x)$ and $H_{n-1}(x)$:

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x). \tag{4.6}$$

4.1.2 Orthogonality

Our goal is to present the orthogonality of Hermite polynomials,

$$\int_{-\infty}^{\infty} H_n(x)H_m(x)e^{-x^2}dx = 2^n n! \sqrt{\pi}\, \mathbb{1}_{\{m=n\}}(m,n). \tag{4.7}$$

where $\mathbb{1}$ is the indicator function, if n = m it corresponds to one, if not it corresponds to zero. Note that m and n can be interchanged. The term e^{-x^2} is important since it allows the integral to converge, as a consequence of the product of the polynomials $H_n(x)$ and $H_m(x)$.

In order to analyse orthogonality, consider the integral

$$(-1)^m \int_{-\infty}^{\infty} H_n(x) \frac{d^m(e^{-x^2})}{dx^m} dx, m > n. \tag{4.8}$$

For simplification in the following calculus, we ignore the $(-1)^m$ term, that will be added later.

Considering n $\neq$ m, the integral from (4.8) and integrating it by parts, we have

$$H_n(x) \left[\frac{d^{m-1}(e^{-x^2})}{dx^{m-1}} \right]_{-\infty}^{+\infty} - \int_{-\infty}^{\infty} \frac{d(H_n(x))}{dx} \frac{d^{m-1}(e^{-x^2})}{dx^{m-1}} dx. \tag{4.9}$$

In (4.9), the first term will be equal to zero since the $(m-1)$-derivative includes the exponential term. Therefore, this term converges, and it is equal to zero. We are left with the second term, which we can again integrate by parts; however, we have again a derivative of e^{-x^2}. When we do m integration's by parts we get

$$(-1)^m \int_{-\infty}^{\infty} \left(\frac{d^m(H_n(x))}{dx^m} \right) e^{-x^2} dx.$$

The term on brackets is the m-th derivative of the n-th Hermite polynomial. Since $m > n$, this term will be equal to zero, and we can conclude that

$$(-1)^m \int_{-\infty}^{+\infty} H_n(x) \frac{d^m(e^{-x^2})}{dx^m} dx = 0.$$

Considering $n = m$,

$$\int_{-\infty}^{+\infty} H_n(x)\frac{d^n(e^{-x^2})}{dx^n}dx = (-1)^n \int_{-\infty}^{+\infty} \left(\frac{d^n(H_n(x))}{dx^n}\right) e^{-x^2}dx.$$

We will now compute the parameter in brackets. Applying successively the recurrence relation (4.5), we obtain

$$\frac{d^n(H_n(x))}{dx^n} = 2n\frac{d^{n-1}(H_{n-1}(x))}{dx^{n-1}} = 2n \times 2(n-1) \times \frac{d^{n-2}(H_{n-2}(x))}{dx^{n-2}} = \cdots = 2^n n!,$$

so the integral will be,

$$(-1)^n 2^n n! \int_{-\infty}^{+\infty} e^{-x^2}dx = (-1)^n 2^n n!\sqrt{\pi}.$$

To finish this deduction, we just need to multiply this for $(-1)^n$, and now we can conclude that the Hermite polynomials are equal to zero, for $n \neq m$; and equal to $2^n n!\sqrt{\pi}$, for $n = m$.

4.1.3 Generating Function

Now, we are going to present e^{2xt-t^2} as a generating function for Hermite polynomials, in the sense that Taylor series coefficients for the function e^{2xt-t^2} are Hermite polynomials, as follows,

$$e^{2xt-t^2} = \sum_{n=0}^{\infty} \frac{H_n(x)}{n!}t^n. \tag{4.10}$$

This is similar to claiming that

$$H_n(x) = \left[\frac{d^n\left(e^{2xt-t^2}\right)}{dt^n}\right]_{t=0}. \tag{4.11}$$

By simplifying (4.11) we get

$$H_n(x) = \left[\frac{d^n\left(e^{2xt-t^2}\right)}{dt^n}\right]_{t=0} = \left[\frac{d^n\left(e^{-(x-t)^2+x^2}\right)}{dt^n}\right]_{t=0} = e^{x^2}\left[\frac{d^n\left(e^{-(x-t)^2}\right)}{dt^n}\right]_{t=0}.$$

Taking into account that $\frac{df(x-t)}{dt} = -\frac{df(x-t)}{dx}$, we have,

$$H_n(x) = e^{x^2}\left[\frac{d^n\left(e^{-(x-t)^2}\right)}{dt^n}\right]_{t=0} = (-1)^n e^{x^2}\left[\frac{d^n\left(e^{-(x-t)^2}\right)}{dx^n}\right]_{t=0} = (-1)^n e^{x^2}\frac{d^n\left(e^{-x^2}\right)}{dt^n},$$

as stated by (4.3).

Therefore, we can conclude, from equation (4.10), that the n-th Hermite polynomial corresponds to the n-th coefficient in the Taylor series for the function e^{2xt-t^2}. Equation (4.10) allows to identify the generating function for the Hermite polynomial as e^{2xt-t^2}.

4.1.4 Option pricing formula

Following Jondeau, Poon, Rockinger (2007) [17] we start, by considering a change of variable. We will define the standardized log-return z as a function of the actual prices, denoting the volatility over the horizon of the option as $s = \sigma\sqrt{\tau}$,

$$S_T = S_t exp\left(\mu\tau - \frac{1}{2}s^2 + sz\right) \Rightarrow z = \frac{log\left(\frac{S_T}{S_t}\right) - \left(\mu\tau - \frac{s^2}{2}\right)}{s}.$$

Let q be the RND for the underlying asset, S_t. The following change of variable allows to obtain $q(.)$ in terms of $q_z(.)$:

$$q(S_T)dS_T = q_z\left(\frac{log\left(S_T/S_t\right) - \left(\mu\tau - s^2/2\right)}{s}\right) \times \frac{1}{S_T} \times \frac{1}{s} \times dS_T. \tag{4.12}$$

Focusing on a call option, its payoff could be given, as a function of z,

$$c(z, S_t, K, \mu, s, \tau) = max\left(S_t\left(\mu\tau - \frac{1}{2}s^2 + sz\right) - K, 0\right).$$

The price of a call option is given by

$$C(z, S_t, K, \mu, s, \tau) = e^{-r\tau}\int_0^{+\infty} c(z, S_t, K, \mu, s, \tau)q_z(z)dz, \tag{4.13}$$

where $q_z(.)$ denotes the risk neutral density for z. In addition, we can generalize this function so that it holds for an option having a generic payoff, $g(z)$, to which corresponds the price

$$C(S_t, K, \mu, s, \tau, r) = e^{-r\tau}\int_0^{+\infty} g(z)q_z(z)dz. \tag{4.14}$$

It is assumed that $g(z)$ may be defined by basis functions constituted by Hermite standardized polynomials $h_k(z)$, written as

$$g(z) = \sum_{k=0}^{\infty} a_k h_k(z), \tag{4.15}$$

where $h_k(z)$ is given by,

$$h_k(z) = H_k(z)/\sqrt{k!},$$

see Madan and Milne (1994) [21].

Following Abken, Madan and Ramamurtie (1996) [1], we present the first five terms:

$$\begin{aligned}
h_0(z) &= 1, \\
h_1(z) &= z, \\
h_2(z) &= \frac{1}{\sqrt{2}}(z^2-1), \\
h_3(z) &= \frac{1}{\sqrt{6}}(z^3-3z), \\
h_4(z) &= \frac{1}{\sqrt{24}}(z^4-6z^2+3).
\end{aligned}$$

By replacing (4.15) in (4.14), we have

$$C(S_t,K,\mu,s,\tau,r) = e^{-r\tau}\sum_{k=0}^{\infty} a_k \int_z h_k(z)q_z(z)dz. \tag{4.16}$$

In addition, the risk-neutral density $q_z(.)$ may be obtained by multiplying $\phi(z)$ by the risk-neutral change of measure $\lambda(z)$, given in terms of Hermite polynomials expansion, as follows, see Madan and Milne (1994) [21],

$$\lambda(z) = e^{r\tau}\sum_{l=0}^{\infty} \pi_l h_l(z), \tag{4.17}$$

hence,

$$q_z(z) = \phi(z)\lambda(z). \tag{4.18}$$

Therefore, by replacing (4.17) in (4.18) we obtain

$$q_z(z) = \phi(z)\left(e^{r\tau}\sum_{l=0}^{\infty} \pi_l h_l(z)\right), \tag{4.19}$$

where π_l is interpreted as the implicit price of non-traded risk, given by $h_l(z)$. In what follows, we will truncate the infinite sum up to the fourth order and to insure $q_z(z)$ is a density we impose $\pi_0 = e^{-r\tau}$. We can estimate the mean μ and the variance σ^2 of log-returns and set the mean π_1 and the variance π_2 equal to 0, since z is the standardized log-return. π_3 and π_4 are the implicit prices of skewness and kurtosis, respectively. Accordingly,

$$\begin{aligned}
q_z(z) &\approx \phi(z)\left(e^{r\tau}\sum_{l=0}^{4} \pi_l h_l(z)\right) = \phi(z)e^{r\tau}\left(e^{-r\tau} + \pi_3 h_3(z) + \pi_4 h_4(z)\right) \\
&= \phi(z)\left(1 + \frac{b_3}{\sqrt{6}}(z^3-3z) + \frac{b_4}{\sqrt{24}}(z^4-6z^2+3)\right),
\end{aligned} \tag{4.20}$$

where the $b_i = e^{r\tau}\pi_i, i = 3,4$, are the future value of the i-th implicit price of the risk coefficient. Parameters b_3 and b_4 are the skewness and kurtosis, z is assumed to follow a normal distribution. Parameters μ, s, b_3 and b_4 are estimated through a nonlinear objective function. The skewness and

kurtosis for expanded density of z are

$$\begin{aligned} Sk[z] &= \sqrt{6}b_3, \\ Ku[z] &= 3+\sqrt{24}b_4, \end{aligned} \tag{4.21}$$

see Jondeau and Rockinger (2001) [18].

Substituting (4.19) in (4.16) we obtain

$$C(S_t,K,\mu,s,\tau,r) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} a_k\pi_l \int_z h_l(z)h_k(z)\phi(z)dz = \sum_{k=0}^{\infty} a_k\pi_k,$$

where the second equality results from the orthogonality of Hermite polynomials. Truncating this sum to the fourth order, we have

$$C(S_t,K,\mu,s,\tau,r) = e^{-r\tau}a_0 + \pi_3 a_3 + \pi_4 a_4. \tag{4.22}$$

Abken, Madan and Ramamurtie (1996) [1], use a call option generating function

$$G(t,S_t,K,\mu,s,\tau) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} c(z,S_t,K,\mu,s,\tau)e^{-(z-t)^2/2}dz,$$

to prove that

$$\begin{aligned} a_0(z) &= S_t e^{\mu\tau}\Phi(d_1) - K\Phi(d_2), \\ a_1(z) &= sS_t e^{\mu\tau}\Phi(d_1) + S_t e^{\mu\tau}\phi(d_1) - K\phi(d_2), \\ a_2(z) &= \frac{1}{\sqrt{2}}\left[s^2 S_t e^{\mu\tau}\Phi(d_1) + 2sS_t e^{\mu\tau}\phi(d_1) + S_t e^{\mu\tau}\phi^{'}(d_1) - K\phi^{'}(d_2)\right], \\ a_3(z) &= \frac{1}{\sqrt{6}}\left[s^3 S_t e^{\mu\tau}\Phi(d_1) + 3s^2 S_t e^{\mu\tau}\phi(d_1) + 3sS_t e^{\mu\tau}\phi^{'}(d_1)\right. \\ &\quad \left. + S_t e^{\mu\tau}\phi''(d_1) - K\phi^{''}(d_2),\right. \\ a_4(z) &= \frac{1}{\sqrt{24}}\left[s^4 S_t e^{\mu\tau}\Phi(d_1) + 4s^3 S_t e^{\mu\tau}\phi(d_1) + 6s^2 S_t e^{\mu\tau}\phi^{'}(d_1) + 4sS_t e^{\mu\tau}\phi^{''}(d_1)\right. \\ &\quad \left. + S_t e^{\mu\tau}\phi^{'''}(d_1) - K\phi^{'''}(d_2),\right. \end{aligned} \tag{4.23}$$

where d_1 and d_2 are given by (2.11) and (2.12). Therefore, (4.22) can be represented as a function of the variables μ, s, π_3 and π_4, denoted by $C(\mu,s,\pi_3,\pi_4)$.

4.2 Gram-Charlier expansions

Later work on Hermite polynomials allowed researchers to reach a new type of semi-nonparametric approach relying on these polynomials: Gram-Charlier expansions. Gram-Charlier expansions have been largely used as a semi-nonparametric approach to overcome the restriction imposed by the usual normality assumption, see Corrado and Su (1996, a) [4], Jondeau and Rockinger (2001) [18], Rompolis and Tzavalis (2007) [24], Rompolis and Tzavalis (2008) [25] and Lin et al. (2015) [19].

A Gram-Charlier series expansion of a density function f is defined as

$$f(x) = \sum_{n=0}^{\infty} c_n H_n \phi(x) \tag{4.24}$$

where ϕ is the standardized normal density function, H_n are Hermite polynomials, and the coefficients c_n are determined by moments of the distribution function F. However, usually the series is truncated to exclude terms beyond the fourth moment. The resulting truncated density provides an approximation that accounts for non-normal skewness and kurtosis. Specifically, after standardizing to a zero mean and unit variance, a truncated series that accounts for skewness and kurtosis yields the following density function where μ_3 and μ_4 denote standardized coefficients of skewness and kurtosis, respectively

$$q(z) = \phi(z)\left[1 + \frac{\mu_3}{3!}(z^3 - 3z) + \frac{\mu_4 - 3}{4!}(z^4 - 6z^2 + 3)\right]. \tag{4.25}$$

Equation (4.25) is similar to equation (4.20), due to the fact that (4.24) is expressed through Hermite polynomials.

From the density function (4.25) we can present the following expected values: $E(z) = 0$, $E(z^2) = 1$, $E(z^3) = \mu_3$, $E(z^4) = \mu_4$, see Corrado and Su (1996, a) [4]. Thus, the coefficients of skewness and kurtosis for $q(z)$ are explicit parameters in its functional form. Under a normal distribution we have the skewness and kurtosis coefficients $\mu_3 = 0$ and $\mu_4 = 3$, respectively, which substituted into (4.25) correspond to the special case of a standard normal density.

Assuming risk neutrality, we apply equation (4.25) to derive a theoretical European call option price as the present value of an expected payoff at option expiration. This option price is derived from the following expression, see Breeden and Litzenberger (1978) [3],

$$C_{GC} = e^{-rt}\int_k^{\infty} (S_T - K) q(z) dz. \tag{4.26}$$

From the integral, an option price formula is obtained on a Gram-Charlier density expansion:

$$C_{GC} = C_{BSM} + \mu_3 Q_3 + (\mu_4 - 3) Q_4 \tag{4.27}$$

where C_{BSM} is the Black-Scholes-Merton option pricing formula (2.9),

$$Q_3 = S_t e^{-r\tau}\frac{s}{3!}\left[(2s - d_1)\,\phi(d_1) + s^2\Phi(d_1)\right]$$
$$Q_4 = S_t e^{-r\tau}\frac{1}{4!}\left[s\left(d_1^2 - 1 + 3s(s - d_1)\right)\phi(d_1) + s^3\Phi(d_1)\right].$$

In equation (4.27), Q_3 and Q_4 represent the marginal effect of non-normal skewness and kurtosis, respectively, on the option price C_{GC}.

Equation (4.27) is vital when using Gram-Charlier expansions, because when stock returns are normally distributed, then $\mu_3 = 0$ and $\mu_4 = 3$, and it presents the BSM option price formula; on the other hand, if $\mu_3 \neq 0$ and/or $\mu_4 \neq 3$, equation (4.27) is the sum of a BSM option price plus adjustment terms for non-normal skewness and kurtosis.

Chapter 5

Empirical analysis

In this chapter, we will analyse the effectiveness of the Hermite polynomial expansion approach in the estimation of the RND, using theoretical and market option prices. Theoretical option prices will be simulated from the Black-Scholes-Merton model presented in Chapter 2. This simulation analysis is performed using the *blsprice* routine of MATLAB. Afterwards, we will use historical market option prices on the S&P 500 index and two of the most shorted companies during April 2022, quoted in the Chicago Board Options Exchange (CBOE) - Arch Resources and Cassava Sciences.

We will proceed with the implementation of the model proposed by Madan and Milne (1994) [21] to estimate the RND, applying it to both simulated and market data, through a routine constructed in MATLAB. In order to simplify our notation, we will aggregate the parameter values $\theta = (S_t, K, r, \tau)$ and consider (4.22) as a function of variables μ, s, π_3, π_4. The optimization problem is then given by

$$\begin{aligned} \underset{\mu,s,\pi_3,\pi_4}{minimize} \quad & \sum_{n=1}^{NC} \left(C_i - C\left(\mu, s, \pi_3, \pi_4\right)\right)^2 \\ \text{s.t.} \quad & 0 < \mu < 1 \\ & 0 < s < 1 \\ & -1 < \pi_3 < 2 \\ & 0 < \pi_4 < 5 \end{aligned}$$

where $C_i\,(i = 1, ..., NC)$ are the observed call option prices.

Comparisons between the estimated and theoretical BSM distributions will be done throughout this chapter.

5.1 Black-Scholes-Merton data analysis

This first section aims to present the result of using Hermite polynomials to estimate RND functions, with data generated from the BSM model.

To analyse the data generated by the BSM model we set the parameters as follows: the underlying asset price, $S_t = 40$; the risk-free interest rate, $r = 0.01$; the volatility, $\sigma = 0.25$; and, $\tau = \left\{\frac{1}{12}, \frac{3}{12}, \frac{6}{12}, 1\right\}$. We considered fifty strike prices, K, equally spaced between twenty and eighty, for each maturity.

To reproduce market's behaviour, namely price fluctuations, random noise, ξ, is added to the theoretical option prices, C_{BSM}, so that the perturbed prices, C^*_{BSM}, are given by

$$C^*_{BSM} = C_{BSM} + 0.01 \times C_{BSM} \times \xi_i, \tag{5.1}$$

where $\xi_i \sim N(0,1)$.

Firstly, we present the results without noise, followed by the results using the (5.1) noise condition. After this, we will analyse skewness and kurtosis for the estimated RND from the BSM data, with and without the noise condition.

5.1.1 Risk-Neutral Density estimation - without noise

We started by computing call and put option prices through the BSM model, with the parameters presented before. After RND estimation using Hermite polynomials, we were able to compare theoretical and estimated RNDs for the standardized log-returns, (4.20), and for the underlying asset price, S_t. The estimative obtained for the log-returns is compared to the theoretical Normal distribution, and the estimative obtained for the underlying asset price is compared to the Lognormal distribution.

Analysing Figures 5.1 to 5.4 we can observe they are, apparently, well fitted. They are related to the four maturities presented before.

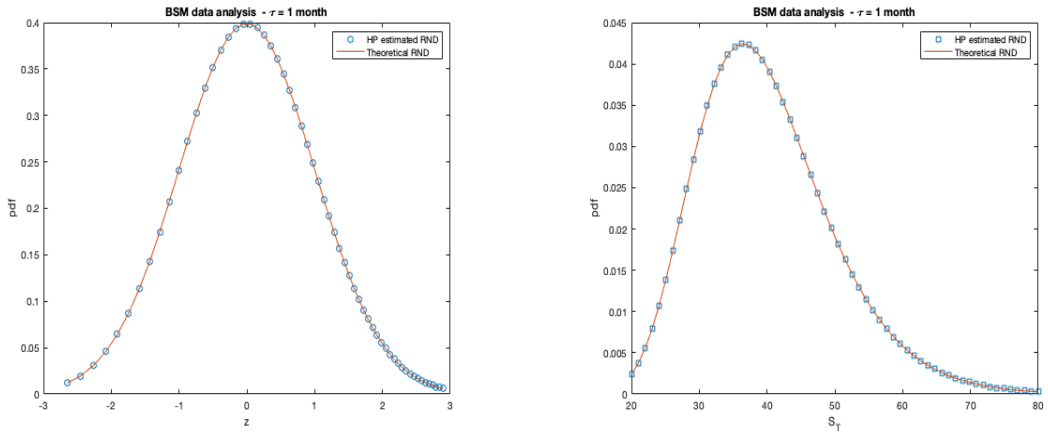

Figure 5.1 Theoretical and estimated RND functions from BSM data, without noise - Maturity: 1 month. On the left, for the standardized log-returns; and, on the right, for the underlying asset price.

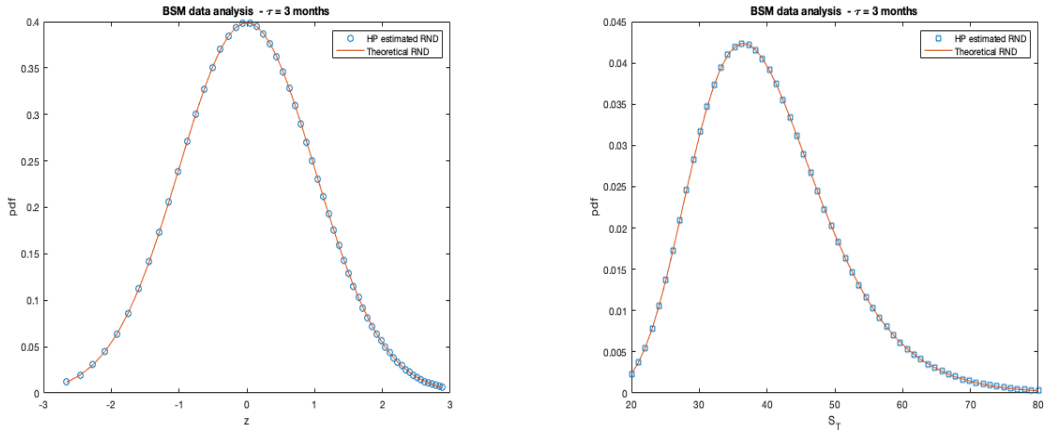

Figure 5.2 Theoretical and estimated RND functions from BSM data, without noise - Maturity: 3 months. On the left, for the standardized log-returns; and, on the right, for the underlying asset price.

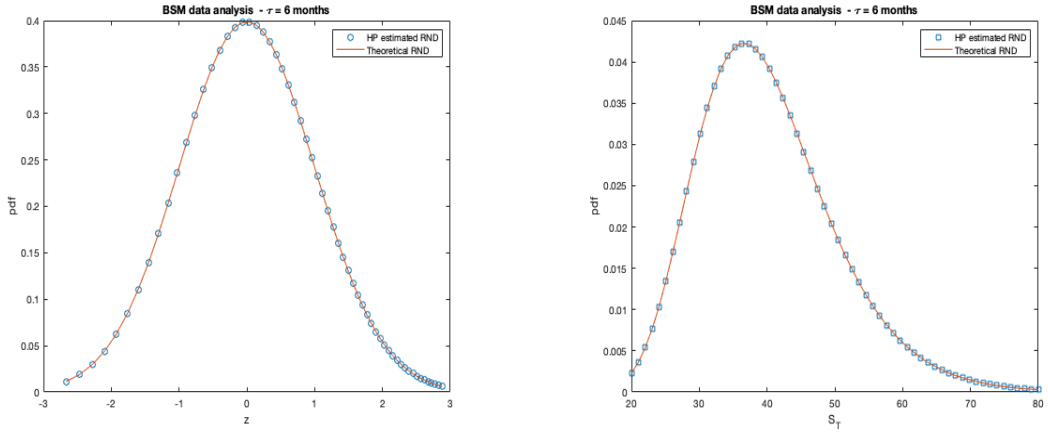

Figure 5.3 Theoretical and estimated RND functions from BSM data, without noise - Maturity: 6 months. On the left, for the standardized log-returns; and, on the right, for the underlying asset price.

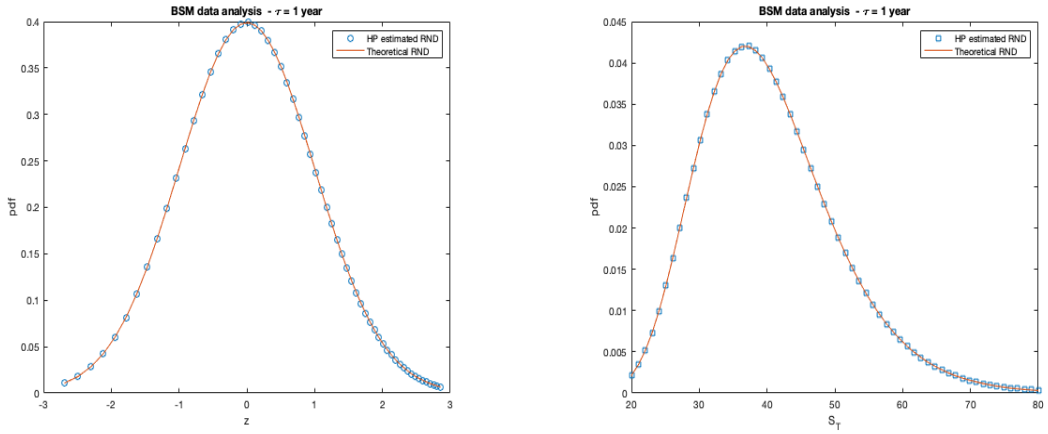

Figure 5.4 Theoretical and estimated RND functions from BSM data, without noise - Maturity: 1 year. On the left, for the standardized log-returns; and, on the right, for the underlying asset price.

5.1.2 Risk-Neutral Density estimation - with noise

Afterwards, we implemented a similar process to the one in the previous section, by adding a noise condition, (5.1).

Analysing Figures 5.5 - 5.8 we can observe that they are apparently slightly less well fitted as maturity increases. Estimated RNDs with maturities of 1 month and 3 months are better fitted, however we can detect a slight deviation from the theoretical density. This deviation from theoretical RND is apparently more prominent for estimated RNDs with maturities of 6 months and 1 year. In these cases, uncertainty valution associated to prices is also greater, given the fact that the noise condition apparently amplifies fluctuations. The estimates obtained with noise were run multiple times corroborating the theoretical intuition referred.

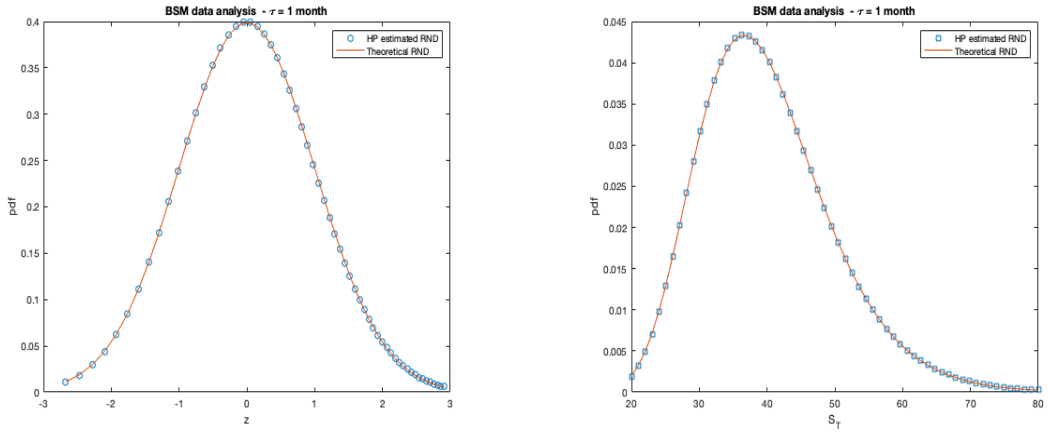

Figure 5.5 Theoretical and estimated RND functions from BSM data, with noise - Maturity: 1 month. On the left, for the standardized log-returns; and, on the right, for the underlying asset price.

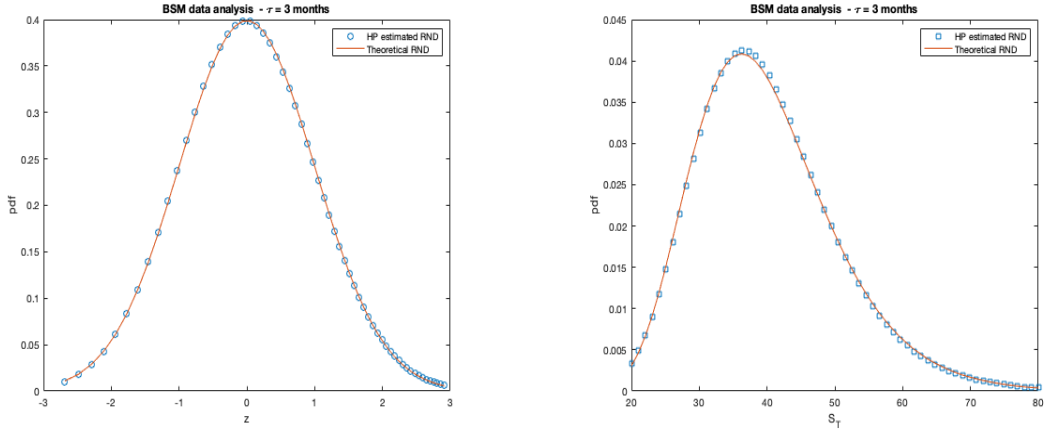

Figure 5.6 Theoretical and estimated RND functions from BSM data, with noise - Maturity: 3 months. On the left, for the standardized log-returns; and, on the right, for the underlying asset price.

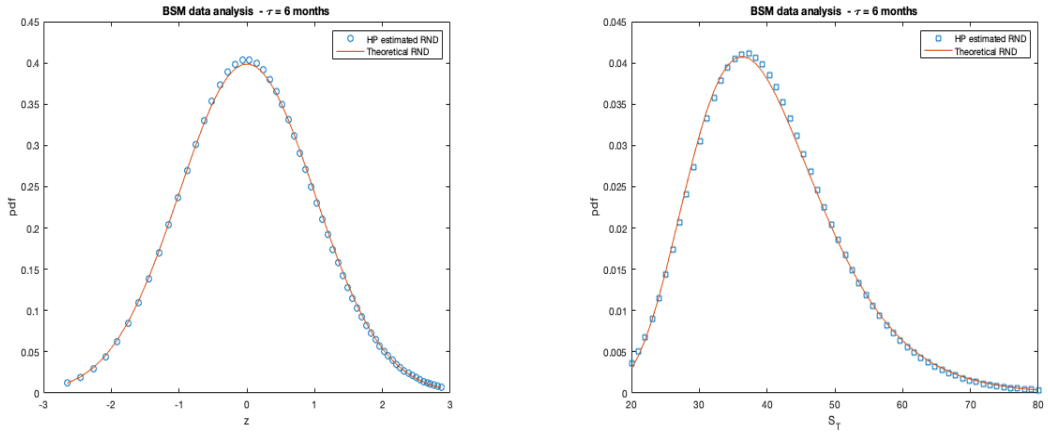

Figure 5.7 Theoretical and estimated RND functions from BSM data, with noise - Maturity: 6 months. On the left, for the standardized log-returns; and, on the right, for the underlying asset price.

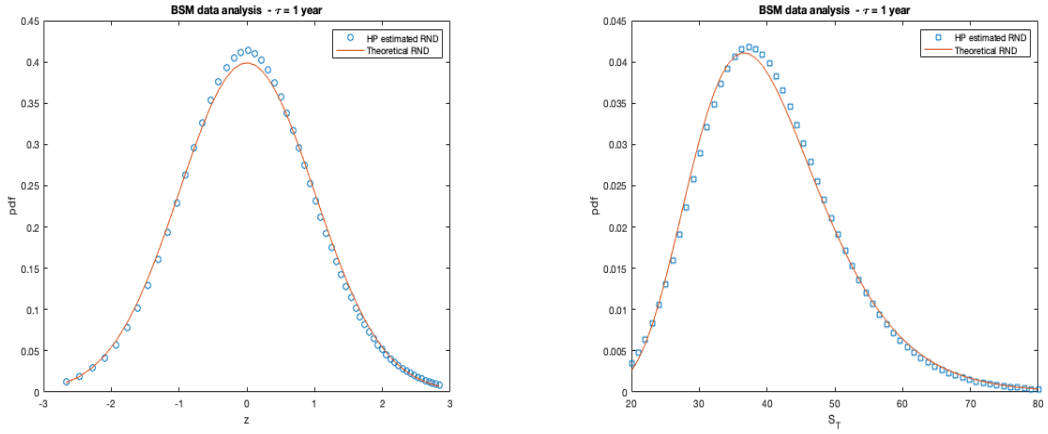

Figure 5.8 Theoretical and estimated RND functions from BSM data, with noise - Maturity: 1 year. On the left, for the standardized log-returns; and, on the right, for the underlying asset price.

5.1.3 Skewness and Kurtosis

Skewness and Kurtosis were obtained through formulas (4.21), which we implemented for BSM data with and without noise conditions. Skewness depicts the asymmetry of the distribution, while kurtosis captures the tail thickness of the distribution.

Table 5.1 aggregates skewness and kurtosis for RND estimated from BSM data. Skewness for RND with and without noise decreases as maturity increases. Skewness for RND with maturity of 1 month is positive, however very close to zero. Furthermore, we observe that kurtosis is approximately 3 when noise is not considered, and greater when noise is considered, increasing as maturity increases.

Skewness and Kurtosis for RND estimated from BSM data				
	Skewness		Kurtosis	
	Without Noise	Noise	Without Noise	Noise
1 month	0.0146	0.0132	3.00	3.00
3 months	-1.0117e^{-4}	-0.0428	3.0004	3.1308
6 months	-1.3838e^{-4}	-0.0962	3.0004	3.1592
1 year	-2.9070e^{-4}	-0.1402	3.0008	3.2619

Table 5.1 Skewness and kurtosis for RND estimated from the BSM data, with and without the noise condition.

5.2 Market data analysis

5.2.1 Data set description

Numerical results will be obtained from the S&P500 (SPX) index and two of the most shorted companies from April 2022; initially, we included five companies, however due to similar estimation performance, we present here only two of them. The companies are: MicroStrategy Incorporated (MSTR) - business intelligence, mobile software, and cloud-based services; Cassava Sciences (SAVA) – pharmaceutical company; Arch Resources (ARCH) – coal mining and processing company; Costco Wholesale Corporation (COST) – retail; and, Blink Charging (BLNK) – electric vehicles equipment and stations. We will only present results for S&P 500, ARCH and SAVA. This data was obtained from CBOE exchange market and relates to European style options for SPX and American style options for ARCH and SAVA. Furthermore, it should have been considered out-of-money options, in the sense that, they approximate better European style options. Therefore, this estimation is an approximation to the proper RND.

For each sample, we considered the arithmetic mean of bid and ask, call and put, option prices. The interest rate data is collected from the Central Bank of the Federal Reserve System of the United States of America, $r = 0.0233$.

Data was collected from August 15th, 2022 to August 25th, 2022, with four times to maturity - approximately one month, three months, six months and a year, depending on the day it was retrieved and the suitable correspondent maturity available. Using Madan and Milne's method, we obtained the estimated RND for each of these four maturities, for each underlying asset. Given that similar observations were obtained for these 10 days, we will only present and analyze the results for August 18th, 2022; similar conclusions can be derived for the remaining days. Table 5.2 summarizes the information on the data. We will consider call option prices in our estimations, however, in order to illustrate the behaviour of the call and put options we will also present the information related to puts.

Day	Underlying asset	S_t	Maturity	τ (days)
August 18th	SPX	4283,7402	September 16th, 2022	29
			November 18th, 2022	92
	ARCH	158,59	February 17th, 2023	183
	SAVA	26,25	August 18th, 2023	365

Table 5.2 Data set information for SPX, ARCH and SAVA.

5.2.2 RND estimation results

In this section, we present the results of implementing Madan and Milne (1994) [21] approach applied to SPX, Arch Resources and Cassava Sciences. Call and put option strikes, and respective open interests are presented in Figures 5.9, 5.10, 5.12, 5.13, 5.15, 5.16, 5.18, 5.19, 5.21, 5.22, 5.24, 5.25,5.27, 5.28, 5.31, 5.32, 5.33, 5.34, 5.36 and 5.37. RND estimatives were defined by considering log-returns and, thereafter, the underlying asset price, S_T. The latter is obtained through a change of variable. Open interest displays key information regarding the liquidity of an option. The larger the open interest, the easier it will be to trade that option at a reasonable bid-ask spread. In this context, it could be seen as a measure of the relevance of an option price, at a given strike.

SPX data

Observing Figures 5.9, 5.12 and 5.15 we can detect several peaks in call options open interest with maturities of 29, 92 and 183 days, with periodicity of 1000; also, we can see a concentration of values in the central zones of the plotting. Observing Figures 5.10, 5.13 and 5.16 we note that open interest values for put option are relevant for strikes until 5000. Finally, for the maturity of 365 days, open interest values for call options are relevant for strike prices approximately between 4000 and 5000; as for put options, they are relevant for strike prices approximately between 2500 and 5000.

Analysing Figure 5.11 the estimated RND is apparently well fitted against the Normal theoretical RND. As maturity increases, in Figures 5.14 and 5.17, we observe that the Hermite polynomial estimated RND seems to be leptokurtic, away from the Normal RND. This happens for both cases, log-returns and the underlying asset price. This may be due to greater uncertainty valuation of markets, as maturity increases.

We must note that there exists numerical problems when estimating the RND for $\tau = 365$ days. Also, as maturity increases, there is a greater valuation deviation from normality.

Finally, Table 5.3 presents information regarding the number of strikes for each maturity.

Underlying asset	Maturity	Number of Strikes
SPX	September 16th, 2022	810
	November 18th, 2022	507
	February 17th, 2023	140
	August 18th, 2023	128

Table 5.3 Number of Strikes - SPX.

$\tau = 29$ **days**

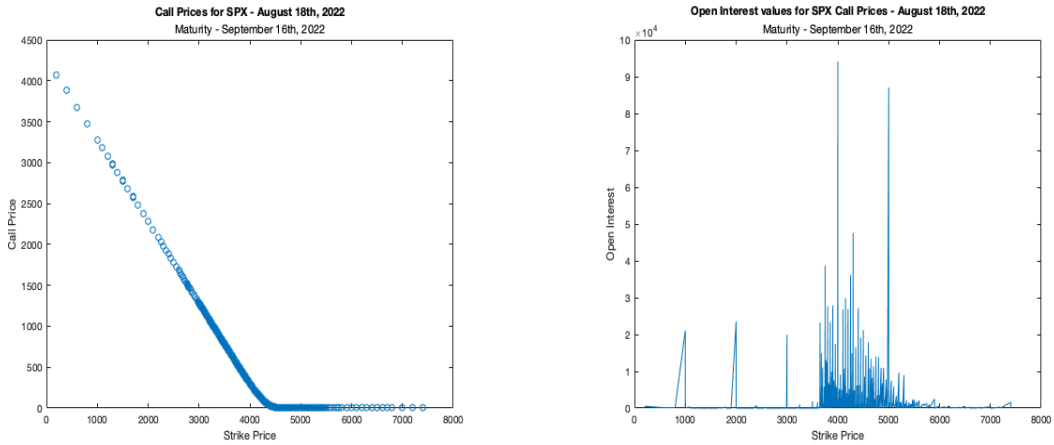

Figure 5.9 Call Options and Open interest from SPX data - Maturity: 29 days

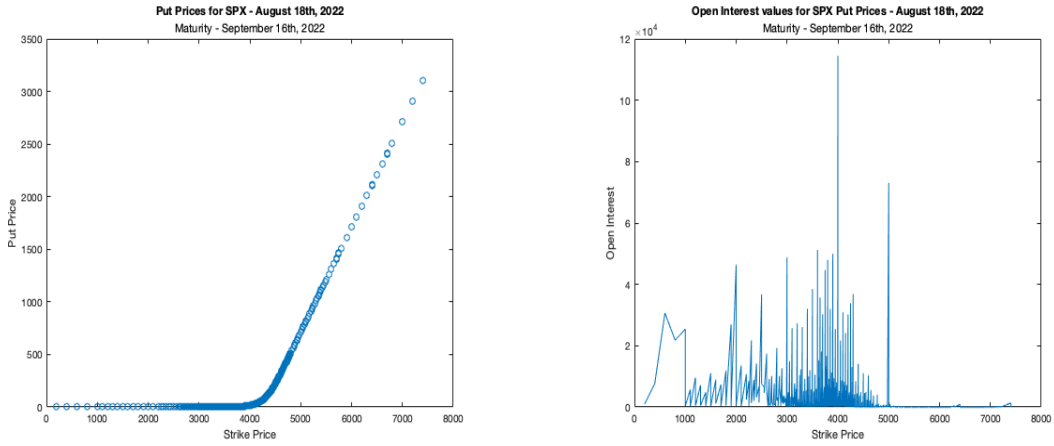

Figure 5.10 Put Options and Open interest from SPX data - Maturity: 29 days

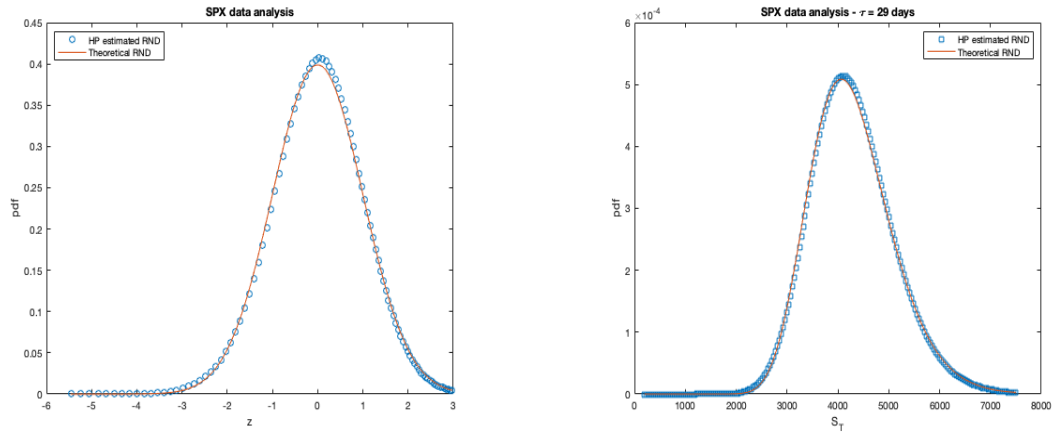

Figure 5.11 RND from SPX data - Maturity: 29 days.

$$\tau = 92 \textbf{ days}$$

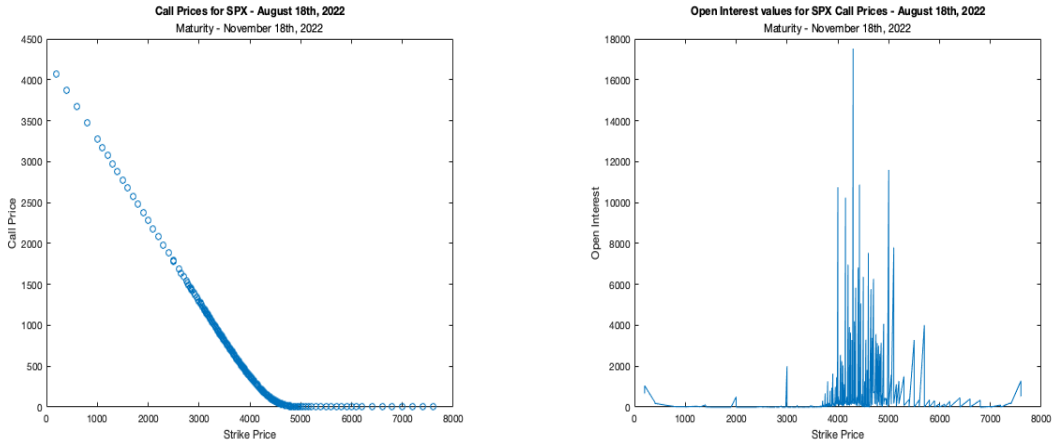

Figure 5.12 Call Options and Open interest from SPX data - Maturity: 92 days

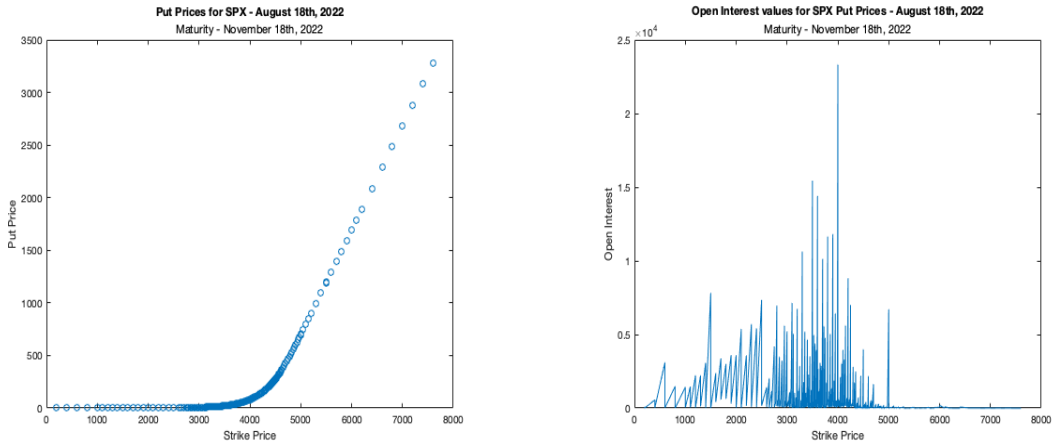

Figure 5.13 Put Options and Open interest from SPX data - Maturity: 92 days

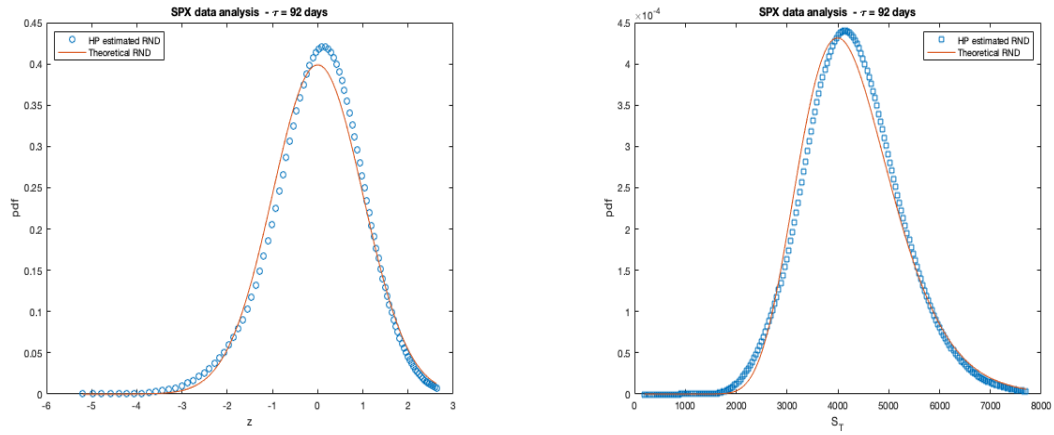

Figure 5.14 RND from SPX data - Maturity: 92 days.

$\tau = 183$ **days**

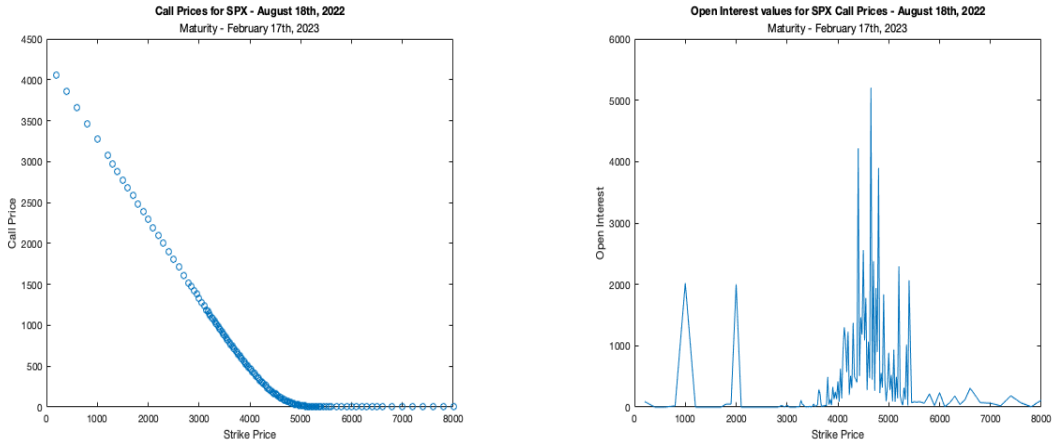

Figure 5.15 Call Options and Open interest from SPX data - Maturity: 183 days

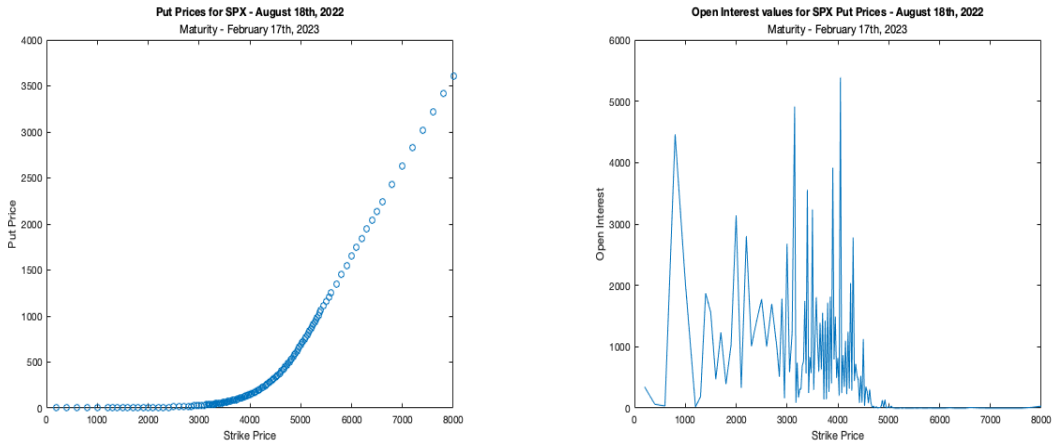

Figure 5.16 Put Options and Open interest from SPX data - Maturity: 183 days

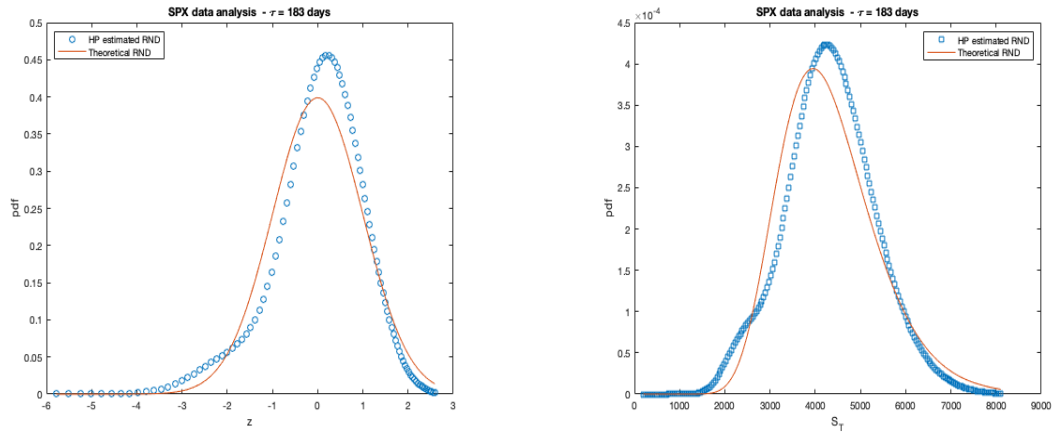

Figure 5.17 RND from SPX data - Maturity: 183 days.

$$\tau = 365 \textbf{ days}$$

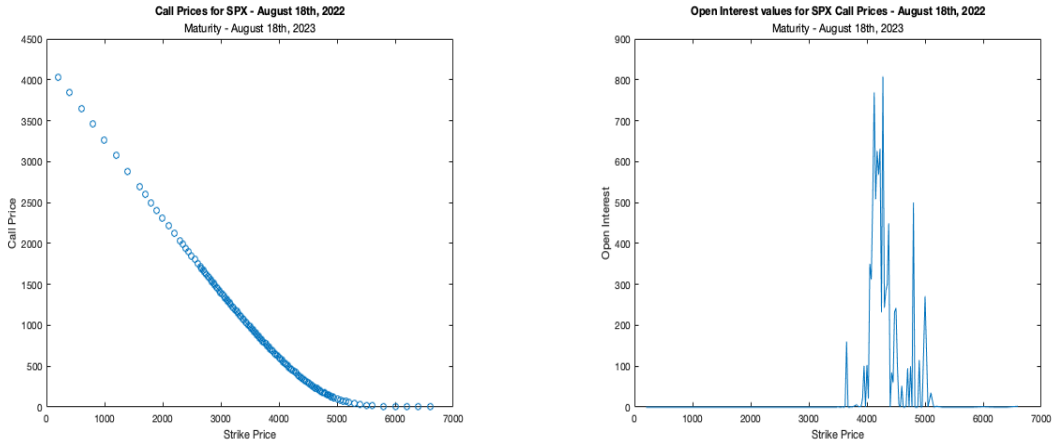

Figure 5.18 Call Options and Open interest from SPX data - Maturity: 365 days

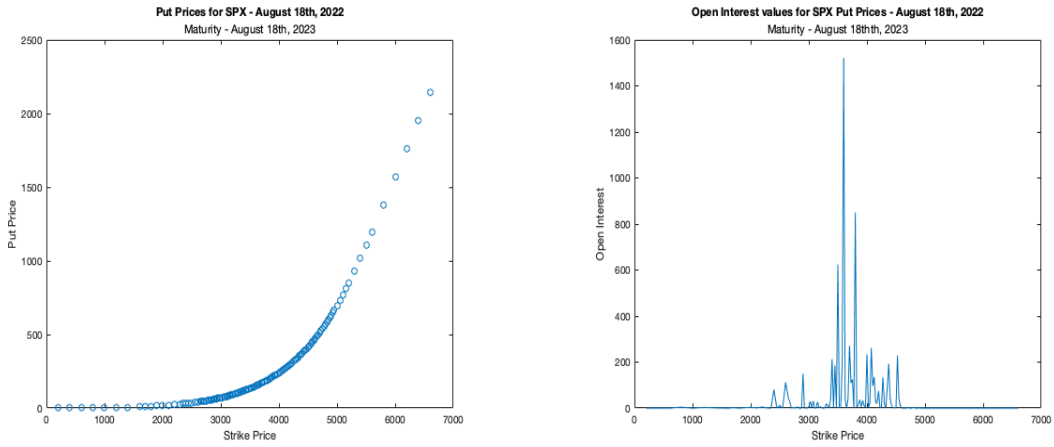

Figure 5.19 Put Options and Open interest from SPX data - Maturity: 365 days

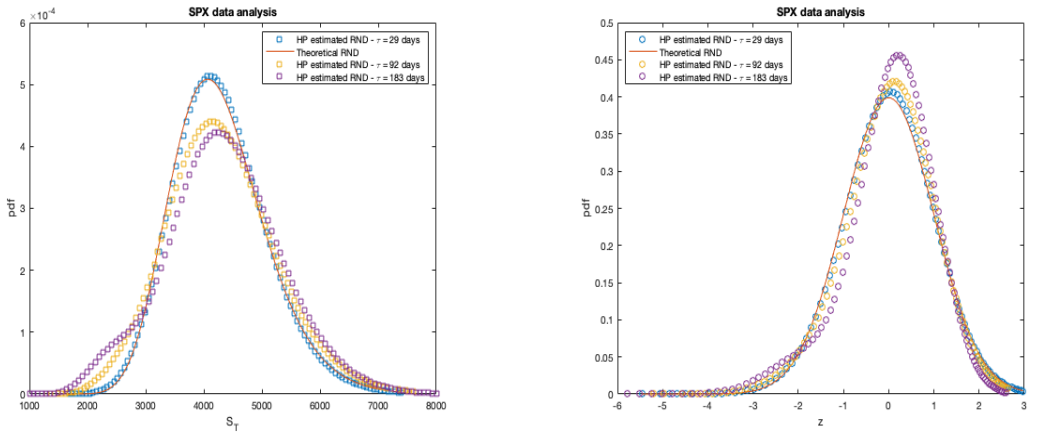

Figure 5.20 RND from SPX data.

ARCH data

Observing Figure 5.21 we can see that open interest increases, as strike prices increase; for Figure 5.22, we detect high open interest for $Strike = 65$, spiking again between 120 and 160, registering a downfall on 140. In Figure 5.24 there is a concentration for strikes between 140 and 180, with two major spikes. In addition, for Figure 5.27 we detected 4 major considerable strikes, and in Figure 5.28 the relevant strikes are from 60 to 140.

Analysing Figures 5.23, 5.26 and 5.29, we observe that the Hermite polynomial estimated RND moves away from the Normal RND. This happens for both log-returns and the underlying asset price. This occurs due to greater uncertainty of markets, as maturity increases. These results are more similar to the Normal density, than the ones we obtained before for SPX.

We didn't had any types of convergence problems estimating RND for ARCH.

Finally, Table 5.4 presents information regarding the number of strikes for each maturity.

Underlying asset	Maturity	Number of Strikes
ARCH	September 16th, 2022	28
	November 18th, 2022	34
	February 17th, 2023	33

Table 5.4 Number of Strikes - ARCH.

$$\tau = 29 \textbf{ days}$$

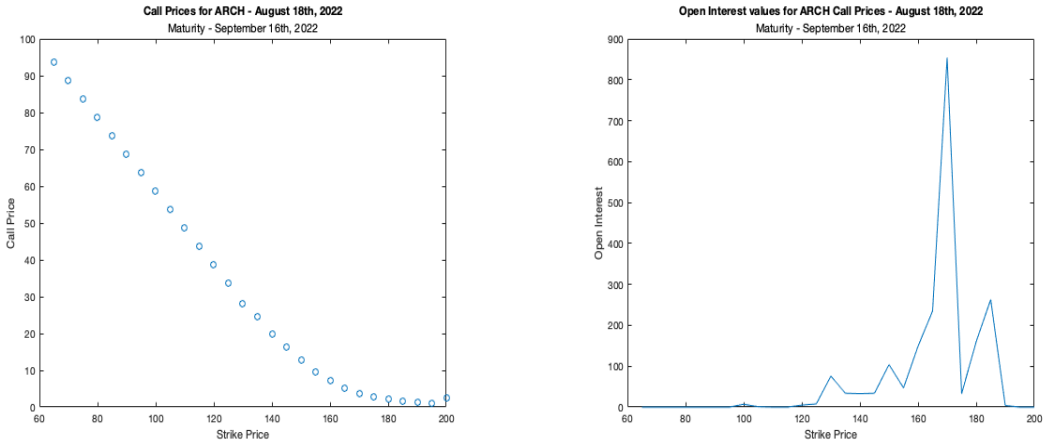

Figure 5.21 Call Options and Open interest from ARCH data - Maturity: 29 days

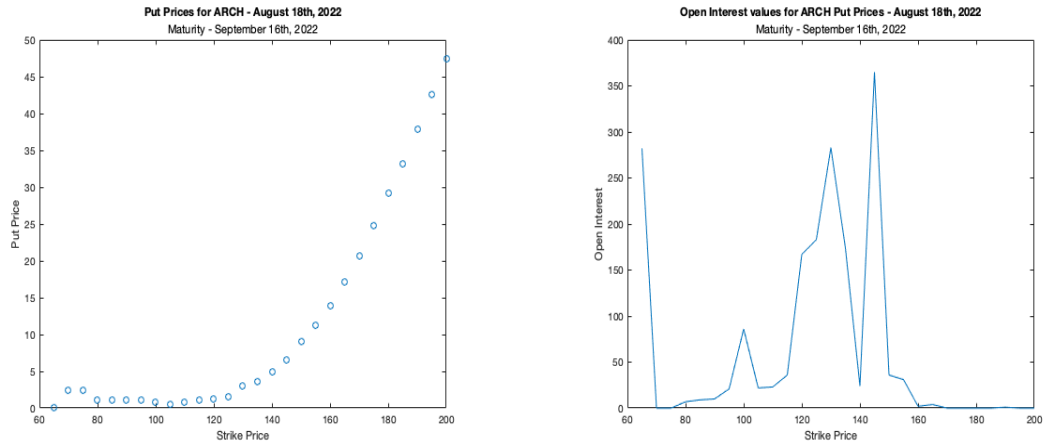

Figure 5.22 Put Options and Open interest from ARCH data - Maturity: 29 days

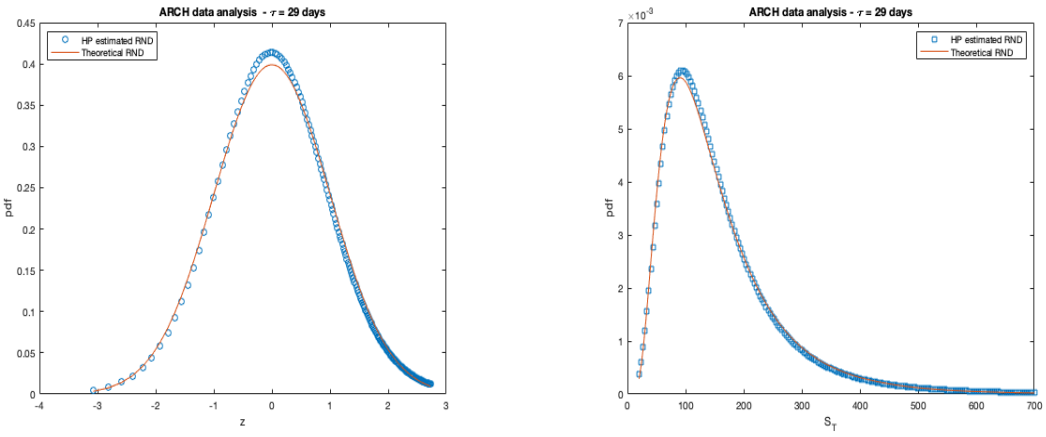

Figure 5.23 RND from ARCH data - Maturity: 29 days.

$\tau = 92$ **days**

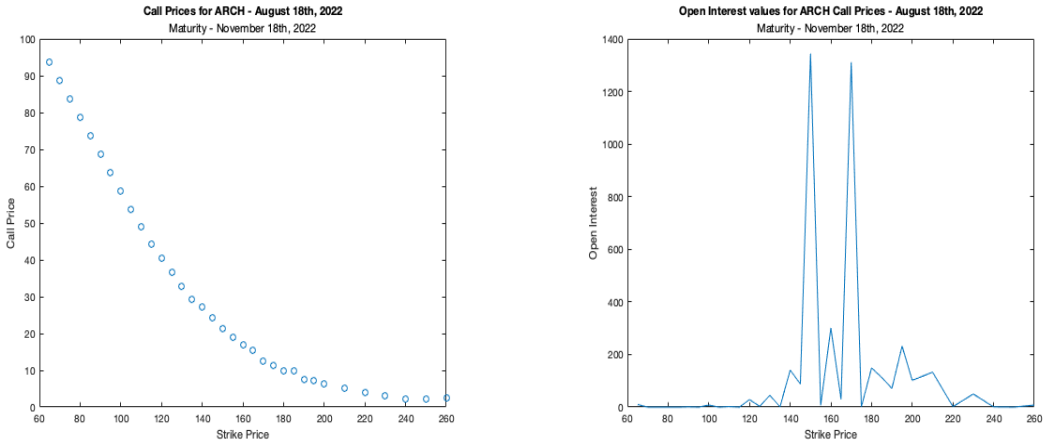

Figure 5.24 Call Options and Open interest from ARCH data - Maturity: 92 days

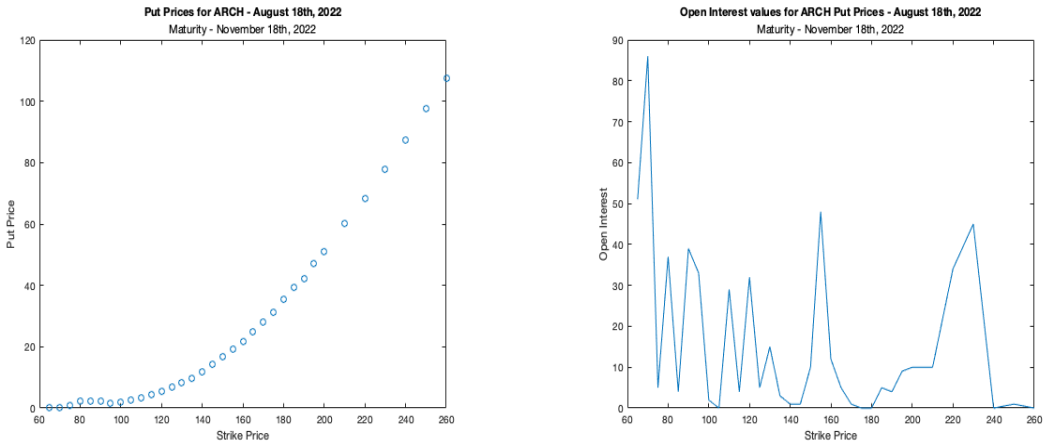

Figure 5.25 Put Options and Open interest from ARCH data - Maturity: 92 days

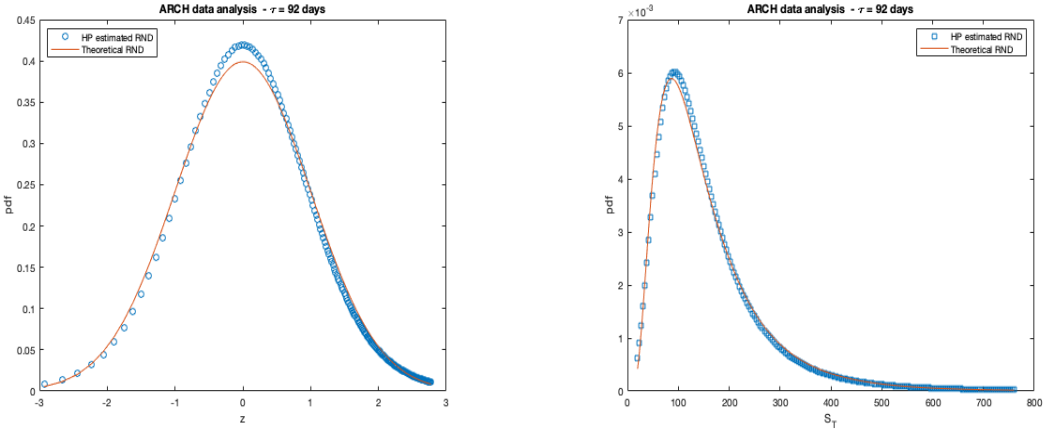

Figure 5.26 RND from ARCH data - Maturity: 92 days.

$$\tau = 183 \textbf{ days}$$

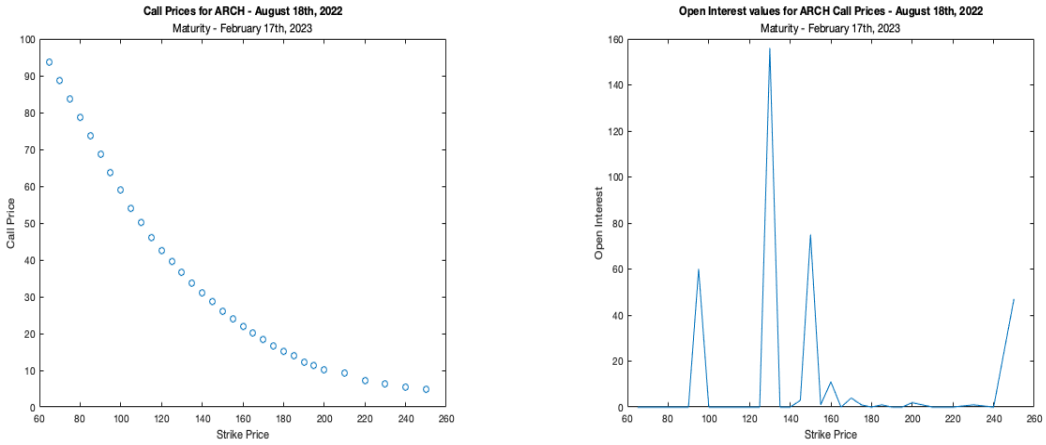

Figure 5.27 Call Options and Open interest from ARCH data - Maturity: 183 days

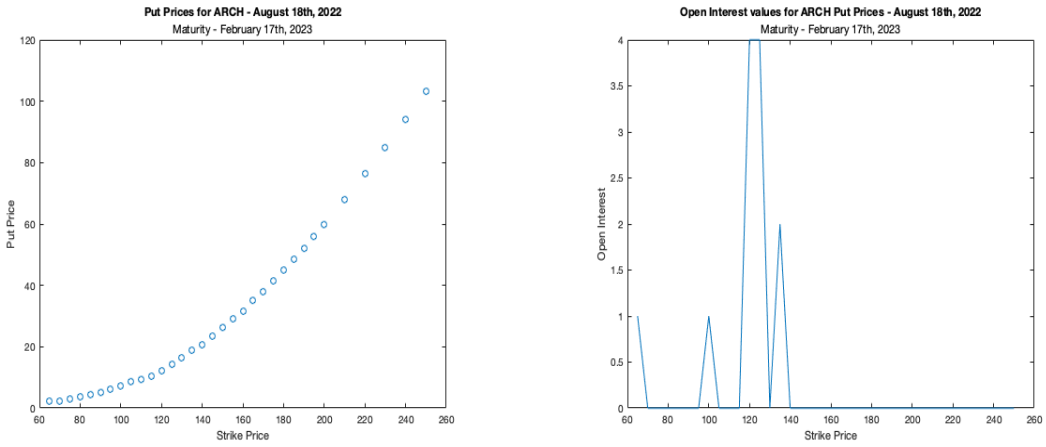

Figure 5.28 Put Options and Open interest from ARCH data - Maturity: 183 days

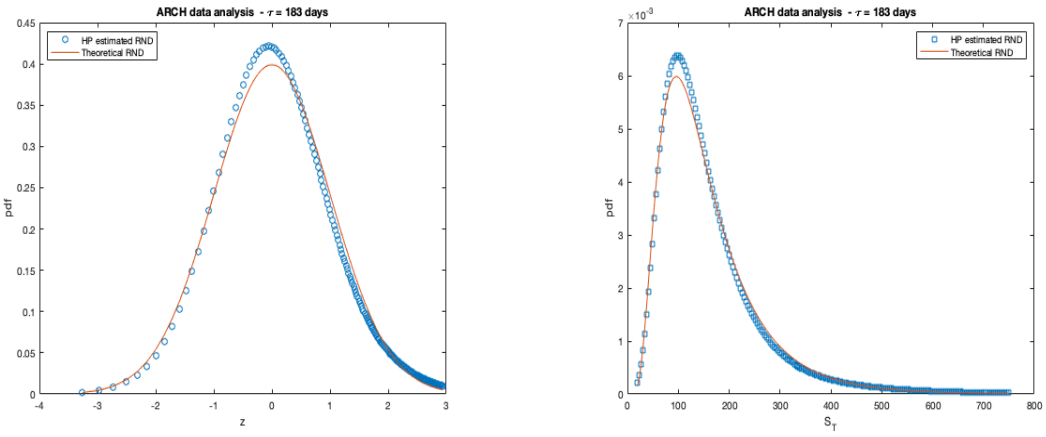

Figure 5.29 RND from ARCH data - Maturity: 183 days.

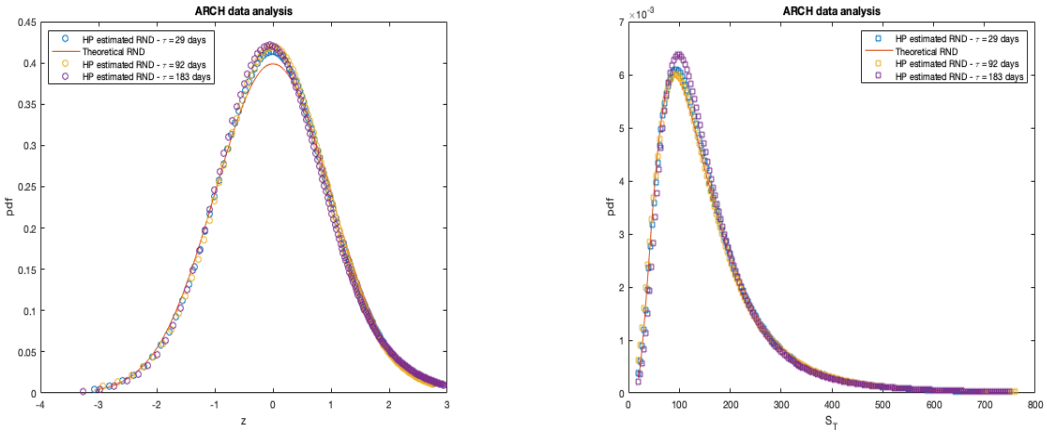

Figure 5.30 RND from ARCH data.

SAVA data

Observing Figures 5.31 and 5.36 we observed that the open interest assumes its maximum at $Strike = 25$. From Figures 5.32, 5.34 and 5.37 and 5.15 we see that put options are more desired than call options for this company.

In addition, we detected lack of convergence for the RND with $\tau = 29$ days, hence, the corresponding plot is not shown. Even with a selection of strikes with higher open interest, convergence was not achieved. On the other hand, analysing Figures 5.35 and 5.38, we observe that the Hermite polynomial estimated RND moves away from the Normal RND. This happens for both standardized log-returns and the underlying asset price. This maybe occurs due to greater uncertainty of markets, as maturity increases.

Finally, Table 5.5 presents information regarding the number of strikes for each maturity.

Underlying asset	Maturity	Number of Strikes
ARCH	September 16th, 2022	16
	November 18th, 2022	17
	February 17th, 2023	16

Table 5.5 Number of Strikes - SAVA.

$$\tau = 29 \textbf{ days}$$

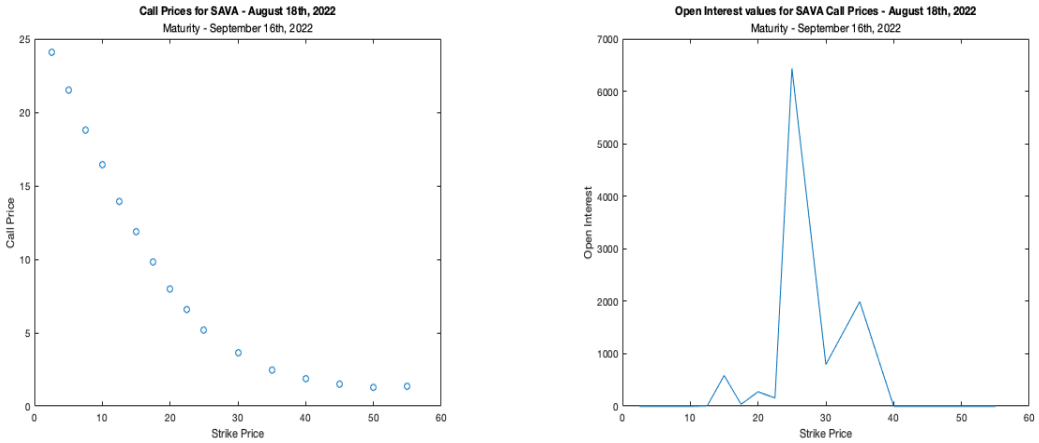

Figure 5.31 Call Options and Open interest from SAVA data - Maturity: 29 days

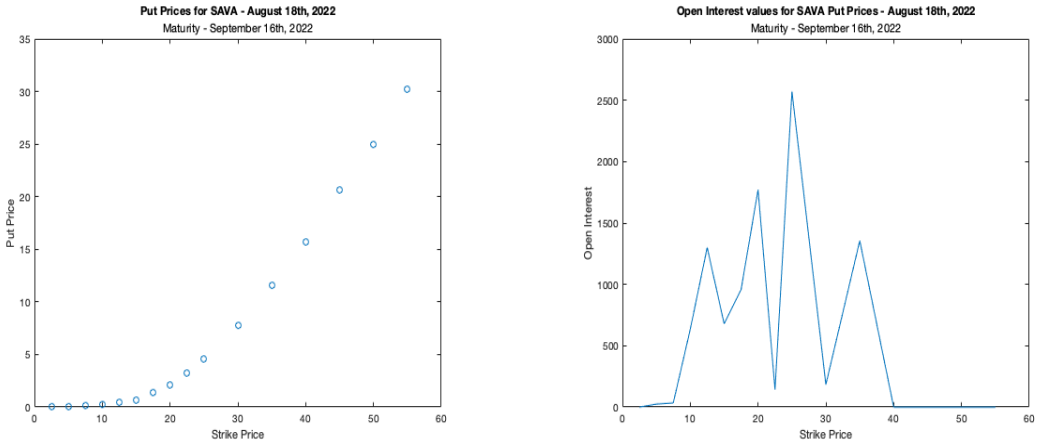

Figure 5.32 Put Options and Open interest from SAVA data - Maturity: 29 days

$\tau = 92$ **days**

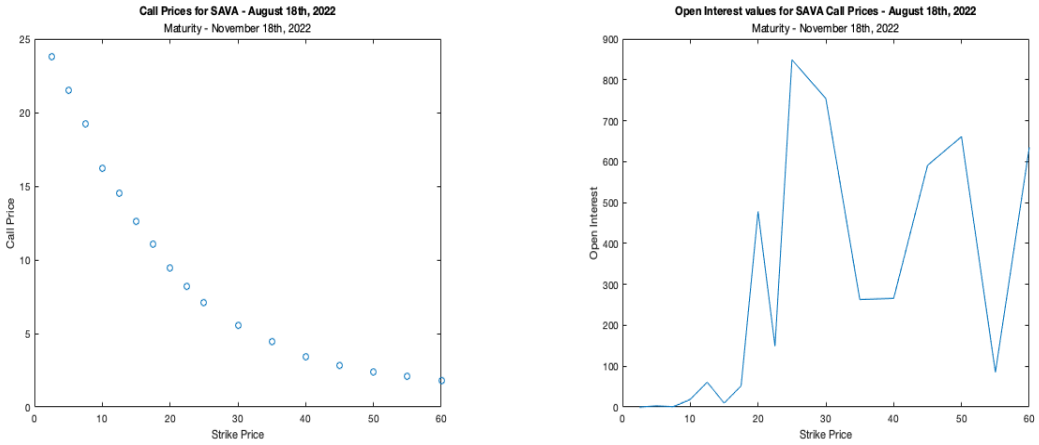

Figure 5.33 Call Options and Open interest from SAVA data - Maturity: 92 days

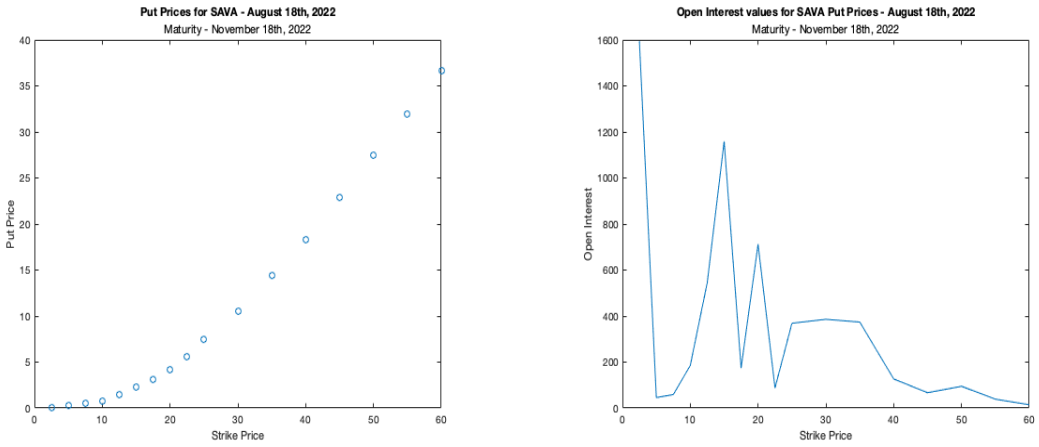

Figure 5.34 Put Options and Open interest from SAVA data - Maturity: 92 days

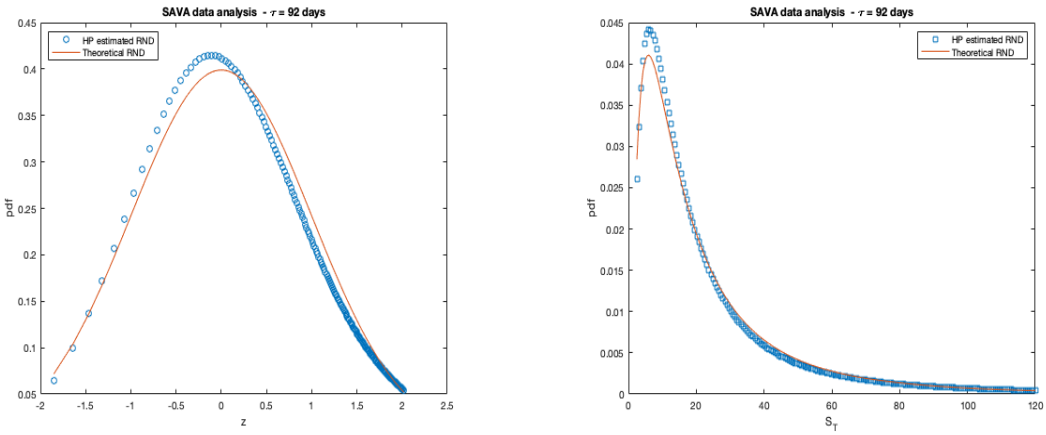

Figure 5.35 RND from SAVA data - Maturity: 92 days.

$$\tau = 183 \textbf{ days}$$

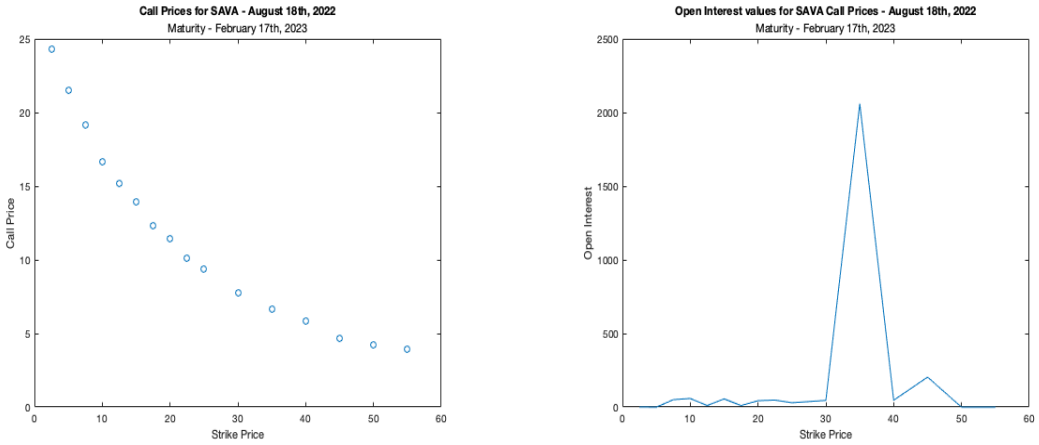

Figure 5.36 Call Options and Open interest from SAVA data - Maturity: 183 days

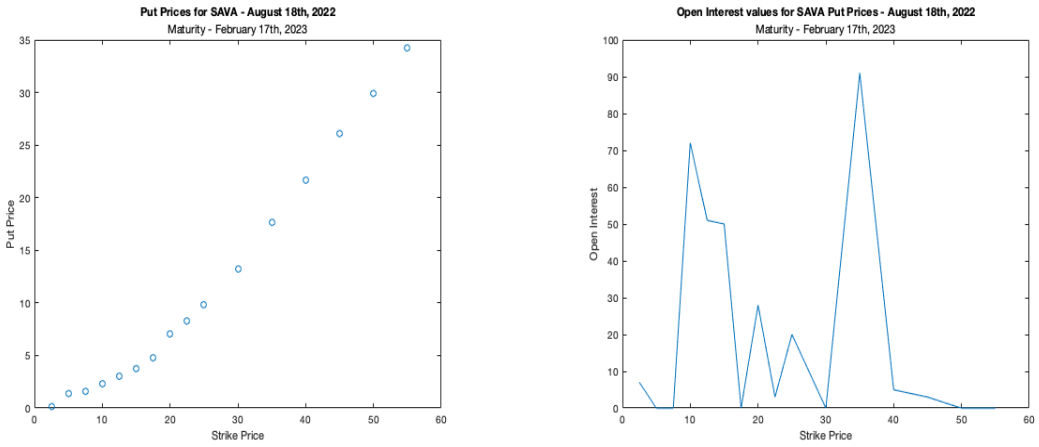

Figure 5.37 Put Options and Open interest from SAVA data - Maturity: 183 days

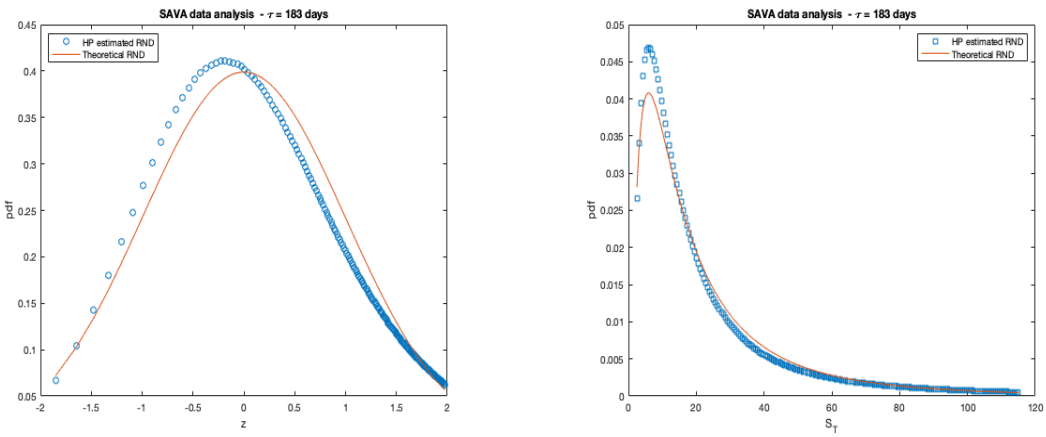

Figure 5.38 RND from SAVA data - Maturity: 183 days.

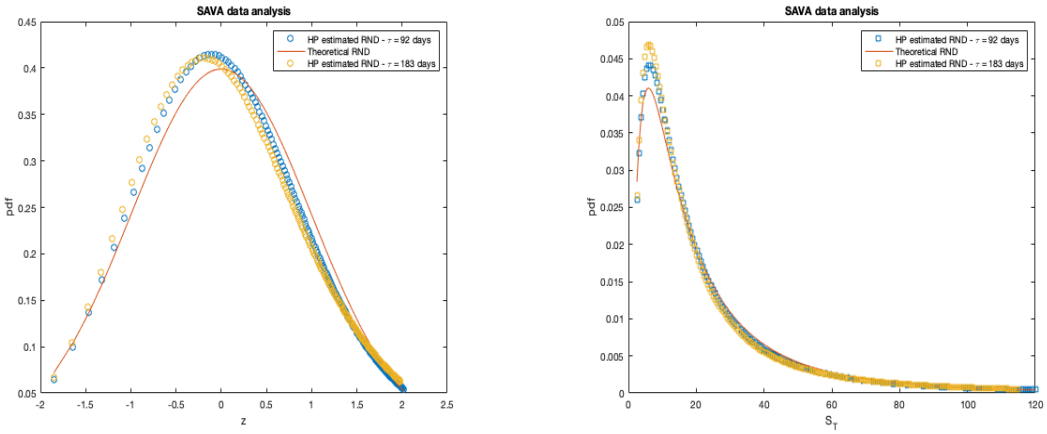

Figure 5.39 RND from SAVA data.

5.2.3 Skewness and Kurtosis

Considering SPX, we observe that skewness is negative, which means that left tail is heavier, and therefore the probability of strongly negative returns is higher than for positive returns. There seems to be a relationship between skewness and kurtosis. When skewness goes up, in absolute value, so does the kurtosis.

Considering ARCH, we detect that skewness is only negative for $\tau = 92$ days, hence similar conclusions to the previous paragraph arise. On the other hand, for $\tau = 29$ and $\tau = 183$ days, skewness is positive.

Negative skewness may indicate that investors are somewhat pessimistic.

The kurtosis coefficient is greater than 3 for every underlying asset (leptokurtic distribution), which can suggest there is a large probability of extreme values occurrence on both sides.

Day	Underlying asset	Maturity	Skewness	Kurtosis
August 18th	SPX	September 16th, 2022	-0,0975	3,1495
		November 18th, 2022	-0,3172	3,3490
		February 17th, 2023	-0,6928	3,8431
	ARCH	September 16th, 2022	0,0273	3,2908
		November 18th, 2022	-0,0037	3,4119
		February 17th, 2023	0,1570	3,4266
	SAVA	November 18th, 2022	0,2603	3,2536
		February 17th, 2023	0,4243	3,0745

Table 5.6 Skewness and Kurtosis for SPX, ARCH and SAVA.

Chapter 6

Conclusion

Derivative markets provide investors with a rich source of information for evaluating market sentiment, concerning the underlying asset evolution. Options prices are an efficient tool to express market perceptions of underlying asset prices in the future. From this data it is possible to estimate the RND. In this dissertation, we used Hermite polynomials as a semi-nonparametric approach on the estimation of the RND. This expansion method was employed on theoretical prices obtained through the BSM model and thereafter applied to market data from the S&P 500 index and two short-sold companies.

Through graphical analysis, we observe that simulated data from the BSM model the obtained estimates, when a noise condition is introduced, only deviates from the theoretical densities for longer maturities; also, as skewness decreases, kurtosis increases. Once applied to either BSM model data and market data, Hermite polynomials seem to behave well to obtain proper RND estimates. We concluded that if maturity increases, apparently the quality of the estimation decreases. Also, if there is a small number of strikes, this difficults the estimation process.

In addition, we dealt with American options, instead of European options, therefore, the obtained results can be seen as an approximate estimative to the proper estimated density.

Furthermore, open interest is a relevant source of information regarding the relevance of an option contract. High open interest leads to higher relevance of the option contract. Moreover, it is also used as a criteria for strike selection.

Finally, for $\tau = 92$ days, when skewness is negative we can conclude that investors are pessimistic about market behaviour of the S&P 500 index and of ARCH. On the other hand, for $\tau = 29$ days and $\tau = 183$ days, investors seem to be more confident about SAVA and ARCH. Since we obtained leptokurtic distributions we may expect higher risk.

Overall, Hermite polynomials seem to be an effective semi-nonparametric approach to retrieve proper risk-neutral densities estimates.

Bibliography

[1] Abken, P. A., Madan, D. B., and Ramamurtie, S. (1996). Estimation of risk-neutral and statistical densities by hermite polynomial approximation: with an application to eurodollar futures options. Working paper.

[2] Aït-Sahalia, Y. and Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *The journal of finance*, 53(2):499–547.

[3] Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of business*, pages 621–651.

[4] Corrado, C. J. and Su, T. (1996a). Skewness and kurtosis in s&p 500 index returns implied by option prices. *Journal of Financial research*, 19(2):175–192.

[5] Corrado, C. J. and Su, T. (1996b). S&p 500 index option tests of jarrow and rudd's approximate option valuation formula. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 16(6):611–629.

[6] Coutant, S. (1999). Implied risk aversion in options prices using hermite polynomials ', estimating and interpreting probability density functions. *Bank for International Settlements*.

[7] Coutant, S., Jondeau, E., and Rockinger, M. (2001). Reading pibor futures options smiles: The 1997 snap election. *Journal of Banking & Finance*, 25(11):1957–1987.

[8] Cox, J. C. and Ross, S. A. (1976). The valuation of options for alternative stochastic processes. *Journal of financial economics*, 3(1-2):145–166.

[9] Figlewski, S. (2018). Risk neutral densities: A review. *Annual Review of Financial Economics*, 10.

[10] Guasoni, P. et al. (2004). Estimating state price densities by hermite polynomials: theory and application to the italian derivatives market. *Banca d'Italia*, 507.

[11] Hentschel, L. (2003). Errors in implied volatility estimation. *Journal of Financial and Quantitative analysis*, 38(4):779–810.

[12] Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *The journal of finance*, 42(2):281–300.

[13] Hull, J. and White, A. (1988). An overview of contingent claims pricing. *Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration*, 5(3):55–61.

[14] Ingersoll Jr, J. E. (1976). A theoretical and empirical investigation of the dual purpose funds: An application of contingent-claims analysis. *Journal of Financial Economics*, 3(1-2):83–123.

[15] Jackwerth, J. C. (2000). Recovering risk aversion from option prices and realized returns. *The Review of Financial Studies*, 13(2):433–451.

[16] Jarrow, R. and Rudd, A. (1982). Approximate option valuation for arbitrary stochastic processes. *Journal of financial Economics*, 10(3):347–369.

[17] Jondeau, E., Poon, S.-H., and Rockinger, M. (2007). *Financial modeling under non-Gaussian distributions*. Springer Science & Business Media.

[18] Jondeau, E. and Rockinger, M. (2001). Gram–charlier densities. *Journal of Economic Dynamics and Control*, 25(10):1457–1483.

[19] Lin, S.-H., Huang, H.-H., and Li, S.-H. (2015). Option pricing under truncated gram–charlier expansion. *The North American Journal of Economics and Finance*, 32:77–97.

[20] MacBeth, J. D. and Merville, L. J. (1980). Tests of the black-scholes and cox call option valuation models. *The Journal of Finance*, 35(2):285–301.

[21] Madan, D. B. and Milne, F. (1994). Contingent claims valued and hedged by pricing and investing in a basis. *Mathematical Finance*, 4(3):223–245.

[22] Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of economics and management science*, pages 141–183.

[23] Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of finance*, 29(2):449–470.

[24] Rompolis, L. S. and Tzavalis, E. (2007). Retrieving risk neutral densities based on risk neutral moments through a gram–charlier series expansion. *Mathematical and Computer Modelling*, 46(1-2):225–234.

[25] Rompolis, L. S. and Tzavalis, E. (2008). Recovering risk neutral densities from option prices: A new approach. *Journal of Financial and Quantitative Analysis*, 43(4):1037–1053.

[26] Rubinstein, M. (1985). Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active cboe option classes from august 23, 1976 through august 31, 1978. *The Journal of Finance*, 40(2):455–480.

[27] Rubinstein, M. (1999). *Rubinstein on derivatives*. Risk Books.

[28] Scholes, M. and Black, F. (1973). The pricing of options and corporate liabilities. *Journal of political Economy*, 81(3):637–654.