The Option Implied Risk-Neutral Distribution

A comparison study for obtaining risk-neutral densities from observed option prices

Author:

Jonatan Isakas

Supervisor:

Prof. Dr. Bertrand Melenberg

Master Thesis

M.Sc. Finance

Tilburg School of Economics and Management

Department of Finance

SNR: 2031855

ANR: 856826

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BM	Brownian motion
JD	Jump diffusion
KR	Kernel regression model
MM	Mixture model
NN	Neural network model
ReLu	Rectified linear unit function
RND	Risk-neutral density
RNDs	Risk-neutral densities
S&P 500	Standard and Poor's 500
SPX	Ticker symbol for the S&P 500 index
Std. Dev.	Standard Deviation
SV	Stochastic volatility
SVJD	Stochastic volatility and jump diffusion
Tanh	Hyperbolic tangent function

Statement of Originality

This document is written by student Jonatan Isakas who declares to take full responsibility for the contents of this document. I declare that the text and the work presented in this document are original and that no source other than those mentioned in the text and its references have been used in creating it.

I am aware that the violation of this regulation will lead to failure of the thesis.

Abstract

This thesis examines and compares the performance of the parametric mixture model and the two nonparametric, kernel regression and neural network models, with the intention of estimating option implied risk-neutral densities. Pseudo option prices are generated using Monte Carlo simulations for four different data-generation processes over three maturities. The simulated option prices are further used to estimate the risk-neutral densities of the underlying asset using the three different models. The models' performance is then evaluated by comparing the estimated risk-neutral densities with the theoretical densities corresponding to the known underlying parameters used in data-generation processes. The results show that the nonparametric models outperform the parametric model, while the neural network provides more consistent results over the majority of test cases.

1. Introduction

Option pricing theory is widely accepted as a tool for assessing valuable information of future expectations and risk preferences in the financial markets. The risk-neutral density (RND) is a fundamental concept in asset pricing theory and can be applied to many different financial frameworks. The information incorporated in option implied RNDs can be used for determining future market beliefs, pricing of illiquid derivatives, hedging purposes and monetary policies (Aıt-Sahalia & Lo, 2000; Jackwerth, 2004). The amount of information that can be analyzed from intraday data has increased tremendously in recent years, allowing researchers to find new, heavily data driven approaches or further develop existing methods to estimate option implied RNDs.

RND estimation is a fairly new research topic within the field of asset pricing and risk-neutral valuation theory. The concept of risk-neutral measures for pricing derivatives was first introduced by Black and Scholes (1973) and Merton (1973) with their novel, closed-form solution for pricing option contracts. Their model demonstrated that the price of a European option is equivalent to the expected payoff of the underlying asset and a risk-free bond, discounted by the risk-free. Based on Black, Scholes and Merton's revolutionizing theoretical framework, Cox, Ross, and Rubinstein (1979) introduced their binomial tree model for pricing derivatives in a complete and arbitrage free market. Based on the concept of risk-neutral valuation, the model introduced a convenient way to replicate future payoffs of the underlying stock in discrete time. However, these models require rigid assumptions regarding the underlying process of the asset, thus Rubinstein (1985) proposed a new methodology of obtaining implied parameters, such as the RND and implied volatility, from observed option contracts in the market, and subsequently utilizing these observed parameters to estimate option prices.

In a complete market, where every contingent claim can be replicated, Ross (1976) demonstrated how one can recover the complete RND from a set of European option prices. Breeden and Litzenberger (1978) continued the work of Ross and showed how the RND can be obtained from option prices with a continuum of strike prices, by taking the second partial derivative of the call option price with respect to the strike price. Since the findings of Breeden and Litzenberger, numerous methods of estimating the option implied RND have been proposed. The different methods can be classified into two subcategories: parametric and nonparametric, both with their respective advantages and drawbacks. Here, we will briefly cover some of the more prominent methods proposed in earlier literature.

For the parametric methods, Jarrow and Rudd (1982) successfully presented how one can retrieve the full option implied RND using a parametric Edgeworth expansion, where the authors adjusted the well-known Black-Scholes model to account for different statistical moments in the return distribution, such as skewness and kurtosis. Coutant, Jondeau, and Rockinger (2001) investigated the performance of an expansion method based on Hermite polynomials, where they could approximate option prices based on skewness and kurtosis and subsequently obtain the RND. Rompolis and Tzavalis (2007) introduced an expansion method based on an exponential form of a Gram-Charlier series expansion, which allowed the authors to constrain the RND and further incorporate statistical moments, similar to Jarrow and Rudd (1982), to account for known return probability distribution features such as skewness and kurtosis. Similarly to the method used in this thesis, Ritchey (1990), Bahra (1997), Söderlind and Svensson (1997) and Gemmill and Saflekos (2000) utilized parametric Gaussian mixture models to estimate the RND, where the authors assume the asset return probability distribution can be modeled with a combination of normal densities, fitted to the observed option data.

Shimko (1993) was one of the first to introduce the nonparametric approach, where one first transforms the option prices into implied volatilities, interpolate the implied volatilities and subsequently inverse the fitted implied volatilities back to option prices. The RND is then derived by taking the second partial derivative of the option price with respect to the strike price, proposed by Breeden and Litzenberger (1978). This simple, but yet effective approach paved the way for new intriguing nonparametric approaches to be used for estimating the RND. Aït-Sahalia and Lo (1998) used a kernel regression model to estimate the state-price density across five independent variables: stock price, strike price, time to maturity, interest rate and dividend yield.[1] However, the authors argue that the approach is requires a substantial amounts of data and only becomes feasible when combining multiple days of data. Pritsker (1998) implemented a kernel method for estimating the RND obtained from options derived from interest rates. The author found that the kernel estimator’s choice of optimal kernel bandwidth is sensitive to persistence in the US interest rates, but quite insensitive to the frequency at which the data is sampled. Thus, Pritsker argue that the use of kernel methods might be problematic due to the fact that asymptotic distributions do not depend on persistence, while the sample might do. Rockinger and Jondeau (2002) utilized a maximum entropy method for obtaining the RND, by characterizing the skewness and kurtosis features and subsequently fitting the option prices and the estimated RNDs. Yatchew and Härdle (2006) utilized nonparametric regression to estimate a constrained call price function over strike prices. The authors incorporated monotonicity and convexity constraints to ensure that the call function is a decreasing convex function over strike prices and

[1] See section 2.2.2 for a further explanation on state-prices.

further ensures that the stat price density is a valid density function, or more specifically that the function is non-negative and integrate to one. A more recent study by Feng and Dang (2016), where the authors presented a constrained support vector machine as a regression technique for interpolating the bid-ask spread and subsequently estimating the RNDs incorporating the Breeden-Litzenberger approach.

The complexity and variety of the option implied RND makes it practically impossible to derive a single method that can perfectly estimate a profound and well-behaved RND for every asset, at different points in time. Thus, several comparison studies have been made to evaluate the most prominent methods for obtaining the RND. However, the issue of comparing the methods is that the theoretical or true RND cannot be observed from market data, since it is impossible to accurately determine the underlying asset return distribution. Cooper (1999) was one of the first to effectively study the performance of these different approaches by generating pseudo option prices using Monte Carlo simulations, and thus, he remained in full control over the underlying distribution and could adequately evaluate the performance of the different methods.

Since the novel approach of simulating option prices, numerous studies have been conducted to compare different methods and their ability to estimate the option implied RND. Jondeau and Rockinger (2000) compared three different methods, a parametric, semi-parametric and nonparametric, with varying results. The nonparametric approach provided a good fit, but was unable to provide as much information as the parametric. Bu and Hadri (2007) evaluated the performance of the nonparametric method of smoothing the implied volatility and the semi-nonparametric confluent hypergeometric method, based on pseudo-priced options using the stochastic volatility model proposed by Heston (1993). The authors found that the semi-nonparametric method outperformed the nonparametric method both in terms of accuracy and stability. Bouden (2007) evaluated the performance of four parametric and two nonparametric approaches. The author compared the methods based on their ability to replicate the true density, pricing of options and forecasting performance and found that the nonparametric methods were superior. Grith, Härdle, and Schienle (2012) compared the performance between kernel-based methods, where two of the methods used local features (local polynomials), while a third method utilized global curve fitting. The authors found that all three approaches performed similarly. At the same time, they pointed out that the local polynomial methods are highly sensitive to bandwidth selection, similar to the findings of Pritsker (1998). Lai (2014) compared three different nonparametric methods. The three different methods used were kernel regression, spline interpolation and a neural network model, where the kernel regression proved to be the best approach. On the other hand, the author concludes that the estimation error depends on both the nature of the option data and that specific models are better suited for different set of options. Even more recent work, Celis, Liang, Lemmens, Tempere, and Cuyt (2015) investigated the performance of the parametric mixture method and the two nonparametric, interpolating methods: smoothing of the implied volatility surface and the rational interval interpolation. The authors found that the rational interval interpolation method proved to be more robust and more adequate for estimating tail distributions, especially for longer maturities. Santos and Guerra (2015), examined the performance of four different methods, based on parametric, semi-parametric and nonparametric approaches to estimate the RND. They found that the best performing methods were surprisingly the two parametric methods: Hypergeometric function and Mixture Lognormal method. Hence, we recognize that earlier literature is divided regarding the preferred approach for obtaining the RND. For a comprehensive overview of the earlier research and theoretical framework within the field of risk-neutral distribution and option pricing, see Jackwerth (1999, 2004) or the more recent work by Figlewski (2018).

The main purpose of this thesis is to investigate and examine how one can efficiently estimate and measure the performance of option implied RNDs, and further determine which of the selected models is best suited for the purpose of estimating them. In the next segment, *section 2,* we will cover the theoretical framework behind risk-neutral valuation and densities, option pricing and how certain market aspects influence how market agents price derivatives. In *section 3*, we review the methodology and the different approaches for estimating the option implied RND and in detail define the three models: the mixture model, the kernel regression model and the neural network model. Based on the work of Cooper (1999), *section 4* introduces the four known option pricing models used in this thesis, namely the famous Black and Scholes (1973) model, the stochastic volatility model introduced by Heston (1993), the jump diffusion model derived by Merton (1976) and lastly the stochastic volatility and jump diffusion model proposed by Bakshi, Cao, and Chen (1997). The option pricing models are subsequently used to evaluate the theoretical or 'true' RNDs and simulate corresponding option prices, which in turn are used to train our models. *Section 5* presents the model estimated RNDs and how well they correspond to the true RNDs. We have divided the result section into four segments, where we want to answer the following questions,

1) Which model can best replicate the statistical moments of the true RND?
2) Are the model estimated RNDs statistically identical to the true RND?
3) Which model's RND function can most efficiently price hypothetical options?

We proceed to answer these questions in *section 5.1*, where we begin with comparing the RNDs' characteristics in terms of mean, standard deviation, skewness and kurtosis. *Section 5.2* introduces the nonparametric Mann-Whitney U-test, where we investigate if two specific distributions statistically originates from the same population distribution. In *section 5.3* we compare the model estimated RNDs' option valuation capabilities. Lastly, in *section 5.4* we evaluate the models' performance of real market data consisting of historical S&P 500 index option prices. In the last section, *section 6*, we discuss the concluding findings from the thesis and further the limitations and future research within the field of risk-neutral asset return probability distributions.

2. Theoretical Framework

Obtaining the option implied RND is often a challenging task, due to the complexity of the underlying processes. Today's research, conducted on option implied information, mainly focuses on extracting, understanding and making use of the valuable information regarding investors' expectations and risk preferences contained in option prices, where volatility has attracted by far the most attention (Figlewski, 2018). However, obtaining the RND is considered to be a crucial step in pricing derivatives. With a comprehensively estimated RND, one can price any derivative deriving from the same underlying security with the same time to expiration. This is especially useful for illiquid derivatives and derivatives with low trading volume. Further, one can gain useful insights about the actual probabilities and future risk premiums from the RND in conjunction with the pricing kernel, obtaining future market beliefs that are necessary for fiscal policy makers, market agents and other financial institutions (Jackwerth, 2004).

2.1 Risk-Neutral Valuation and Underlying Economics

It is necessary to understand the theory behind risk-neutral valuation, how the risk premium is incorporated in derivative prices and the complexity of a representative investor when modeling derivatives, before one proceeds to estimate and interpret the option implied RNDs.

A derivative market is constituted by two representative agents, hedgers and speculators. Keynes (1930) argued that hedgers utilize the derivative contracts to hedge their exposure in the spot market, while speculators buy the particular derivative contract with the associated risk from the hedgers and in return gets compensated with a profit. The interplay between these two parties subsequently results in the so called risk premium, which is paid by the hedgers and received by the speculators, and is dependent on a magnitude of factors that change over time (Figlewski, 2018). However, modern rational pricing theory debates that in an arbitrage free and frictionless market, there is no additional risk-premium for derivative contracts. For explanation purposes, consider a basic forward contract. A forward contract locks in the current price F_t at which the underlying asset S will be purchased on a future date T. According to the law of one price, two assets with the same payoff ought to have the same market price and assuming the underlying is not paying any dividends, the price of a forward contract has to be equivalent to the spot price and the continuously compounded risk-free interest rate r,

$$F_t = e^{r(T-t)} S_t. \qquad (2.1)$$

If $F_t > e^{r(T-t)}S_t$, an investor could profit from an arbitrage by buying the asset and short the forward contract of the asset. If $F_t < e^{r(T-t)}S_t$, the investor could short the asset and buy the forward contract on that asset (Hull, 2017). Due to this reasoning, there is no risk premium incorporated in the price of a forward contract. When the payoff of a derivative can be exactly replicated by a portfolio of securities available today, rational investors force the prices of the derivatives into a state of risk-neural equilibrium. This concept of replicating a payoff under a riskless arbitrage scenario is the fundamental concept of risk-neutral valuation and the reason why the representative investors' risk preferences are assumed to be irrelevant when pricing derivatives.

2.1.1 Option Pricing

A European call option contract gives the holder the right, but not the obligation to buy the underlying asset at a predetermined date and price. The predetermined date is called the expiration date, which is the date when the option contract can be exercised to buy the underlying asset for the predetermined price, which in turn is called the strike price. The payoff of an option, compared to the payoff of forward, is significantly more complex, since the option's payoff is non-linear and thus difficult to replicate with other securities. It was not until the groundbreaking work of Black and Scholes (1973) and Merton (1973), who introduced their closed-form solution, the Black-Scholes-Merton model, here referred to as the Black-Scholes model, for estimating option prices. Black and Scholes showed that in a complete and frictionless market, an investor can replicate the option payoff with the underlying stock and a bond, which means that all the associated risk with holding the option contract can be hedged away, also referred to as delta hedging. Hence, since an investor can hedge all their risk associated with the option, the expected return of an option should be equal to the risk-free rate. Based on this theory, the expected payoff of the option can be estimated by integrating the payoff function over a risk-neutral density function[2]. Further, the model proposes that any parameters implied from the arbitrage free relationship between the stock and option are not affected by investors' risk preferences, and therefore, all option payoffs have to be discounted with the risk-free rate, which is the discount rate a risk-neutral investor would apply. By introducing the concept of risk-neutral valuation and asset returns, the model permanently changed the fundamental perception of asset pricing theory.

However, the Black-Scholes model has its limitations and has been criticized on multiple occasions. The Black-Scholes model, as many other theoretical valuation models, makes rigid assumptions about the market, including lognormal return distributions, frictionless markets and the absence of arbitrage

[2] See *section 2.3* for an example.

opportunities. Academics have tried to propose countless improvements and expansions to the Black-Scholes formula since its publication. Merton (1974) was quick to re-formulate the model to incorporate a stochastic interest rate. Other extensions include jump-diffusion and pure jump models presented by Merton (1976), Bates (1991) and Madan, Carr, and Chang (1998). However, two of the most prominent and most cited extensions to the model are the stochastic volatility model, proposed by Heston (1993), and stochastic volatility jump-diffusion model, introduced by Bakshi et al. (1997).

2.1.2 Implied Volatility

The models developed based on the work of Black and Scholes, generally relied on the same fundamental procedure for pricing options. One began with making assumptions regarding the underlying stochastic process, for example a geometric Brownian motion, which described the movement of the underlying asset's price. Then one continues to evaluate the stochastic process under risk-neutral conditions and lastly discount the expected option payoff under the risk-neutral measures to price the option (Jackwerth, 1999). However, Rubinstein (1985) argued that these conventional models neglect real world characteristics associated with observed asset prices. His findings showed that out-of-the-money options with shorter time to maturities were overpriced in the market compared to the prices estimated by theoretical models. These findings indicated that the market puts more emphasis on extreme probabilities than assumed under risk-neutral measures. Rubinstein therefore proposed a reversed methodology for option pricing, where one instead observes the market prices and subsequently estimates the underlying implied distribution and stochastic processes based on the observed prices. This novel approach changed the way researchers and investors priced options, and instead began to derive the implied model parameters from observed prices in the market. This approach further addressed the problem of the implied volatility, a necessary parameter for the Black-Scholes formula that was assumed to be constant, and further cannot be observed in the market. The implied volatility could now be estimated by reverse engineering the Black-Scholes model using observed option prices.

The original theory behind the Black-Scholes model stated that the implied volatility is constant across strike prices and over the life span of an option, hence plotting the implied volatility across strike prices would result in a straight line. However, Rubenstein (1985) showed that plotting market implied volatilities across strike prices resulted in a particular pattern, where one could observe that the implied volatility was considerably greater for options with strike prices deep out-of-the-money or in-the-money, referred to as the 'volatility smile'. After the U.S. stock market crash of 1987, researchers found that the volatility smile had changed its shape and sloped downwards as the strike price increased, creating a skewed volatility smile. This new pattern is nowadays more commonly referred to as the 'volatility skew', where the out-of-the-money puts experience a higher volatility than out-of-the-money calls. Mayhew (1995) and Toft and Prucyk (1997) found evidence for the negatively skewed volatility pattern for individual stock options for the American market. Further, Tompkins (2001) studied the volatility for international markets and found convincing patterns for a downward sloping volatility skew for Japanese, German and British index options as well.

There are many theories for the observed volatility skew. One more notable explanation is the leverage effect, where researchers have found a negative correlation between equity returns and volatility. The leverage effect was first proposed by Black (1976), Christie (1982) and Schwert (1989) who argued that the negative correlation is due to an increased debt-to-equity ratio since the company's equity falls, but its debt stays constant. This increase in leverage will subsequently increase the volatility of the stock. Another proposed explanation of the leverage effect is the insurance characteristics of an out-of-the-money put option. These relatively expensive out-of-the-money put options provide a hedge for risk-averse investors in case of a market crash and are therefore willing to pay a premium. Franke, Stapleton, and Subrahmanyam (1999), Mayshar and Benninga (1997), Grossman and Zhou (1996) and Bates (2008) tried to explain this increased investor risk aversion in bear markets by modeling certain investor behavior by introducing a group of heterogeneous investors with another exogenous group who demanded portfolio insurance. However, these results were not consistent with the volatility skew, but rather resulted in moderately sloped volatility smiles. Lastly, it is evident that only a few market participating agents, such as banks and institutions are willing to write more risky contracts, such as deep out-of-the-money put options. Thus, the price of these contract increases, which consequently increases the implied volatility, resulting in a more prominent volatility skew (Jackwerth, 2004).

Recognizing that the implied volatility is skewed is a crucial step for estimating the RND, since the RND is implied from option prices, which subsequently are determined by the implied. Due to the skewed volatility across strikes, the option implied RND will subsequently become skewed and leptokurtic (Jackwerth, 2004). Skewed asset return distributions, which are observed to frequently be left skewed, would hypothetically indicate to a certain extent the markets anticipation of more extreme negative returns. Thus, fatter left tails distribute more probability to lower asset returns, while fatter right tailed distributions indicate a higher probability for particularly higher asset returns. Leptokurtosis, or simply kurtosis, allow the distribution to allocate more probability to the more extreme events in the tails, while remain a certain 'peakedness' around the mean. This would hypothetically imply that there is a higher probability of small and large price changes, while there is a lower probability for intermediate sized price changes (Jackwerth, 2004). Dennis and Mayhew (2002) found convincing evidence of a negative skew in the RND implied from individual stock options. They conclude that stock option prices cannot be exclusively determined by no-arbitrage arguments and thus market risk should be incorporated when theoretically evaluating option prices. The observation of skewed risk-neutral distributions paved the way for further research based on the characteristics of the RNDs. Lynch and Panigirtzoglou (2002) investigated the relationship between RND features and macroeconomic events, but did not manage to find any significant results. Steeley (2004) on the other hand, investigated the relationship between the features of the RND and interest rate announcements and found that there is a relationship between the two. Han (2007) investigated the statistical moments of the RND and investor sentiment. The author concluded that the RND becomes more left skewed, or negatively skewed, when market professionals exhibit bearish characteristics. Conrad, Dittmar, and Ghysels (2013) explored the relationship between the implied volatility, skewness and kurtosis of risk-neutral distributions implied from individual stock options and their respective equity returns. They could conclude that there is a strong evidence between returns and the different characteristics of the particular RND. Martin (2017) and Martin and Wagner (2018) further derived a formula estimating the expected return on the market and also individual stocks, by incorporating the certain characteristics of the option implied RND. They confirmed that there is a relationship between the features of risk-neutral densities used in their model and future expected returns, which they argued, can be applied to many real-world applications and frameworks.

2.2 Option Implied Risk-Neutral Distribution

Security prices can be thought of as the expected value of future cash flows. These expected payoffs constitute a certain probability distribution that indicates the likelihood of these expectations coming true. In case of only one priced security, and under risk-neutral measures, that price would yield one location of that probability distribution, where the expected return on the distribution would be equal to the risk-free rate. However, if one were to observe multiple securities, like option contracts, that derive from same underlying, with the same time to maturity, but different strike prices, one would have information on multiple locations of the distribution, which can be utilized to specify the security’s expected return distribution, which we further refer to as the option implied return probability distribution. We further know, based on the theory presented by Black and Scholes (1973), that options are ought to be priced under risk-neutral measures, since the contract holder can hedge away all risk,[3] and thus, the option implied return distribution should in theory be risk-neutral or referred to as the RND. Therefore, it is important to recognize that options offer a prominent tool for obtaining securities’ RNDs (Jackwerth, 2004). However, asset

[3] Assuming a frictionless and arbitrage-free market, as well as no transaction costs and homogenous, risk-neutral representative market agents.

pricing theory defines two probability distributions, the actual and the risk-neutral probability distribution, where the interplay between these two is constantly subjected to market forces such as representative agents' risk-aversion.

2.2.1 Actual and Risk-neutral Valuation

To fully understand the difference between actual and risk-neutral probabilities, we need to recognize the underlying factors, such as utility functions and risk-aversion that are altering the actual world from being a risk-neutral world. The utility function is a common term within the field of rational choice theory and behavioral finance that tries to measure the utility or satisfaction an individual receives from making a choice over another (Munk, 2013). For instance, why an individual prefers a sure payment, also known as the certainty equivalent, over a fair gamble with the same expected payoff. The shape of the utility function is further determined by the agent's so-called risk aversion. For instance, an individual that dislikes risk, or referred to as a risk-averse agent, would rather receive a certainty equivalent of $45 than a fair gamble with the expected payoff of $50. Thus, we can interpret the agent's utility function as concave, where the utility of the expected value of the certainty payment is increasing at a decreasing rate for higher uncertainty payments. The more risk-averse the agent is, the greater the curvature of the utility function and subsequently less appreciation for higher uncertainty payoffs. For risk-seeking agents on the other hand, the utility function is convex, where the utility of the expected value of the certainty payment is increasing at an increasing rate for higher uncertainty payments. The difference in value between agent's utility of the certainty equivalent and the expected payoff of the fair gamble is referred to as the risk premium (Munk, 2013). In a risk-neutral world, however, the utility function is simply constituted by a linear function, where the investors are indifferent between the certainty equivalent and the fair gamble with the same expected payoff and thus the risk-premium is equal to zero (Jackwerth, 2004). Determining the market representative utility function is a challenging task but can in turn provide insightful knowledge about market beliefs and future asset returns. As proposed by Figlewski (2018), by obtaining the market implied RND and further assume the asset return process or the representative agents utility function, one can subsequently obtain the full actual return distribution.

The first academics to explore the concept of risk-neutral pricing were Cox and Ross (1976) who showed, by introducing stochastic processes and jumps, that the risk-neutral option valuation problem itself is equivalent to the problem of determining the distribution of the terminal stock price, and subsequently presented the link between option valuation and the stochastic processes of stocks. Rubinstein (1976), and further improvements by Brennan (1979), developed a risk-neutral valuation method of securities and uncertain income streams, with respect to the risk-averse investor utility function. They argued that the market representative investor's risk aversion does matter and thus proved that only under a constant investor utility function and when returns are log normally distributed, the Black-Scholes model is a sufficient method for pricing European options.

Black and Scholes (1973) showed that derivatives traded on an arbitrage free market ought to be priced based on the assumption that all investors are risk-neutral. Assuming a risk-neutral world has two beneficial features: first, the expected return on a stock is the risk-free rate and secondly, the discount rate of an expected payoff on any derivative is equivalent to the risk-free rate (Hull, 2017). However, if we instead incorporate actual probabilities, constituted by agents' utility and risk-aversion, the derivative associated with more risk in its payoff distribution should have a greater expected return, and thus a lower price, than a derivative associated with less risk. Hence, the relationship between the risk-neutral and actual probabilities depends on how much the market representative agent is willing to pay for risk (Conrad et al., 2013).

To further investigate the relationship between actual and risk-neutral distributions, we can construct an example, as shown by Jackwerth (2004), based on the influential work of Cox et al. (1979) and their fundamental "binomial model" for pricing options. Here we assume a complete and arbitrage free market, which can only take on two states in one year, perceptually equivalent to a binomial tree.[4] In our example, we define the current stock price (S_0) as \$100, in the up-state (S_T^u) the price as \$120, and in the down-state (S_T^d) as \$80. The actual probabilities of the two future states are 0.8 (p^u) and 0.2 $\left(p^d\right)$, respectively. The bond (B_0) is priced at \$100 today and \$105 in the two future states (B_T), which is equivalent to a risk-free rate (r) of 5%. To price a call option on that particular stock with the strike (K) of \$100, one can estimate the discounted payoff under the given actual probabilities. With only these two states, up and down, the price of the call option (c) at time $T = 1$ can be defined as,

$$c = \frac{1}{(1+r)^T}\left[p^u C_T^u + p^d C_T^d\right], \tag{2.2}$$

for,

$$0 \leq p^u \leq 1, \quad p^d = 1 - p^u,$$

where C_T^i is the payoff of the call option at time T expressed as,

$$C_T = \max(0, S_T - K). \tag{2.3}$$

[4] For a complete definition of binomial trees for option pricing, see the inspiring work of Cox et al. (1979).

Thus, the expected payoff is equivalent to $\$20 * 0.8 + \$0 * 0.2 = \$16$. Further discounting the payoff, one can obtain the option price $\$16/(1 + 0.05) = \$15.2381$. However, for this option the market price is $\$11.9048$. The reason for these two different call prices is the representative market investors' utility function and corresponding risk-aversion. The representative investor has less appreciation for the payoff in the good state when she is already wealthy, thus is willing to pay less for the call option than implied by the actual probabilities. Hence, to evaluate the market price of the option, one has to utilize the concept of risk-neutral pricing and state-prices.

The state-price is simply referred to as the price an investor is willing to pay for a certain payoff in that particular state of the economy. Here we denote π^u as the state-price in the good state and π^d as the state-price in the bad state. Elaborating on the example given earlier, one can set up the equation system for the stock and the bond to calculate the state-prices in both states as follows,

$$S_0 = \pi^u S_T^u + \pi^d S_T^d, \qquad B_0 = \pi^u B_T + \pi^d B_T. \tag{2.4}$$

Solving the equation systems, results in $\pi^u = 0.5952$ and $\pi^d = 0.3571$. The sum of the state-prices has to be equal to the price of a zero-coupon bond that pays $1 in both states (Jackwerth, 2004). Continuing, multiplying the state-prices with the inverse of this unit bond results in the risk-neutral probabilities q^i,

$$q^i = (1 + r)^T \pi^i. \tag{2.5}$$

Thus, the risk-neutral probabilities are $q^u = 0.6250$ and $q^d = 0.3750$. Using the risk-neutral probabilities in *equation* (*2.2)*, substituting the actual probability p for the risk-neutral q, one can evaluate the price of the call option at time T to be equivalent to $\$11.9048$. In this simple, complete market scenario, any security with payoff at time T can be evaluated once the risk-neutral distribution is obtained.

Assuming no arbitrage, the state-prices and actual probabilities contain valuable information about the representative market agent (Syrdal, 2002). The ratio between the two probabilities, referred to as the pricing kernel (m) or the stochastic discount factor, which represents the investor's expected marginal utility growth (Jackwerth, 2004). The pricing kernel is defined as,

$$m^i = \frac{\pi^i}{p^i}. \tag{2.6}$$

In the example given, the pricing kernel is equivalent to $m^u = 0.7441$ in the up-state and $m^d = 1.8750$ in the down-state. Thus, the investor has a larger marginal utility in the down-state than in the up-state and, thus, has more appreciation for a certainty equivalent payoff in the down-state than in the up-state. The pricing kernel here reveals that the market agent in this example is risk averse and want to be compensated for the down-state exposure.

As stated earlier, common asset pricing theory is based on the assumption that market agents are risk-averse. That means, in complete markets, with risk-averse investors and common true beliefs, the pricing kernel can be interpreted as a convex function or a decreasing function over increased wealth. However, Jackwerth (2000) showed in his work of recovering risk aversion from option prices, which the implied pricing kernel of the S&P 500 is not decreasing, but rather increasing with aggregated resources, to a certain point of wealth, after a financial crash. Thus, proving the opposite of what common financial behavior theory would suggest. This pricing kernel puzzle implies that instead of investors being risk-averse, they are risk-seeking and would rather pay for a fair gamble than pocketing the certainty equivalent. The pricing kernel puzzle was further observed by Rosenberg and Engle (2002) for the S&P 500 index and later Jackwerth (2004) showed that the puzzle seem to appear internationally on the indices of Germany, United Kingdom and Japan.

2.3 Estimation of the Risk-neutral Density

A complete market is present when the complete set of possible future states of the world can be constructed with existing assets, which have linearly independent payoffs. In the example above, there were two states, up and down, as well as two securities, one stock and one bond, thus simulating a complete market. The theory behind complete markets can be traced back to Arrow and Debreu (1954), Debreu (1959) and Arrow (1974), who proved mathematically the existence of a general market equilibrium. Both Arrow and Debreu were awarded the Nobel Prize for their work, mainly due to their work within the field of complete markets and how to apply the theory to the problem of general equilibrium. Ross (1976) was one of the first to investigate the complete market efficiency and how it impacts option prices and subsequently the option implied RND. Ross showed that one can determine the complete European option implied RND from a set of European option prices, while assuming the market is complete. Two years later, Breeden and Litzenberger (1978) showed that option prices, which are only dependent on the uncertainty of the future state of the underlying, can be expressed as a function $C(K)$ across strikes K and were able to derive a formula for the state-prices as a function of future stock prices.

Consider the equation for the price of a call option with the payoff given in *equation (2.3,)*

$$c(S_t, K) = e^{-rT} \int_{S_T=K}^{\infty} (S_T - K) q(S_T) dS_T, \tag{2.7}$$

where the integral is over the distribution when the call has a positive payoff $S_T \geq K$, $q(S_T)$ is the risk-neutral probability density function of the terminal stock price. If we differentiate the equation with respect to K once, we get the following equation,

$$\frac{\partial c}{\partial K} = -e^{-rT} \int_{S_T=K}^{\infty} q(S_T) dS_T. \tag{2.8}$$

Differentiate once more with respect to K,

$$\frac{\partial^2 c}{\partial K^2} = e^{-rT} q(K), \tag{2.9}$$

where the risk-neutral density function $q(S_T)$, for $K = S_T$, is equal to,

$$q(S_T) = e^{rT} \left. \frac{\partial^2 c}{\partial K^2} \right|_{K=S_T}. \tag{2.10}$$

Thus, indicating that the RND is equivalent to the forward value of the second partial derivative of the call option price with respect to its strike price.[5] Rewriting the risk-neutral probability function from *(2.10)* with the result from *equation* (*2.5)*, shows that the state-price is equivalent to the second derivative of a call option with respect to its strike price when the future underlying asset is equal to the strike price

$$\pi(S_T) = \left. \frac{\partial^2 c}{\partial K^2} \right|_{K=S_T}. \tag{2.11}$$

Breeden and Litzenberger showed that in a complete market, with a strike price for every possible state of the future stock price, one can obtain the risk-neutral density with the *equation* (*2.10)*. However, under actual market conditions strike prices are far apart, thus, option prices have to be interpolated from models based on estimated parameters, such as implied volatility. The interpolation of option prices can further be a challenging task, since a majority of option markets are illiquid and especially deep-out-of-the-money and deep-in-the-money options are rarely traded.

[5] See *appendix a)* for a complete derivation of the Breeden-Litzenberger equation.

3. Methodology

Obtaining the RND is a crucial, but often overlooked step in pricing derivatives. With a comprehensively estimated RND one can price derivatives that derive from the same underlying security and with the same time to expiration. Thus, one can estimate the RND from a set of liquid options and continue to price illiquid options. These options can have exotic payoffs, but they have to follow European exercise restrictions, since RNDs do not accommodate for early exercises (Jackwerth, 1999). Further, one can gain useful insights about the actual probabilities and future risk premiums from the RND in conjunction with the pricing kernel (Jackwerth, 2004). A wide variety of approaches have been proposed to estimate the RND from option prices. In this section we will review the different methods and approaches for obtaining the option implied RND, as well as discuss and elaborate on the three specific methods chosen for the purpose of this thesis, namely the mixture model, kernel regression model and neural network model.

According to Jackwerth (1999) there are three approaches to estimate the option implied RND. The first approach involves establishing a reasonable probability distribution with assumed underlying parameters, then fit the resulting option prices, based on the estimated density, to observed option prices and minimize the error to subsequently optimize the underlying parameters. Second approach is to estimate a function of option prices across strike prices, based on observed option prices. Then use the methodology proposed by Breeden and Litzenberger (1978), discussed in *section 2.3*, to retrieve the RND. The third and more stable approach, first proposed by Shimko (1993), is to instead estimate a function of implied volatilities across strike prices based on the observed market data. Based on the function of implied volatilities, estimate the option prices using the Black-Scholes model (Black & Scholes, 1973) and again use the Breeden-Litzenberger approach to obtain the RND.

The three different approaches are conducted through applying different methods or models to succeed with the desired task. There are numerous methods suggested in academic literature for each of the three approaches mentioned.[6] Jackwerth (1999, 2004) categorize these methodologies into two main categories: parametric and nonparametric methods, based on the survey articles from Cont (1997) and Bahra (1997). Other researchers such as Jondeau and Rockinger (2000) and Bu and Hadri (2007) introduce semi-parametric methods as an intermediate category. However, for the purpose of this thesis, we will focus on the two main categories parametric and nonparametric.

[6] See Jackwerth (2004) for a detailed overview of the different methods for each approach.

3.1 Parametric and Nonparametric Methods

Both parametric and nonparametric methods are used estimate the RND by calibrating the underlying parameters associated with the desired model. However, parametric methods define the model as a function of a finite number of parameters and thus have to make assumptions regarding the specific characteristics of the implied RND (Lai, 2014). In general, a parametric model starts with assuming a predetermined RND with designated features such as skewness or excess kurtosis, as well as defining the associated parameters to optimize for the model. The model calibrates these parameters by minimizing the error between the model's estimated option prices and observed option prices (Jackwerth, 1999). Thus, with a set of optimized parameters, one can easily obtain the estimated RND. Parametric methods are usually a popular alternative, since they are generally intuitive and easy to interpret. Further, parametric methods are less computationally demanding and do not require a substantial amount of data, due to their small set of parameters over which to optimize. However, the models require predetermined assumptions regarding the underlying data-generation process, which is often difficult to validate (Santos & Guerra, 2015).

Compared to parametric methods, nonparametric methods have hypothetically an infinite number of parameters, with more degrees of freedom and are thus a lot more flexible (Lai, 2014). Nonparametric methods try to directly estimate the function $\hat{y}(\cdot)$ on the underlying data Z^*, without the need of making any major assumptions regarding the underlying data-generation process and regression function. This makes nonparametric methods easy to utilize in scenarios where extensive data is available and one can only make vague assumptions regarding the underlying data process. When it comes to estimating option implied RNDs, nonparametric methods are in general used to fit the implied volatility $\sigma(\cdot)$ on the underlying data Z that is represented by a set for features $Z = \{K, S, \tau, r, \delta\}$, such as strike price ($K$), spot price ($S$), time to maturity ($\tau$), risk-free rate ($r$) and dividends ($\delta$). However, nonparametric methods are in general more computationally demanding and often require extensive datasets. Nonparametric methods are again usually harder to interpret and tend to be regarded as a black-box.

3.2 Mixture Model

Black and Scholes (1973) assumed that the underlying asset price follows a geometric Brownian motion, which in turn generates a simple log-normal return probability distribution, inadequate for real world scenarios. A mixture model is essentially an extension of the original Black-Scholes model, where the model utilizes the added flexibility from multiple simple distributions to capture the features of a certain return distribution, such as skewness and leptokurtosis. Most famously Bahra (1997), but also other academics like Melick and Thomas (1997), Söderlind and Svensson (1997), Gemmill and Saflekos (2000) and Schittenkopf and Dorffner (2001), argued that the option implied RND is simply constituted by a linear combination of weighted average log-normal densities, for $K = S_T$, defined as,

$$q(S_T) = \sum_{i=1}^{n} [w_i L_i(\alpha_i, \beta_i; S_T)] \Big|_{K=S_T} \tag{3.1}$$

where $q(S_T)$ is the RND evaluated for S_T, $L_i(\alpha_i, \beta_i; S_T)$ is the lognormal density with the unknown parameters α_i and β_i, $i = 1, \ldots, n$, where n is the number of densities used to evaluate the RND. We further can define the i-th lognormal density as,

$$L_i(\alpha_i, \beta_i; S_T) = \frac{1}{S_T \beta_i \sqrt{2\pi}} e^{[-(\ln(S_T) - \alpha_i)^2 / 2\beta_i^2]},$$

$$\alpha_i = \ln(S_t) + \left(\mu_i - \frac{1}{2}\sigma_i^2\right)\tau,$$

$$\beta_i = \sigma_i \sqrt{\tau}, \tag{3.2}$$

where μ_i is the instantaneous drift, σ_i is the volatility of the associated return distribution and w_i is the weight of the distribution, where $w_i \geq 0$ and $\sum_{i=1}^{n} w_i = 1$. By optimizing these parameters, μ_i, σ_i, w_i, and subsequently summarize the number of n-distributions, one can achieve enough flexibility to obtain a well behaved RND from a cross-sectional set of option prices. Following the methodology of Bahra (1997), we can optimize our parameters by first pricing the respective call and put options, defined as follows,[7]

$$c(K, \tau) = e^{-r\tau} \sum_{i=1}^{n} w_i \left[\exp\left(\alpha_i + \frac{1}{2}\beta_i^2\right) N(d_i) - KN(b_i)\right],$$

$$p(K, \tau) = e^{-r\tau} \sum_{i=1}^{n} w_i \left[KN(-b_i) - \exp\left(\alpha_i + \frac{1}{2}\beta_i^2\right) N(-d_i)\right], \tag{3.3}$$

where we define,

$$d_i = \frac{-\ln(K) + \alpha_i + \beta_i^2}{\beta_i},$$

$$b_i = d_i - \beta_i, \tag{3.4}$$

[7] See Jondeau, Poon, and Rockinger (2007) for a more detailed derivation of the equation.

where $N(d_i)$ is the cumulative density function of the standard normal distribution. Thus, the combination of log-normal densities results in a linear combination of Black-Scholes equations, where the call option price depends on the parameter vector $\theta_i = \{\mu_i, \sigma_i, w_i\}$. In order to find the optimal parameters for associated distribution, one has to minimize the following loss function,

$$\theta_i = arg \min_{\mu_i,\sigma_i,w_i} \sum_{j=1}^{N} \left[c_{t,j} - \hat{c}(K,\tau)_{t,j}\right]^2 + \left[p_{t,j} - \hat{p}(K,\tau)_{t,j}\right]^2, \qquad (3.5)$$

where c_t and p_t are the market observed call and put prices at time t and $\hat{c}(K,\tau)_t$, $\hat{p}(K,\tau)_t$ are the estimated call and put prices at time t from *equation (3.3)*. The optimized parameter vector θ is then used in *equation (3.1)* to obtain the RND.

Bahra (1997) utilized a relatively simple mixture model, consisting of only two mixture distributions, but theoretically one could use a greater number of distributions to estimate the RND. However, the added flexibility of more distributions comes at a cost of more parameters to be optimized. For instance, utilizing three distributions would require one to estimate eight parameters, two for each distribution and two weights, where the third weight is the result from the constraint that the sum of weights has to be equal to one. Further, increase the number of distributions and the model becomes prone to overfitting the in-sample data, where Giamouridis and Tamvakis (2002) argue that more than three distributions is enough for a mixture model to start overfitting the data. Thus, the model developed for the purpose of this thesis will constitute of three mixture distributions to be able to accommodate for the complexity in the return distributions of option prices, without overfitting the in-sample data. Hence, the model evaluates a total of eight parameters for every time to maturity τ at time t.

3.3 Kernel Regression

Kernel regression is nonparametric technique that is utilized to estimate the conditional expectation of a random variable. In the case of pricing options, the model is used to fit the nonlinear function of the call option price to the associated, underlying factors Z. However, it is more common to estimate the implied volatility function $\sigma(Z)$ due to the frequent changes in option prices compared to the minor changes in implied volatility over a trading day (Jackwerth, 2004). Thus, kernel regression models are constructed to estimate the true function for implied volatility $\sigma(Z)$ in high dimensional data using identical weighted functions called kernels (k). Suppose we want to estimate the relationship between two variables, Z_i and σ_i that satisfy the following nonlinear relationship,

$$\sigma_i = \sigma(Z_i) + \varepsilon_i, \qquad i = 1, \dots, n,$$

where $\sigma(\cdot)$ is an nonlinear function and ε_i is white noise. The kernels are used to estimate the unknown function $\sigma(\cdot)$ at each specific point $Z_{i_0} = z_0$, by taking the weighted average of observations $\sigma_{i_0}^{(j)}, j = 1, \dots, n$ around point z_0. The further away the observations are from the point z_0, the lower the likelihood that the true function goes through that distant point, and thus, is assigned less weight (Aït-Sahalia & Lo, 1998). Hence, to obtain the estimated function, we have to compute a weighted average of the observations σ_i for each value of z in the domain of $\sigma(\cdot)$.

For the purpose of this thesis, we will only cover the basic derivation of the kernel estimator used for regression purposes. For a thorough derivation, see the work of Efromovich (2008) and Chacón and Duong (2018). We begin by defining the kernel regression as the conditional expectation of Y given X as,

$$\begin{aligned} E(Y \mid X) &= \int y f(y \mid x) dy, \\ &= \frac{\int y f(x, y)}{f(x)} dy, \end{aligned} \tag{3.6}$$

where $f(y|x)$ is the conditional density function of Y given $X = x$, $f(x, y)$ is the joint probability density function of Y and X and $f(x)$ is the marginal density function of X. With the given observations X_i, $i = 1, \dots, n$, with each $X_i \in \mathbb{R}$, we can estimate the density function $\hat{f}(x)$ using a so-called kernel estimator,

$$\hat{f}(x; h) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right), \tag{3.7}$$

where we for simplicity denote our kernel k as,

$$k_h(x) = \frac{1}{h} k\left(\frac{x}{h}\right), \tag{3.8}$$

so we can further simplify our kernel estimator as,

$$\begin{aligned} \hat{f}(x; h) &= \frac{1}{n} \sum_{i=1}^{n} k_h(x - X_i), \\ k_h(x - X_i) = \phi(x; X_i, h^2) &= \frac{1}{h\sqrt{2\pi}} exp\left[-\frac{(x - X_i)^2}{2h^2}\right], \end{aligned} \tag{3.9}$$

where kernel k_h, dependent on the bandwidth h, is the density of a normal distribution $N(X_i, h^2)$, where the bandwidth h for this univariate kernel estimator is the standard deviation, that determines the spread of the kernel around the data point X_i. Too high bandwidth and the function becomes too smooth, while a too low bandwidth will result in overfitting the data.[8] Further, we need to define the kernel estimator that can be extended for multivariate data. We can specify the multivariate kernel estimator $\hat{f}$ evaluated at x, for a d-dimensional data sample X_i for $i = 1, \dots, n$, with each $X_i \in \mathbb{R}^d$ as,

$$\hat{f}(x; H) = \frac{1}{n|H|^{1/2}} \sum_{i=1}^{n} k\left(H^{-1/2}(x - X_i)\right), \tag{3.10}$$

where k is referred to as a multivariate kernel, that depends on the $d \times d$ bandwidth matrix H. To simplify the expression, similar to *equation (3.8)*, we can denote the multivariate kernel as,

$$k_H(x) = |H|^{-\frac{1}{2}} k\left(H^{-\frac{1}{2}} x\right), \tag{3.11}$$

hence, we can express the multivariate kernel estimator, similarly to *(3.9)*, as,

$$\hat{f}(x; H) = \frac{1}{n} \sum_{i=1}^{n} k_H(x - X_i),$$
$$k_H(x - X_i) = \phi(x; X_i, H) = \frac{d}{\sqrt{2\pi|H|}} \, exp\left[-\frac{(x - X_i)^2}{2H}\right], \tag{3.12}$$

where K_H is the normal density centered at the mean X_i and with variance matrix H. The bandwidth matrix provides increased flexibility, but with an increased number of bandwidth parameters to be selected. Thus, to increase computational efficiency, we instead consider a positive diagonal bandwidth matrix $H = diag(h_1^2, \dots, h_d^2)$ (Epanechnikov, 1969), where we incorporate dot-product kernels for the multivariate kernel estimator as follows,

$$\hat{f}(x; h) = \frac{1}{n} \sum_{i=1}^{n} k_{h_1}\left(x_1 - X_{i,1}\right) \times \dots \times k_{h_d}\left(x_d - X_{i,d}\right), \tag{3.13}$$

where we recognize that $h = [h_1, \dots, h_d]$ is the vector of bandwidths. Utilizing the multivariate kernel estimator *(3.13)*, we can further express the joint density function $f(x, y)$, for a given two dimensional sample X_i, Y_i for $i = 1, \dots, n$ and where our bandwidths are defined as $h = [h_1, h_2]$, as follows,

$$\hat{f}(x, y; h) = \frac{1}{n} \sum_{i=1}^{n} k_{h_1}(x - X_i) k_{h_2}(y - Y_i), \tag{3.14}$$

Substituting *(3.14)* in the numerator of *equation (3.6)*, we can rewrite the numerator as,

---

[8] Brief explanation of overfitting is covered in section 3.4.

$$\begin{aligned}\int y\hat{f}(x,y;h)dy &= \int y\left[\frac{1}{n}\sum_{i=1}^{n} k_{h_1}(x-X_i)k_{h_2}(y-Y_i)\right]dy \\ &= \frac{1}{n}\sum_{i=1}^{n} k_{h_1}(x-X_i)\int yk_{h_2}(y-Y_i)dy \\ &= \frac{1}{n}\sum_{i=1}^{n} k_{h_1}(x-X_i)\int y\hat{f}(y)dy.\end{aligned} \tag{3.15}$$

As shown by Nadaraya (1964) and Watson (1964), we can write the function, after some manipulations, as,

$$\int y\hat{f}(x,y;h)dy = \frac{1}{n}\sum_{i=1}^{n} k_{h_1}(x-X_i)\,Y_i. \tag{3.16}$$

Using *(3.16)* and *(3.9)* in *equation (3.6)*, we obtain the kernel regression formula, also known as the Nadaraya-Watson estimator, which is given by,

$$\hat{E}(Y \mid X=x) = \frac{\frac{1}{n}\sum_{i=1}^{n} k_h(x-X_i)\,Y_i}{\frac{1}{n}\sum_{j=1}^{n} k_h(x-X_j)}. \tag{3.17}$$

Using the Nadaraya-Watson estimator we can define our initial framing of the problem, where the expectation of the implied volatility σ conditional on underlying data Z, to find $\sigma(Z) = \hat{E}(\sigma|Z)$,

$$\begin{aligned}\hat{E}(\sigma|Z) &= \frac{\frac{1}{n}\sum_{i=1}^{n} k_h(Z-Z_i)\,\sigma_i}{\frac{1}{n}\sum_{j=1}^{n} k_h(Z-Z_j)} \\ &= \sum_{i=1}^{n}\left[\frac{k_h(Z-Z_i)}{\sum_{j=1}^{n} k_h(Z-Z_j)}\right]\sigma_i.\end{aligned} \tag{3.18}$$

If we define the weight as follows,

$$W_i = \frac{k_h(Z-Z_i)}{\sum_{j=1}^{n} k_h(Z-Z_j)}, \tag{3.19}$$

we can further simplify the notation as,

$$\hat{E}(\sigma|Z) = \sum_{i=1}^{n} W_i\sigma_i. \tag{3.20}$$

where the set of weights sum to one, $\sum_{i=1}^{n} W_i(Z) = 1$. Thus, we can see that the Nadaraya-Watson estimator estimates the conditional expectation of the implied volatility as a weighted average of the observed implied volatilities at each point z_i.

As stated earlier, bandwidth selection is a crucial part of kernel regression and there are numerous suggestions for choosing the optimal bandwidth. For the purpose of this thesis, we will use one of the more common alternatives, the so-called least squares cross-validation. We can look at selecting the bandwidth h by minimizing the error between observed implied volatilities σ_i and the function $\hat{\sigma}(Z;h)$ such that,

$$\hat{h} = arg\min_{h>0}\frac{1}{n}\sum_{i=1}^{n}[\,\sigma_i - \hat{\sigma}(Z_i;h)]^2. \quad (3.21)$$

However, this will in the majority of cases lead to $\hat{h} = 0$. Thus, we have to restrict $\hat{h} > 0$ and instead use a so called leave-one-out cross-validation, where we compare σ_i with $\hat{\sigma}_{-i}(Z_i;h)$, where we leave out the i-th starting point, resulting in the least square cross-validation error,[9]

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}[\,\sigma_i - \hat{\sigma}_{-i}(Z_i;h)]^2, \quad (3.22)$$

where we define $\hat{h}$ as,

$$\hat{h} = arg\min_{h>0} CV(h). \quad (3.23)$$

With the estimated function $\hat{\sigma}(Z)$, we can evaluate the implied volatility at each strike price K_i. Utilizing the estimated implied volatility, we use the Black-Scholes formula for our option pricing function and lastly, estimate the implied RND using the Breeden-Litzenberger equation *(2.13)*, by taking the partial derivative of the option pricing function with respect to the strike price twice.

The kernel regression method requires only a few assumptions other than the smoothness of the function and certain specifications and boundaries in the data used to estimate it. However, the method is data-intensive and requires a considerable amount of data to properly fit the estimated function (Aït-Sahalia & Lo, 1998). Further, kernel regressions tend to perform poorly on data that exhibit gaps, where the function is unable to fit a smooth continuation of the function across gaps observed for implied volatilities distributed over a discrete set of strike prices (Jackwerth, 2004).

3.4 Artificial Neural Network

An artificial neural network, or simply referred to as a neural network, is a significantly simplified version of the extremely convoluted biological neural networks and is one of many subcategories in the

[9] For a detailed explanation on least squares cross-validation, see the works of Efromovich (2008) Chacón and Duong (2018).

field of machine learning. A neural network model is composed of weighted interconnections between different nodes, which together have the benefit of evaluating nonlinear functions and find relationships in the data. Similar to the kernel regression, the neural network model for the purpose of regression analysis is trying to estimate the implied volatility function $\hat{\sigma} = f(x)$ directly to the underlying data Z, without the need for any major assumptions regarding the underlying distribution or regression itself. There are multiple different types of neural networks, with different architectures for specific tasks. However, for the purpose of this thesis, we will cover the multilayer perceptron model, or more commonly referred to as the feedforward network, for regression analysis.

The interconnected nodes in the feedforward network are arranged in three distinct sections: The input layer, the hidden layers and the output layer, where each layer is vector-valued. The input layer consists of input nodes, where each node X_f corresponds to one of the associated features included in the dataset. The nodes $X_f, f \ = 1 \ldots, n$, where n is the number of features, in the input layer are connected through weights $W_{j,k}^{(l)}$ to the nodes in the hidden layer, where the notation order is reversed, thus k represents the node in the previous layer, j the node in the following layer, where l indicates the layer. Next, we have the hidden layers, which are constituted by activation nodes, where one node $A_k^{(l)}$, in layer l, is connected to every node $A_j^{(l+1)}, j = 1, \ldots, m$, in layer $l + 1$, as well as to every node $A_h^{(l-1)}, h = 1, \ldots, M$, in layer $l - 1$, thus we can think of the layers as representing a vector-to-scalar function, where each layer represents a function $f^{(l)}(x; W, b)$, where b is the bias from previous layer. Further, for the input layer and each hidden layer we add one bias node with no input and connected with its own weight to the activation nodes in the next layer. The bias node is a simple node that only outputs the value 1 and thus provide more flexibility to the model as it allows the network to fit the data when all input features are equal to zero (Goodfellow, Bengio, & Courville, 2016). The activation nodes and bias in the last layer in the hidden layer $L - 1$, are in turn connected through weights to the last layer, the output layer L. The output layer is represented by one node Y, which combines the linear input from the hidden layers to subsequently produce the estimated regression value $\hat{\sigma}$. The name feedforward network indicates that the information flows forward through the network, from the input nodes, through the hidden layers and is ultimately evaluated at the output layer.

Each activation node $\left(A_k^l\right)$ consists of two parts. The first part is constituted by a summarization function, denoted z_k^l, that evaluates the linear input from the previous layer as follows,

$$z_k^{(l)} = b^{(l)} + \sum_{h=1} W_{k,h}^{(l)} a_h^{(l-1)}, \tag{3.24}$$

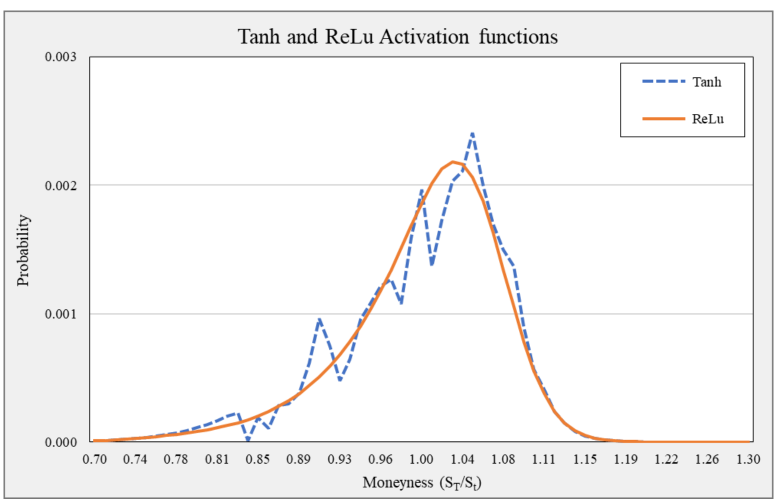
where b^l represents the bias in layer l, $a_h^{(l-1)}$ is the activation function from the previous layer $l-1$ and $W_{k,h}^{(l)}$ is the weight between the summarization function $z_k^{(l)}$ and activation function $a_h^{(l-1)}$. The second part of the activation node is the activation function $[a_k^l(z_k^l)]$ that takes the summarization function as an argument and proceeds to evaluate a linear or nonlinear function. The specific activation function is decided upon in advance, with multiple variations of activation functions to choose from. For the purpose of this thesis we will cover the hyperbolic tangent function or Tanh function, $Tanh(z)$. While other activation functions, such as the rectified linear unit function (ReLu), are more commonly used in today's models, the Tanh function is suitable for predicting implied volatility, since we are generally estimating a value between zero and one. Thus, the selected Tanh function will in return predict a smoother option implied RND. The comparison between the RNDs estimated by the Tanh activation function and ReLu activation function is illustrated in *figure 1*. We can define the Tanh function as follows,

$$Tanh(z) = 2\vartheta(2z) - 1, \tag{3.25}$$

where ϑ is a simple sigmoid function,

$$\vartheta(z) = \frac{1}{1+e^{-z}}. \tag{3.26}$$

***FIGURE 1*. DENSITY ESTIMATION FOR THE TANH AND RELU ACTIVATION FUNCTION**



The figure illustrates the difference between the Tanh and ReLu activation function, where one can observe the clear difference in density generalization made by the two functions.

To measure the performance of our model, we need to measure its prediction capabilities. We can do this by defining a simple loss function, in this case the mean squared error loss function, where we set to minimize the mean squared error between the estimated and the observed implied volatilities. We can define our loss function as,

$$J(x) = \frac{1}{n}\sum_{i=1}^{n}[\,\sigma_i(x) - \hat{\sigma}_i(x; W, b)]^2. \tag{3.27}$$

where $\sigma_i(x)$ is the observed implied volatility at point x. To effectively reach a sufficiently small error for our loss function, we need to implement an optimization algorithm, more specifically a gradient-based optimizing algorithm. A gradient-based algorithm is used to find the gradient, or the first partial derivative of the cost function with respect to every weight and bias in the network. By obtaining the gradient, one can efficiently optimize the corresponding weights and biases and subsequently minimize the desired cost function. To compute the gradient of the cost function, neural networks utilize the back-propagation algorithm introduced by Rumelhart, Hinton, and Williams (1988). For the purpose of this thesis we will only briefly touch upon the subject of backpropagation. For a thorough coverage of backpropagation, see the works of Rumelhart et al. (1988) and Goodfellow et al. (2016).

The back-propagation algorithm operates by utilizing the chain rule of calculus, which states that one can calculate the derivatives of functions formed by other composite functions whose derivatives are known. Thus, by back propagating through the network, starting at the output layer's scalar cost function, the algorithm can evaluate the first partial derivative of that cost function with respect to every weight and bias in the network. By finding the first partial derivative of the cost function, $\nabla J(W, b)$, for each weight and bias, the model can incrementally adjust these weights and biases accordingly to minimize the desired cost function. In the proposed model, our output variable $\hat{\sigma}$ is defined by the weighted sum of the activation functions in the last hidden layer $L - 1$, plus a bias, obtained from *equation (3.24)* and denoted as $z^{(L)}$. Hence, we define our cost function J_1 for one training iteration over one observation as,

$$J_1 = (\sigma_1 - z^{(L)})^2. \tag{3.28}$$

Further, we can calculate the partial derivative of the cost function J_1 with respect to the weights $W_j^{(L)}$ in layer L, utilizing the chain rule, as such,

$$\frac{\partial J_1}{\partial W_j^{(L)}} = \frac{\partial z^{(L)}}{\partial W_j^{(L)}}\frac{\partial J_1}{\partial z^{(L)}} \tag{3.29}$$

where,

$$\frac{\partial z^{(L)}}{\partial W_j^{(L)}} = \frac{\partial}{\partial W_j^{(L)}} b^{(L)} + \sum_{j=1}^{m} W_j^{(L)} a_j^{(L-1)} = a_j^{(L-1)}, \quad \frac{\partial J_1}{\partial z^{(L)}} = 2(\sigma_1 - z^{(L)})$$

where $a_j^{(L-1)} = tanh\left(z^{(L-1)}\right)$ and $j = 1, \dots, m$, where m is the total number of nodes in the $L-1$ layer. Further, we can continue to apply the chain rule to calculate the partial derivative of the cost function with respect to the weights $W_{jk}^{(L-1)}$ in the previous layer, $L-1$, that are connected to the k-th node in layer $L-2$ and the j-th node in layer $L-1$, by the same principle,

$$\frac{\partial J_1}{\partial W_{jk}^{(L-1)}} = \frac{\partial z_j^{(L-1)}}{\partial W_{jk}^{(L-1)}} \frac{\partial a_j^{(L-1)}}{\partial z_j^{(L-1)}} \frac{\partial z^{(L)}}{\partial a_j^{(L-1)}} \frac{\partial J_1}{\partial z^{(L)}} \tag{3.30}$$

where,

$$\frac{\partial a_j^{(L-1)}}{\partial z_j^{(L-1)}} = tanh'\left(z_j^{(L-1)}\right), \qquad \frac{\partial z^{(L)}}{\partial a_j^{(L-1)}} = W_j^{(L)},$$

and $k = 1, \dots, M$, where M is the total number of nodes in the $L-2$ layer. The chain rule can be applied to any of the functions in the network. On the other hand, as one evaluates the activation functions further back in the network, for instance calculating the partial derivative of $z_j^{(L-1)}$ with respect to $a_k^{(L-2)}$, one has to acknowledge that the z_j functions is influenced by all the activation functions in layer $L-2$, since $a_k^{(L-2)}$ is connected to all summarization functions $z_j^{(L-1)}$ in the next layer. Generally speaking, the activation functions in layers further away from the output layer indirectly affect the cost function through multiple paths throughout the network. Thus, we have to evaluate the partial derivative of the cost function with respect to $a_k^{(L-2)}$, to calculate the activation function's total impact on the cost function. We do this by summarizing the chain rule of partial derivatives of the functions in layers $L-1$ and L. We define the partial derivative of the cost function J_1 with respect to $a_k^{(L-2)}$ as,

$$\frac{\partial J_1}{\partial a_k^{(L-2)}} = \sum_{j=1}^{m} \frac{\partial z_j^{(L-1)}}{\partial a_k^{(L-2)}} \frac{\partial a_j^{(L-1)}}{\partial z_j^{(L-1)}} \frac{\partial z^{(L)}}{\partial a_j^{(L-1)}} \frac{\partial J_1}{\partial z^{(L)}}, \tag{3.31}$$

where,

$$\frac{\partial z_j^{(L-1)}}{\partial a_k^{(L-2)}} = W_{jk}^{(L-1)},$$

and m is the number of nodes in layer $L-1$. Hence, when we know the gradient of the cost function with respect to the activation functions in layer $L-2$, we can further repeat the process of calculating the partial derivatives of the cost function with respect to the weights and biases that feeds in to activation nodes in layer $L-2$. For instance, the partial derivative of the cost function with respect to the weights in layer $L-2$, can be written as,

$$\frac{\partial J_1}{\partial w_{k,m}^{(L-2)}} = \frac{\partial z_k^{(L-2)}}{\partial w_{k,m}^{(L-2)}} \frac{\partial a_k^{(L-2)}}{\partial z_k^{(L-2)}} \frac{\partial J_1}{\partial a_k^{(L-2)}}. \tag{3.32}$$

The same principle applies for obtaining the partial derivative of the cost function with respect to the biases in the network,

$$\frac{\partial J_1}{\partial b^{(L)}} = \frac{\partial z^{(L)}}{\partial b^{(L)}} \frac{\partial J_1}{\partial z^{(L)}}, \tag{3.33}$$

where,

$$\frac{\partial z^{(L)}}{\partial b^{(L)}} = 1.$$

We can continue this sequence until we find the partial derivative of every weight and bias in our model. However, this is now only conducted for one iteration on one observation in our training data. To get the final gradient ∇J for each weight and bias, we evaluate the average for all the observations in our training data,

$$\frac{\partial J}{\partial W_k^{(L)}} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial J_i}{\partial W_k^{(L)}}, \tag{3.34}$$

where n is the number of observations. After we have solved the partial derivative of the cost function with respect to each weight and bias, we can now use these gradients in order to minimize the cost function by changing the specific weights and biases in the opposite sign of the corresponding gradient. This procedure is subsequently done over a specified interval of iterations to reach a global minimum for the cost function. This methodology of progressively calculating the gradient and further minimizing the cost function is also known as gradient decent and was first introduced as early as 1847, by mathematician Augustin Cauchy (Cauchy, 1847). We can define the gradient decent algorithm in one equation as,

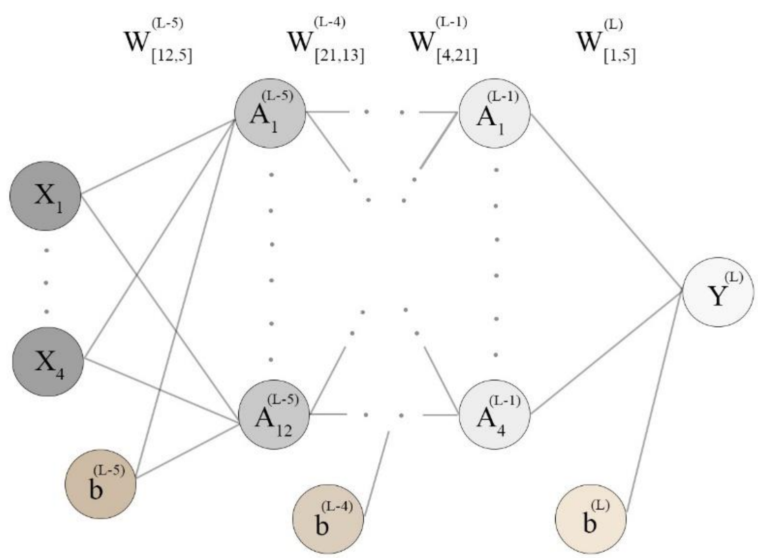
$$\Theta_{i+1} = \Theta_i - \alpha \nabla J(\Theta_i) \tag{3.35}$$

where Θ_i is the current value of the weight or bias and α is the designated learning rate set as a predetermined parameter. Thus, the learning rate regulates the adjustments to the weights and biases and subsequently controls how quickly the cost function converges to a minimum. For more on optimization and variations of gradient decent algorithm, see the works of Goodfellow et al. (2016).

The model presented in this thesis is constituted by 4 input nodes, one for each feature $\theta = \{K, S_t, \tau, r\}$, five hidden layers, first hidden layer $(L-4)$ is constituted by twelve activation nodes, the last hidden layer $(L-1)$ by four activation nodes and the intermediate layers by twenty nodes each. The activation function used for the hidden layers is the Tanh activation function, covered earlier. An overview of the architecture of the model is illustrated in *figure 2*. Further, we recognize the crucial challenge of not overfitting the training data when working with neural network models, meaning that the model maintains a too high degree of flexibility and tries to fit the function to every single data point, instead of generalizing the function over all data points (Goodfellow et al., 2016). Thus, to prevent overfitting, the model is restricted by a simple early stopping method. Essentially, early stopping prevents the data from iterating after a certain threshold for the loss function on the out-of-sample data has been fulfilled over a specified number of iterations.

To estimate the RND, we follow the same procedure as for the kernel regression in *section 3.3*. We estimate the implied volatility function, then proceed to calculate the option prices using the implied volatility function in the Black-Scholes model. Finally, to obtain the RND, we take the second partial derivative of the option pricing function with respect to the strike price.

FIGURE 2. **NEURAL NETWORK MODEL ARCHITECTURE**



Graphical representation of the neural network model's architecture. There is a total of seven layers, including the input layer and output layer. There are subsequently six layers of weights and biases connecting the intermediate layers.

4. Theoretical RND and Data-Generation

In this section we will go through the process of simulating the desired option data, define the different data-generation processes used in this thesis and discuss their different features and desirable characteristics. Further, we will specify how one can obtain the theoretical, or also referred to as the true RND from the different data-generation processes. Additionally, this section will also cover the parameters used for the data-generated processes and briefly present the descriptive statistics of the simulated data.

4.1 Data-Generation Processes

This thesis is conducted with the intention of comparing the efficiency and robustness of the three different models presented in *section 3* for obtaining the option implied RND, which are generated under different data-generation processes. The first data-generation process that we will cover is assumed to follow a geometric Brownian motion, identical to the underlying process in the Black-Scholes model (Black & Scholes, 1973). The second process is a diffusion process with stochastic volatility, presented by Heston (1993). The third process is based on the jump diffusion model proposed by Merton (1976). Lastly, the fourth process is a diffusion process with both stochastic volatility and jump(s), introduced by Bakshi et al. (1997). The different data-generating methods capture certain features associated with observed, historical stock returns, such as skewed and leptokurtic probability distributions. Further, by generating the data ourselves through Monte Carlo simulations, we have knowledge of the underlying parameters used in the data-generation processes and, thus, we can obtain the true RNDs by utilizing the close-form option pricing formulas incorporating the different data-generation processes. The true RND will subsequently be used to evaluate the performance of the proposed models. The RNDs generated for these tests are generally dependent on multiple factors. However, for the purpose of this thesis we will only focus on the RND as a function of the terminal spot price S_T.

4.1.1 Geometric Brownian Motion

The data-generation process for the underlying asset assumed by Black and Scholes (1973) follows a geometric Brownian motion with the following notation,[10]

$$dS_t = rS_t dt + \sigma S_t dW_t, \qquad (4.1)$$

[10] For simplicity, we will refer to the geometric Brownian motion as simply the Brownian motion in the remaining sections of this thesis.

where S_t constitutes the price process of the underlying asset with constant volatility σ, r is the constant drift under risk-neutral measures, where the underlying is not paying any dividends and W_t is a standard Wiener process. To generate our option prices through Monte Carlo simulations, we need to define the stock process in discrete time. We can do that by applying Ito's formula to the stochastic differential equation *(4.1)*.[11] Hence, we get the stock price process in terms of $\ln(S)$ given as,

$$dln(S_t) = \left(r - \frac{\sigma^2}{2}\right)dt + \sigma dW_t. \tag{4.2}$$

We can further integrate the equation and hence obtain the analytic solution for the stock process as follows,

$$S_t = S_0 e^{(r-0.5\sigma^2)t+\sigma W_t}. \tag{4.3}$$

Following the Brownian motion data-generation process, we can define the corresponding option pricing formula according the Black-Scholes model, where the European call option c with time to maturity $\tau = T - t$, assuming the underlying is not paying any dividends, with strike K and constant volatility σ and risk-free rate r, as follows,

$$c_{BS}(S_t, K, \tau, r, \sigma) = S_t N(d_1) - Ke^{-r\tau} N(d_2)$$

$$d_1 = \frac{ln\left(\frac{S_t}{K}\right) + (r + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}$$
$$d_2 = d_1 - \sigma\sqrt{\tau}. \tag{4.4}$$

Following the approach of Aït-Sahalia and Lo (1998), who showed that by acknowledging the findings of Breeden and Litzenberger (1978) and *equation* (*2.13)*, one can conclude that the RND obtained from the Black-Scholes formula is simply a log-normal density with mean $\left[\left(r + \frac{1}{2}\sigma^2\right)\tau\right]$ and variance $(\sigma^2\tau)$ for $\ln\left(\frac{S_T}{S_t}\right)$. Thus, we can evaluate the RND for S_T, when $K = S_T$, as follows,

$$q_{BS}(S_T) = e^{r\tau}\frac{\partial^2 c}{\partial K^2}\bigg|_{K=S_T}$$
$$= \frac{1}{S_T\sqrt{2\pi\sigma^2\tau}} exp\left[-\frac{\left[ln\left(\frac{S_T}{S_t}\right) - \left(r + \frac{1}{2}\sigma^2\right)\tau\right]^2}{2\sigma^2\tau}\right]. \tag{4.5}$$

[11] See Hull (2017) for a complete derivation of the SDE.

4.2.2 Stochastic Volatility

The diffusion process with stochastic volatility, introduced by Heston (1993), adds a stochastic volatility component to replicate the non-constant volatility observed in markets, here referred to as the variance process. Thus, Heston proposed the following stochastic differential equation, with time-varying volatility as,

$$
\begin{aligned}
dS_t &= rS_t dt + \sqrt{V_t} S_t dW_{1,t}, \\
dV_t &= \kappa(\theta - V_t)dt + \sigma_v \sqrt{V_t} dW_{2,t}, \\
cov\left(dW_{1,t}, dW_{2,t}\right) &= \rho dt,
\end{aligned}
\tag{4.6}
$$

where S_t is the price process, V_t is the variance process, which is a mean reverting process with θ as the long-run mean value and κ is the reversion rate towards the mean. r denotes the constant drift under risk-neutral measures, σ_v is the corresponding volatility of the variance process. Further, the additional stochastic Wiener process $W_{2,t}$, is correlated with the other Wiener process $W_{1,t}$ by parameter ρ. By changing ρ we can replicate the negative correlation between the asset returns and volatility, and thus, obtain skewness and leptokurtosis. An example of the specific features can be observed in *figure 3b*. The figure illustrates the difference between a normal distribution and a risk-neutral density estimated based on a stochastic volatility process for an option with a volatility of 20% and with one month until maturity. The density based on the stochastic volatility process has a skewness of -0.77 and kurtosis of 3.56, while the normal distribution has a skewness of 0 and kurtosis of 3.

For the purpose of our Monte Carlo simulations, we need to define the stock process in discrete time. Thus, we incorporate the Euler scheme (Rouah, 2011) to derive a discrete solution for both the variance and stock price processes. The Euler approach defines the Wiener process in discrete time as $\Delta W_t = W_{t+dt} - W_t$, which is equal in distribution to $\sqrt{dt}Z$, where $Z \sim N(0,1)$. According to the Euler scheme, the variance process can be written as follows,

$$V_{t+dt} = V_t^+ + \kappa(\theta - V_t)dt + \sigma_v \sqrt{V_t^+ dt} Z_2, \tag{4.7}$$

where $Z_2 \sim N(0,1)$ and $V_t^+ = \max(0, V_t)$. The Euler scheme further defines the stock price process as,

$$S_{t+dt} = S_t exp\left[\left(r - \frac{1}{2}V_t\right)dt + \sqrt{V_t dt} Z_1\right], \tag{4.8}$$

where $Z_1 = \rho Z_2 + \sqrt{1 - \rho^2} Z_3$, with Z_3 being equivalent to a standard normal variable, $Z_3 \sim N(0,1)$.

However, to obtain the true RND, one needs to define the option pricing model as a closed-form solution. Heston (1993) stated that the stochastic differential *equation* (*4.6*) can be solved by using a Fourier transform. He showed that the associated price of a European call option $c_{SV}(S_t, K, \tau, V_t)$ under a diffusion process with stochastic volatility, where the underlying is not paying any dividends and maturing at time $T = \tau + t$, is defined as,

$$c_{SV}(S_t, K, \tau, V_t) = S_t P_1 - K e^{-r_t \tau} P_2, \tag{4.9}$$

where,

$$\begin{aligned}
P_j(S_t, V_t, \tau; ln(K)) &= \frac{1}{2} + \frac{1}{\pi}\int_0^\infty Re\left[\frac{e^{-i\phi\, ln(K)} f_j(S_t, V_t, \tau; \phi)}{i\phi}\right] d\phi,\\
f_j(S_t, V_t, \tau, \phi) &= exp\left[C_j(\tau; \phi) + D_j(\tau; \phi)V_t + i\phi\, ln(S_t)\right],\\
C_j(\tau; \phi) &= ir\phi\,\tau + \frac{a}{\sigma_v^2}\left[\left(b_j - i\rho\sigma_v\phi + d_j\right)\tau - 2\, ln\left(\frac{1 - g_j e^{d\tau}}{1 - g_j}\right)\right],\\
D_j(\tau; \phi) &= \frac{b_j - i\rho\sigma_v\phi + d_j}{\sigma_v^2}\left(\frac{1 - e^{d\tau}}{1 - g_j e^{d\tau}}\right),\\
g_j &= \frac{b_j - i\rho\sigma_v\phi + d_j}{b_j - i\rho\sigma_v\phi - d_j},\\
d_j &= -\sqrt{\left(i\rho\sigma_v\phi - b_j\right)^2 - \sigma_v^2\left(2iu_j\phi - \phi^2\right)},
\end{aligned} \tag{4.10}$$

for $j = 1{,}2$ and where,

$$u_1 = \frac{1}{2}, \qquad u_2 = -\frac{1}{2}, \qquad a = \kappa\theta, \qquad b_1 = \kappa + \lambda - \rho\sigma, \qquad b_2 = \kappa + \lambda, \qquad RE[\cdot] \in \mathbb{R}.$$

Equation (*4.10*) is slightly different compared to the model introduced by Heston (1993), where we add a minus sign in front of the variable d_j. Albrecher, Mayer, Schoutens, and Tistaert (2007) showed that the new solution does not affect the option price, but instead the model now produces stable results for the whole parameter space. We can evaluate the RND for S_T, when $K = S_T$, by taking the partial derivative of *equation* (*4.9)* with respect to the strike price K twice, and thus end up with the following solution for the RND function,

$$q_{SV}(S_T) = \frac{1}{\pi}\int_0^\infty Re\left[\frac{e^{-i\phi\, ln(K)} f_2}{S_T}\right] d\phi\,\Bigg|_{K=S_T}, \tag{4.11}$$

where $f_2(S_t, V_t, \tau, \phi)$ is defined as,

$$f_2(S_t, V_t, \tau, \phi) = exp[C_2(\tau; \phi) + D_2(\tau; \phi)V_t + i\phi\, ln(S_t)]. \tag{4.12}$$

4.2.3 Jump Diffusion

The diffusion process with jumps proposed by Merton (1976), compared to the original Black-Scholes model, allows the underlying asset to have random jump dynamics and can thus replicate the notorious implied volatility smile and subsequently fatter tails and excess kurtosis in the return distribution. Merton's inclusion of jumps to the stock's diffusion process gave rise to a new set of models incorporating unpredictability and random events, more known as jump-diffusion models. Here, Merton assumed that the average interval between and the size of the jumps in spot prices are known, but the exact timing and size of each individual jump is random. Thus, Merton proposed that the jumps follow a compounded Poisson process. He defined the stochastic differential equation as follows,

$$\begin{gathered} dS_t = \left(r - r_J\right)S_t dt + \sigma_t S_t dW_t + S_t dJ_t, \\ r_J = \lambda\left(e^{\mu_J + \delta^2/2} - 1\right), \end{gathered} \tag{4.13}$$

where $\mathrm{S_t}$ is the asset price process, r is the constant drift under risk-neutral measures and σ_t is the volatility of the underlying asset. r_J is the drift correction term for the jump, where λ is the number of jumps annually, μ_J is the expected jump size, δ is the volatility of the jump and W_t is a standard Wiener process. The jump event, which is denoted as the third part in *equation (4.13)*, is determined by the compounded Poisson process J_t,

$$J_t = \sum_{j=1}^{N_t} (Y_j - 1), \tag{4.14}$$

where N_t is the Poisson process with intensity λ and Y_j is the random jump size, where $Y_j \sim N(\mu_J, \delta^2)$ and $\{Y_j\}_{j\geq 1}$ is denoted a sequence of independent random variables. Thus, one can define an average jump size, but at the same time allow for the jumps to remain independent and randomly distributed with the mean μ_J and standard deviation δ^2. Finally, we acknowledge that the processes W_t, Y_j and N_t are presumed to be independent from each other.

The next step is to define the model in discrete time for our Monte Carlo simulations. The discretization of the jump diffusion process, following the Euler discretization scheme, where the Wiener process is defined in discrete time as $\Delta W_t = W_{t+dt} - W_t$, which is equal in distribution to $\sqrt{dt}Z$, where $Z \sim N(0,1)$, is defined as,

$$S_{t+dt} = S_t\left\{exp\left[\left(r - r_j - \frac{1}{2}\sigma^2\right)dt + \sigma\sqrt{dt}Z_1\right] + \left(exp\left[\mu_J + \delta^2 + Z_2\right] - 1\right)y_t\right\}, \tag{4.15}$$

where y_t is a Poisson distributed variable with intensity λ, Z_1 and Z_2 are standard normal variables and y_t, Z_1 and Z_2 are assumed to be independent from each other (Hilpisch, 2015).

Further, we want to estimate the true RND for the jump diffusion process utilizing a closed-form solution. However, for models with discrete compounding, closed-form solutions are generally impossible to derive (Kou, 2002). Thus, we will have to resort to approximation methods to obtain the true RND. For the purpose of this thesis, we will not cover the lengthy derivation of the jump diffusion model. For a thorough derivation of the Merton jump diffusion stochastic differential equation, see the works of Cont and Tankov (2004), Applebaum (2009) and Tankov and Voltchkova (2009). Following the methodology presented by Cont and Tankov (2004), who applied Itô's formula to the stochastic differential equation *(4.13)* provided the following final solution for the $\ln(S_t)$ process as follows,

$$ln(S_t) = ln(S_0) + \sigma W_t + \left(r - r_J - \frac{1}{2}\sigma^2\right)t + \sum_{j=1}^{N_t} ln(Y_j), \tag{4.16}$$

where the process evolves like a Brownian motion between jumps and after each jump the value of $ln(S_t)$ is multiplied by e^{Y_j} (Shonkwiler, 2013). Further, if we consider the log-returns, $ln\left(\frac{S_t}{S_0}\right)$, for the process to be conditional on the event $N_t = i$, we can rewrite the log-return process as,

$$ln\left(\frac{S_t}{S_0}\right) = \sigma W_t + \left(r - r_J - \frac{\sigma^2}{2}\right)t + \sum_{j=1}^{i} ln(Y_j) \sim N\left[\left(r - r_J - \frac{\sigma^2}{2}\right)t + \mu_J, \sigma^2 t + i\delta^2\right]. \tag{4.17}$$

Thus, we can now define the RND of log returns $q_{JD}(x_t)$, where $x_t = \ln\left(\frac{S_t}{S_0}\right)$, to be estimated utilizing a converging series, defined as follows,

$$q_{JD}(x_t) = \sum_{i=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^i}{i!} N\left\{x_t; \left(r - r_J - \frac{1}{2}\sigma^2\right)t + i\mu_J, \sigma^2 t + i\delta^2\right\}, \tag{4.18}$$

where,

$$N\left\{x_t; \left(r - r_J - \frac{1}{2}\sigma^2\right)t + i\mu_J, \sigma^2 t + i\delta^2\right\}$$
$$= \frac{1}{\sqrt{2\pi(\sigma^2 t + i\delta^2)}} exp\left\{-\frac{\left[x_t\left\{\left(r - r_J - \frac{1}{2}\sigma^2\right)t + i\mu_J\right\}\right]^2}{2(\sigma^2 t + i\delta^2)}\right\}. \quad (4.19)$$

Thus, the RND is expressed as a weighted sum of normal densities, with Poisson probability mass function $P(i) = \frac{e^{-\lambda t}(\lambda t)^i}{i!}$, representing the probability of the asset jumping i times during the time interval $(0; t]$.

4.2.4 Stochastic Volatility and Jumps

Lastly, we will cover the diffusion process with stochastic volatility and jumps introduced by Bakshi et al. (1997). The original model is described as a unifying option pricing model, incorporating stochastic volatility, jump(s) and stochastic interest rate. For the purpose of this thesis, we will only cover the stochastic volatility and jump(s), excluding the stochastic interest rate. Bakshi et al. (1997) defined the stochastic differential equation as follows,

$$dS_t = (r - r_J)S_t dt + \sqrt{V_t} S_t\, dW_{1,t} + S_t dJ_t,$$
$$dV_t = \kappa(\theta - V_t)dt + \sigma_v \sqrt{V_t d} W_{2,t},$$
$$r_J = \lambda\left(e^{\mu_J + \delta^2/2} - 1\right),$$
$$J_t = \sum_{j=1}^{N_t} (Y_j - 1),$$
$$cov\left(dW_{1,t}, dW_{2,t}\right) = \rho dt, \quad (4.20)$$

where S_t is the price process, V_t is the variance process, r is the constant drift under risk-neutral measures, $W_{1,t}$ and $W_{2,t}$ are Wiener processes with correlation parameter ρ. θ is the long-run volatility, κ is the reversion rate towards the long-run volatility and σ_v is the corresponding volatility of the variance process. r_J is the drift correction term for the jump, where λ is the number of jumps annually, μ_J is the expected jump size, δ is the volatility of the jump, J_t is the compounded Poisson process, where N_t is a standard Poisson process with intensity λ and Y_j is the random jump size, where $Y_j \sim N(\mu_J, \delta^2)$ and $\{Y_i\}_{i\geq 1}$ is a sequence of independent random variables. Further, we state that N_t, V_t, Y_j are independent from each other, as well as independent from $W_{1,t}$ and $W_{2,t}$.

For the Monte Carlo Simulations we define the stock price process following a stochastic volatility with jump diffusion in discrete time. Similarly to earlier processes, we define the stock process in according to the Euler discretization scheme, where we define the Wiener process in discrete time as $\Delta W_t = W_{t+dt} - W_t$, which is equal in distribution to $\sqrt{dt}Z$, where $Z \sim N(0,1)$. According to the Euler scheme, the discrete stock process S_t and variance process V_t for the stochastic volatility and jump diffusion process can be evaluated as follows,

$$\begin{aligned} S_{t+dt} = S_t \Big\{ exp\left[\left(r - r_J - \frac{1}{2}V_t\right)dt + \sqrt{V_t dt}Z_1\right] \\ + \left(exp\left[\mu_J + \delta^2 + Z_4\right] - 1\right)y_t \Big\}, \\ V_{t+dt} = V_t^+ + \kappa(\theta - V_t)dt + \sigma_v \sqrt{V_t^+ dt}Z_2, \end{aligned} \tag{4.21}$$

where y_t is a Poisson distributed variable with intensity λ, $Z_1 = \rho Z_2 + \sqrt{1-\rho^2}Z_3$, where Z_2, Z_3 and Z_4 are standard normal variables and y_t, Z_3 and Z_4 are assumed to be independent from each other, as well as from Z_1 and Z_2 (Hilpisch, 2015).

The solution for the stochastic differential equation, that has to be solved in order to obtain a closed-form solution and subsequently the RND, is similar to that of the stochastic volatility process proposed by Heston (1993). Bakshi et al. (1997) solved *equation (4.20)* utilizing the Fourier transform method and worked out the following pricing model for a European call option $c_{SVJD}(S_t, K, \tau, V_t)$, assuming not paying any dividends, and maturing at time $T = \tau + t$, as follows,

$$c_{SVJD}(S_t, K, \tau, V_t) = S_t P_1 - K e^{-r_t \tau} P_2, \tag{4.22}$$

where,

$$P_j(S_t, V_t, \tau; ln(K)) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty Re\left[\frac{e^{-i\phi\, ln(K)} f_j(S_t, V_t, \tau; \phi)}{i\phi}\right] d\phi,$$

$$f_j(S_t, V_t, \tau, \phi) = exp\left[C_j(\tau;\phi) + D_j(\tau;\phi)V_t + E_j(\tau;\phi) + i\phi\, ln(S_t)\right],$$

$$C_j(\tau;\phi) = ir\phi\,\tau + \frac{a}{\sigma_v^2}\left[\left(b_j - i\rho\sigma_v\phi + d_j\right)\tau - 2\,ln\left(\frac{1 - g_j e^{d\tau}}{1 - g_j}\right)\right],$$

$$D_j(\tau;\phi) = \frac{b_j - i\rho\sigma_v\phi + d_j}{\sigma_v^2}\left(\frac{1 - e^{d\tau}}{1 - g_j e^{d\tau}}\right),$$

$$\begin{aligned} E_j(\tau;\phi) = \lambda\left(1 + \mu_J\right)\tau\left[\left(1 + \mu_J\right)^{i\phi} exp\left[(i\phi/2)(1 + i\phi)\delta_J^2\right] - 1\right] \\ - \lambda i\phi\mu_J\tau, \end{aligned}$$

$$g_j = \frac{b_j - i\rho\sigma_v\phi + d_j}{b_j - i\rho\sigma_v\phi - d_j},$$

$$d_j = -\sqrt{(i\rho\sigma_v\phi - b_j)^2 - \sigma_v^2(2iu_j\phi - \phi^2)}, \tag{4.23}$$

for $j = 1,2$, where,

$$u_1 = \frac{1}{2}, \qquad u_2 = -\frac{1}{2}, \qquad a = \kappa\theta, \qquad b_1 = \kappa - \rho\sigma_v, \qquad b_2 = \kappa, \qquad RE[\cdot] \in \mathbb{R}.$$

Applying *equation (4.23)* to *equation* (*4.22)*, and by differentiating twice with respect to the strike price K, where $K = S_T$, we obtain the true RND for the option pricing model as follows,

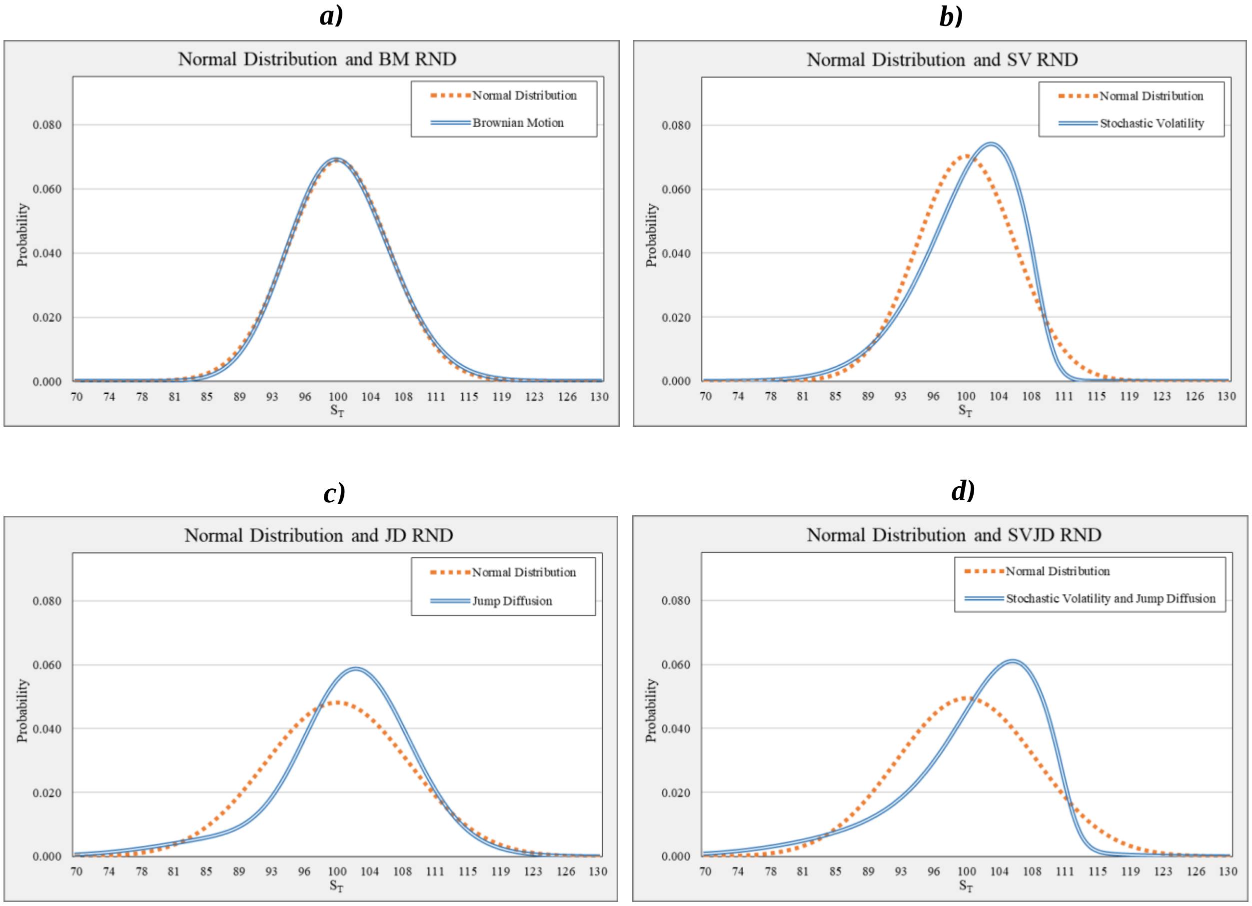
$$q_{SVJD}(S_T) = \frac{1}{\pi}\int_0^\infty Re\left[\frac{e^{-i\phi\, ln(K)}f_2}{S_T}\right]d\phi\Bigg|_{K=S_T} \tag{4.24}$$

where f_2 is defined as,

$$f_2(S_t, V_t, \tau, \phi) = exp[C_2(\tau;\phi) + D_2(\tau;\phi)V_t + E_2(\tau;\phi) + i\phi\, ln(S_t)] \tag{4.25}$$

The closed-form solution for the diffusion process with stochastic volatility and jump(s) is rather similar to the solution for the option price model proposed by Heston (1993) with only stochastic volatility. The only difference is the function $f_j(S_t, V_t, \tau, \phi)$ and the added characteristics function $E(\tau;\phi)$, approximating for the accommodating jump diffusion. Similar to *equation (4.10)* in *section 4.2.2*, we incorporate a minus sign for d_j in *equation (4.23)* for it to be able to produce more stable results.

***FIGURE 3*. DATA-GENERATION PROCESSES RND VS. NORMAL DISTRIBUTION**



The figures illustrates the different data-generation processes and their corresponding RND to a normal distribution, where both densities in each graph have the same mean and standard deviation. The RNDs are generated from a range of options with strike price ranging from 70 to 130, with a underlying spot price of a 100. The time to maturity is 21 days and the annualized volatility is 20%.

4.3 Data Overview

The parameters associated with the data-generating processes are based on the historical option prices deriving from the Standard and Poor's 500 (S&P 500) index. The spot price (S_0), is the underlying S&P 500 index price on December 29th, 2017. The risk-free rate (r) is based on the one-month U.S. treasury bill observed on December 29th, 2017. The underlying volatility (σ) used is based in the seven-year average, annualized volatility of the S&P 500 from December 29th, 2010 to December 29th, 2017. This period is considered to be a quite stable period and hence we obtain a relatively low volatility of 16.6%. The annualized volatility of the volatility (σ_v) is assumed to be 40%.[12] The long-run volatility ($\sqrt{\theta}$) is further assumed to be 20%. The mean reversion rate (κ) for the underlying volatility towards the long run volatility is assumed to be 1. The correlation (ρ) between volatility and returns is assumed to be negative, -0.80, to reflect the leverage effect observed in spot markets. The annual jump frequency (λ) is assumed to occur 1 time annually, while the jump size (μ_j) is -10% to emulate negative market jumps. Lastly the volatility of the jumps (Δ) is assumed to be 10%.[13] An overview of the parameters is presented in *table 1*. Our assumptions regarding the parameters are mainly based on the desired characteristics we are looking for in of our simulated RNDs in terms of skewness and kurtosis.

To verify the performance and robustness of the models, we simulated the option prices over three different maturities: one month, three months and six months, or more specifically 21, 63 and 126 trading days, where there are 252 trading days annually. Further, to obtain a sufficient number of option prices from our Monte Carlo simulations, we simulated options on different strike prices, ranging from 1875 to 3475, with an interval of 25, resulting in 65 option prices for each maturity and data-generation process. We further simulated each set of option prices, per maturity, 100 times to obtain a total of 6,500 option prices for each maturity, thus we end up with total of 19,500 call and put options respectively for each data-generation process. Each option price was simulated using 250 paths and averaged over 100,000 different iterations. To emulate real world mispricing, due to a limit set on the smallest possible tick size for quoted prices, we added a small error term to each simulated option price that consists of a small random number between -0.5 and 0.5 of the smallest tick size allowed for SPX options.[14] Further, we recognized that the Monte Carlo simulations generated unrealistically low implied volatilities for relatively deep out-of-the money and deep in-the-money options with low maturities. Thus, to address these underpriced options, we utilized the put-call parity to price them. The results of the True RNDs and estimated RNDs are all given in terms of simple moneyness $\xi = S_T / S_t$, where S_T is the terminal spot price and S_t is the spot price at time $t = 0$.

[12] The volatility of the volatility over the relatively stable period from December 29th, 2010 to December 29th, 2017 was roughly 35% and thus, we consider our assumption of 40% to be suitable.
[13] Notice the change in notation here compared to earlier sections, due to the introduction of the dividend notation.
[14] Minimum tick for options trading below 3.00 is 0.05 and for all other series 0.10.

***TABLE 1.* PARAMETERS FOR THE DATA-GENERATION PROCESSES**

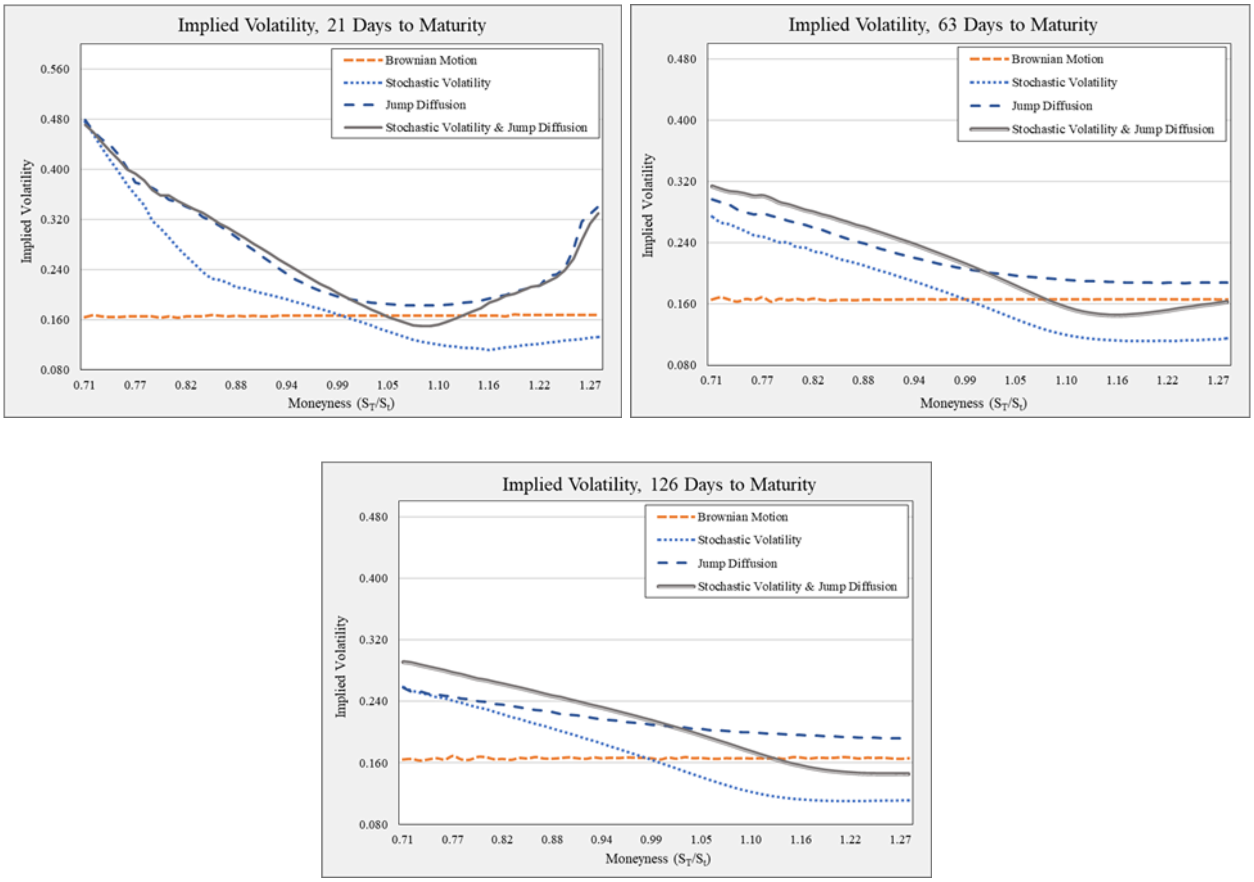
Summary of the parameters used for the four different data-generation processes, where BM is the Brownian motion process, SV is the stochastic volatility process, JD is the jump diffusion process and SVJD is the stochastic volatility and jump diffusion process.

Parameters	BM	SV	JD	SVJD
S_0	2673.61	2673.61	2673.61	2673.61
r_f	0.01299	0.01299	0.01299	0.01299
δ	0.00	0.00	0.00	0.00
σ	0.166	0.166	0.166	0.166
σ_v		0.40		0.40
$\sqrt{\theta}$		0.20		0.20
κ		1.00		1.00
ρ		-0.80		-0.80
λ			1.00	1.00
μ_j			-0.10	-0.10
Δ			0.10	0.10

4.3.1 Implied Volatility and True RND

The simulated average implied volatilities generated through Monte Carlo simulations from the four different data generation processes and the three maturities 21 days, 63 days and 126 days are presented in *figure 4*. We can observe that the Heston (1993) SV data-generation process shows a negative correlation between implied volatility and asset returns, to reflect the leverage effect observed in the markets. The effect of the negative correlation between asset returns and implied volatilities can further be observed in *figure 5* and *table 2*, where we recognize the negative skew for the densities generated from the SV process and Bakshi et al. (1997) SVJD process for all maturities. However, the negative skewness can likewise be observed for densities generated by the Merton (1976) JD process for options with shorter maturities, due to the negative jump component, inducing a negative correlation between implied volatilities and asset returns. Further, in *figure 4* we can observe the distinct volatility skew generated by the JD process as well as the SVJD process. Thus, the densities generated from the two processes also exhibit significantly higher kurtosis for the two processes compared to the SV and BM processes for the options with 21 days to maturity, seen in *table 2*. However, the implied volatility skew diminished for the JD and SVJD process over longer maturities, and thus the densities exhibit lower kurtosis for options with longer maturities. The BM data-generation process exhibits constant implied volatility over the entire range of strike prices and maturities, therefore we can further observe the log-normal density features in *table 2*, where the skewness is positive over all maturities and the density is characterized by a kurtosis close to zero.

***FIGURE 4.* SIMULATED IMPLIED VOLATILITIES**



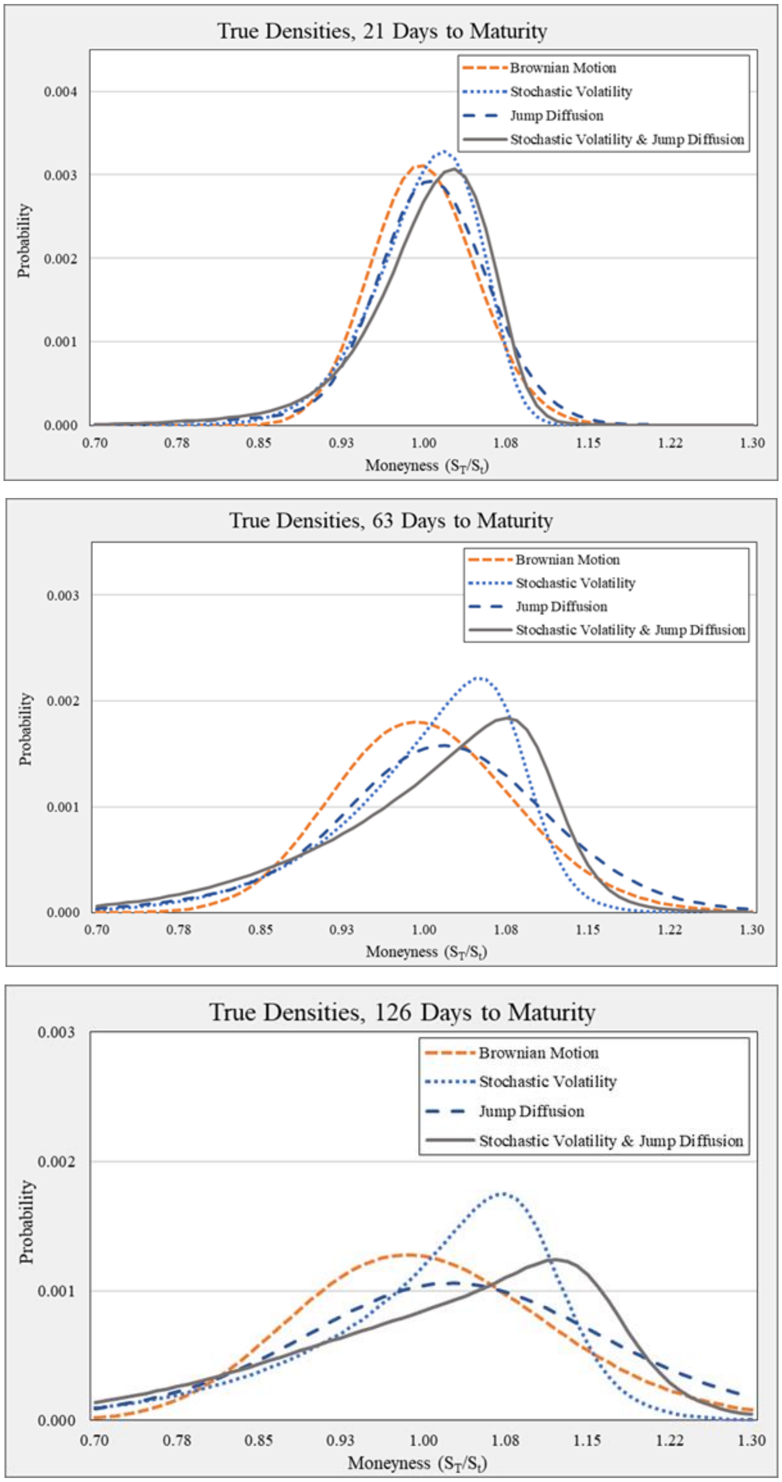
The average implied volatilities simulated by the four different data-generation processes: Brownian motion, stochastic volatility, jump diffusion and stochastic volatility with jump diffusion, over three maturities: 21, 63 and 126 days.

***TABLE 2.* SUMMARY STATISTICS, DATA-GENERATION PROCESSES**

Summary statistics for the four different data-generation processes following a geometric Brownian motion (BM), stochastic volatility (SV), jump diffusion (JD) and stochastic volatility with jump diffusion (SVJD) for 21, 63 and 126 days to maturity. The mean and standard deviation (Std. Dev.) is given in terms of moneyness (S_T/S_t), while skewness and kurtosis are reported in units of statistical moments. For reference, a Gaussian distribution is constituted by a skewness of 0 and a kurtosis of 3.

21 Days to Maturity				
Statistics	BM	SV	JD	SVJD
Mean	1.0011	1.0010	1.0010	1.0025
Std. Dev.	0.0480	0.0477	0.0605	0.0593
Skewness	0.1440	-0.6257	-0.8174	-1.1916
Kurtosis	3.0369	3.3117	5.8663	5.9590
63 Days to Maturity				
Statistics	BM	SV	JD	SVJD
Mean	1.0031	1.0028	1.0014	1.0059
Std. Dev.	0.0833	0.0819	0.1037	0.1009
Skewness	0.2461	-0.8658	-0.1702	-0.7177
Kurtosis	3.0817	3.8996	3.3941	3.4309
126 Days to Maturity				
Statistics	BM	SV	JD	SVJD
Mean	1.0009	1.0014	0.9871	1.0049
Std. Dev.	0.1157	0.1092	0.1399	0.1362
Skewness	0.3460	-0.5739	0.3564	-0.1627
Kurtosis	2.9146	3.2865	2.7304	2.5206

***FIGURE 5.* THEORETICAL RISK-NEUTRAL DENSITIES**



Graphical representation of the true RNDs generated from the four different data-generation processes for three maturities: 21, 63 and 126 days.

5. Empirical Evidence

In this section we present the option implied RNDs estimated by the mixture, kernel regression and neural network model, and compare them with the true RNDs given by the closed-form solutions of the four different data-generation processes. In *section 5.1,* we cover the summary statistics of the first four statistical moments of the RNDs. In *section 5.2,* we investigate the statistical differences between the densities through the nonparametric Mann-Whitney U test. In *section 5.3,* we examine the estimated RND functions' performance on pricing different derivatives. Lastly, in *section 5.4,* we further evaluate the estimated RND's performance on real market data.

To estimate the RNDs, each model is trained on the same set of option prices for a specific data-generation process over the three maturities. Further, each model estimates the risk-neutral density for each maturity 100 times to accommodate for the different optimization outcomes. We then calculate the average of the 100 different RND functions to subsequently end up with the final estimated RND for each maturity. This process is done for each data-generation process, leaving us with a grand total of twelve RNDs for each model.

5.1 Summary Statistics

The summary statistics of the RNDs are an essential measurement for the analysis of the performance of the models presented in this thesis. To effectively measure the efficiency, we evaluate the estimated RNDs' first four statistical moments in terms of *mean, standard deviation, skewness* and *kurtosis* compared to the statistical moments of the true RNDs.[15] The summary statistics results are presented in terms of moneyness, $\xi = S_T / S_t$.

Table 3 presents the summary statistics of the Brownian motion data-generation process, where the implied volatility is assumed to be constant for all maturities and across strike price, resulting in log-normal RNDs. As can be interpreted from *table 3,* accompanied by *figure 6,* one can argue that each model can effectively estimate the RND for all maturities, if the asset's return probability distribution follows a log-normal distribution.

The summary statistics for the stochastic volatility process can be viewed in *table 4*, as well as a graphical representation of the results in *figure 7.* As can be observed from *table 4*, the most prominent result is the extensive excess kurtosis estimated by the mixture model for both 21 and 63 days to maturity.

[15] See appendix *b. Computation of Moments* for a brief overview of the statistical moments.

Thus, we can conclude that the mixture model assumes that the data is characterized by an implied volatility smile or skew and estimates more probability to the more extreme events far out in the tails of the density. However, illustrated in *figure 4*, we can observe that the implied volatility is rather downward sloping than exhibit the shape of a distinct smile or skew. Further, we recognize that both the mixture model and the kernel regression model underestimate the excess skewness generated from the negative relationship between asset returns and implied volatility for the shorter maturities. Meanwhile, we can observe that the kernel regression model achieves the most similar results to the true RND for the 126 days to maturity. Meanwhile, the neural network performs reasonably well for the two shorter maturities compared to the other two models.

Table 5 and corresponding *figure 8* represents the summary statistics for the true RND and estimated RNDs following the jump diffusion data-generation process. Observing *table 5*, we see that the results are varying in terms of performance over time for the different models. The mixture model manages to capture the kurtosis for all three maturities, but again underestimates the negative skewness for all maturities. The kernel model performs sufficiently well for the shorter maturities compared to the other models. However, for the 126 days to maturity we notice that the kernel model underestimates the positive skewness and further exhibits an inaccurately estimated right tail for the density illustrated in *figure 8c*. The neural network cannot accurately estimate the RND for the two shorter maturities, while redeeming itself for the 126 days maturity RND. Lastly, we can observe that the absolute percentage error is considerably higher for all models for the 63 days to maturity RNDs.

Finally, *table 6* represents the summary statistics for the true and estimated RNDs based on the stochastic volatility and jump diffusion data-generation process. The most prominent findings observed in *figure 9a*, is the peaked distribution predicted by the mixture model. However, interpreting the results presented in *table 6*, we notice that the mixture model has a significantly lower kurtosis than the true RND. To understand the relationship between the kurtosis and shape of the mixture model's density, we examine the visual representation of the cumulative probability distribution of the left tail for the 21 days to maturity RNDs in *figure 10*. We see that the mixture model has a considerably thinner left tail than the other densities, and as discussed in the earlier sections, kurtosis is defined by the tail distributions, where fatter tails result in higher kurtosis. Thus, we can conclude that the mixture model cannot sufficiently capture the prominent volatility skew presented in *figure 4* and subsequently estimates less probability for the extreme events. Continuing, both the kernel regression and neural network model are able to sufficiently capture the features of the RND for the options with 21 days to maturity. For the 63 days to maturity, we observe that the neural network model performs adequately well compared to the other models. On the other hand, illustrated in *figure 9b*, the kernel model displays a volatile pattern for the left tail of the distribution, indicating that the model is most likely overfitting the data for $\xi < 0.85$.[16] However, for the longer maturity of 126 days, we see that all three models struggle to fit the RND properly.

As a conclusion, we recognize that the mixture model struggles to estimate the RND for shorter maturities, where the implied volatility skew is most significant. Further, the mixture model tends to underestimate the negative skewness features characterizing the true RND. These findings are aligned with Cooper (1999), who made similar conclusions regarding the mixture model. The Kernel regression model performs consistently over all data-generation methods and maturities, and we find that the model has considerably lower absolute percentage errors, especially for the two jump diffusion processes and can thus estimate the moments of the RND more precisely than the other two models The results are further consistent with the findings of Lai (2014). On the other hand, we acknowledge that the kernel model tends to underestimate the significance of the skewness on a couple of occasions. In particular, we observe that the kernel model further shows signs of overfitting the data on one occasion, thus one suggestion would be to revise the bandwidth optimization. The neural network performs reasonably well for all data-generation processes, but is lacking in precision, especially for the RNDs with longer maturities. Lastly, based on *table 7*, we can determine that the kernel regression model best replicates the summary statistics of the true RND over all maturities and data-generation processes by a small margin to the neural network model. However, the neural network model is superior compared to the other two models when it comes to estimating the skewness of the RND.

[16] See *section 3.4* for a brief explanation on overfitting.

TABLE 3. SUMMARY STATISTICS BROWNIAN MOTION

Summary statistics for the mixture model, kernel regression and neural network compared to the true RND for the Brownian motion data-generation process with 21, 63 and 126 days to maturity. The mean and standard deviation is given in terms of moneyness (S_T/S_t). The absolute percentage error, calculated as the difference between the estimated and true statistic times 100, is presented in parentheses with percentage points as unit of measurement.

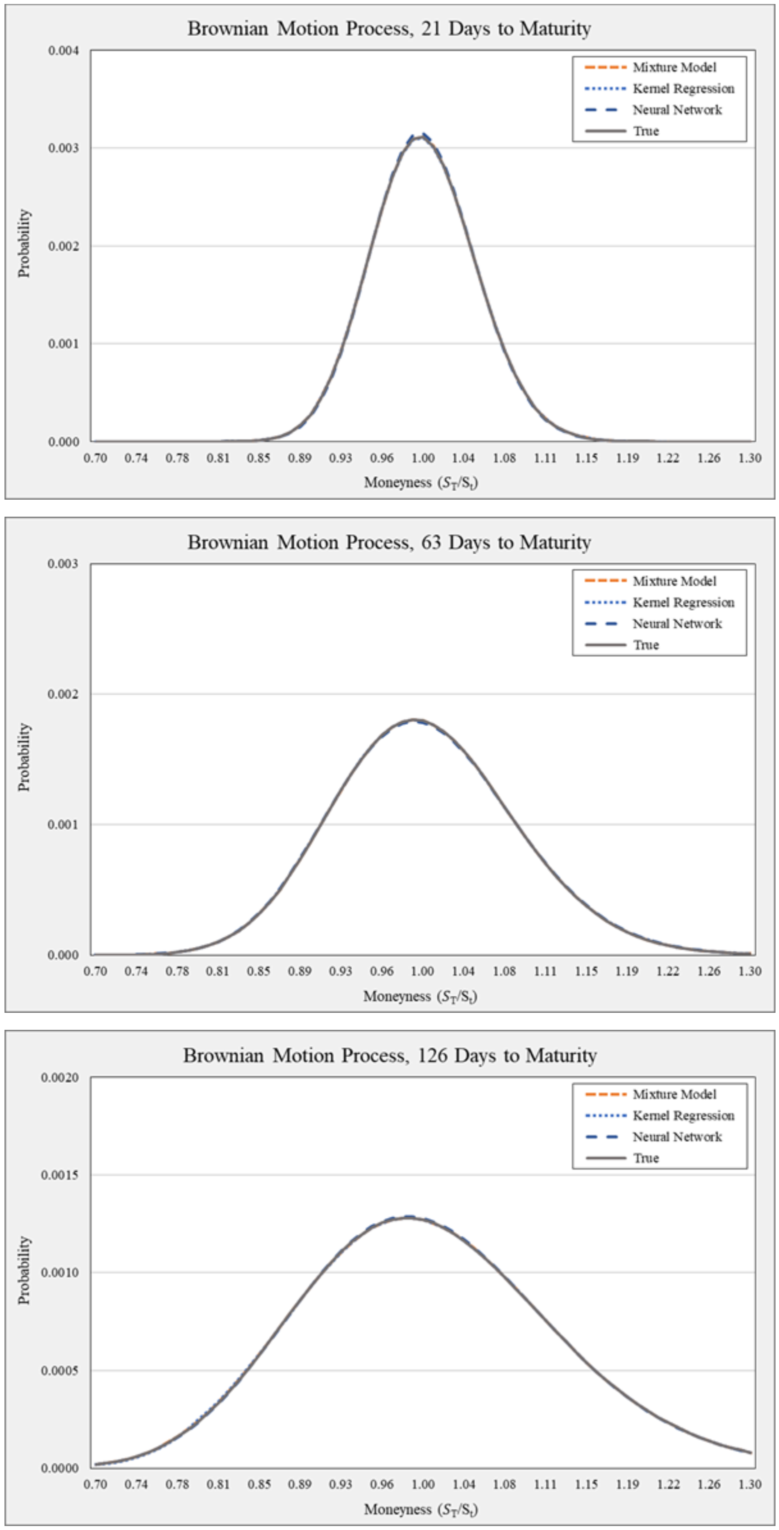
Brownian Motion - 21 Days				
Statistics	True	Mixture Model	Kernel Regression	Neural Network
Mean	1.0011	1.0010	1.0011	1.0011
		(0.004)	(0.000)	(0.000)
Std. Dev.	0.04800	0.04801	0.04795	0.0482
		(0.025)	(0.093)	(0.371)
Skewness	0.1440	0.1428	0.1448	0.1446
		(0.809)	(0.568)	(0.454)
Kurtosis	3.0369	3.0495	3.0372	3.0369
		(0.415)	(0.012)	(0.001)

Brownian Motion - 63 Days				
Statistics	True	Mixture Model	Kernel Regression	Neural Network
Mean	1.0031	1.0031	1.0031	1.0031
		(0.004)	(0.000)	(0.000)
Std. Dev.	0.0833	0.0833	0.0833	0.0836
		(0.007)	(0.065)	(0.364)
Skewness	0.2461	0.2482	0.2489	0.2471
		(0.850)	(1.149)	(0.432)
Kurtosis	3.0817	3.0789	3.0794	3.0807
		(0.093)	(0.076)	(0.032)

Brownian Motion - 163 Days				
Statistics	True	Mixture Model	Kernel Regression	Neural Network
Mean	1.0009	1.0008	1.0009	1.0007
		(0.005)	(0.002)	(0.015)
Std. Dev.	0.1157	0.1157	0.1157	0.1161
		(0.024)	(0.009)	(0.3141)
Skewness	0.3460	0.3467	0.3472	0.3484
		(0.190)	(0.323)	(0.672)
Kurtosis	2.9146	2.9126	2.9111	2.9113
		(0.068)	(0.119)	(0.114)

FIGURE 6. **RISK-NEUTRAL DENSITIES FOR THE BROWNIAN MOTION PROCESSES**

a) ***b)*** ***c)***



Graphical representation of model estimated and true RNDs based on the Brownian motion process for the three maturities: 21, 63 and 126 days.

TABLE 4. SUMMARY STATISTICS STOCHASTIC VOLATILITY

Summary statistics for the mixture model, kernel regression and neural network compared to the true RND for the stochastic volatility data-generation process with 21, 63 and 126 days to maturity. The mean and standard deviation is given in terms of moneyness (S_T/S_t). The absolute percentage error, calculated as the difference between the estimated and true statistic times 100, is presented in parentheses with percentage points as unit of measurement.

Stochastic Volatility - 21 Days

Statistics	True	Mixture Model	Kernel Regression	Neural Network
Mean	1.0010	1.0005	1.0011	1.0011
		(0.045)	(0.010)	(0.010)
Std. Dev.	0.0477	0.0456	0.0473	0.0475
		(4.390)	(0.826)	(0.396)
Skewness	-0.6257	-0.2261	-0.3775	-0.5861
		(63.868)	(39.665)	(6.324)
Kurtosis	3.3117	4.9013	3.1217	3.5021
		(45.000)	(5.735)	(5.750)

Stochastic Volatility - 63 Days

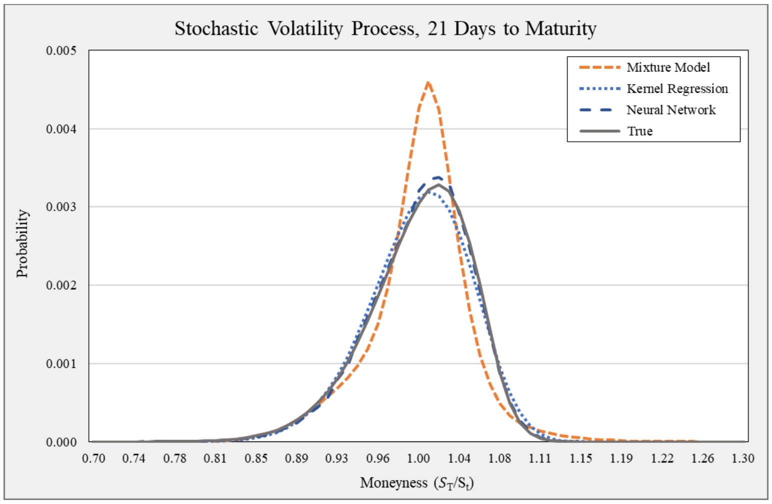
Statistics	True	Mixture Model	Kernel Regression	Neural Network
Mean	1.0028	1.0026	1.0033	1.0030
		(0.018)	(0.045)	(0.021)
Std. Dev.	0.0819	0.0794	0.0823	0.0817
		(2.969)	(0.525)	(0.174)
Skewness	-0.8658	-0.4967	-0.7031	-0.8916
		(42.632)	(18.788)	(2.989)
Kurtosis	3.8996	4.1551	3.7570	3.9702
		(6.553)	(3.657)	(1.811)

Stochastic Volatility - 126 Days

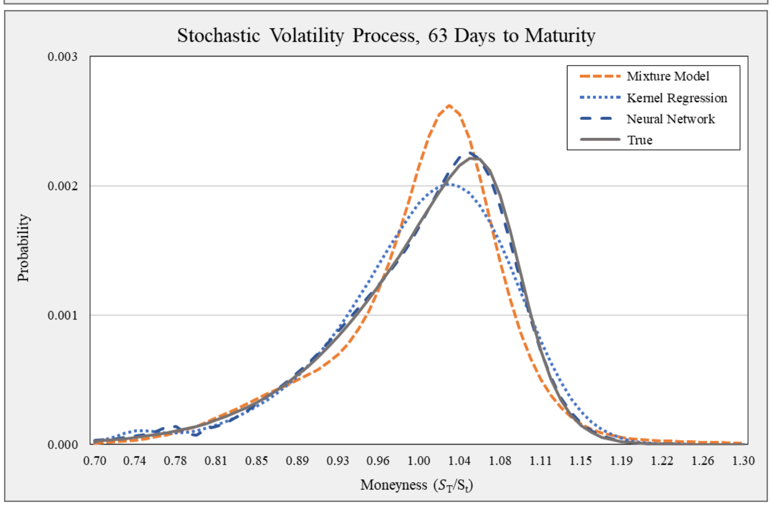
Statistics	True	Mixture Model	Kernel Regression	Neural Network
Mean	1.0014	1.0040	1.0033	1.0023
		(0.262)	(0.197)	(0.091)
Std. Dev.	0.1092	0.1109	0.1087	0.1105
		(1.533)	(0.505)	(1.168)
Skewness	-0.5739	-0.5978	-0.5758	-0.6759
		(4.163)	(0.324)	(17.773)
Kurtosis	3.2865	3.2050	3.2394	3.4885
		(2.480)	(1.434)	(6.146)

FIGURE 7. RISK-NEUTRAL DENSITIES FOR THE STOCHASTIC VOLATILITY PROCESSES

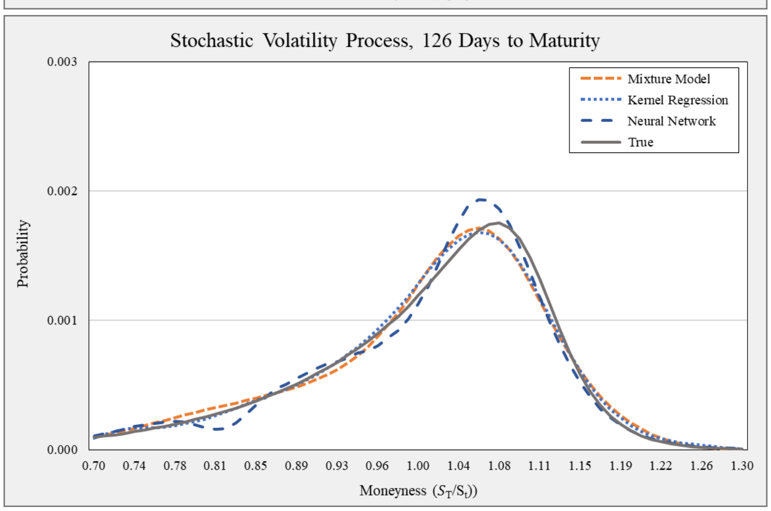
a)



b)



c)



Graphical representation of model estimated and true RNDs based on the stochastic volatility process for the three maturities: 21, 63 and 126 days.

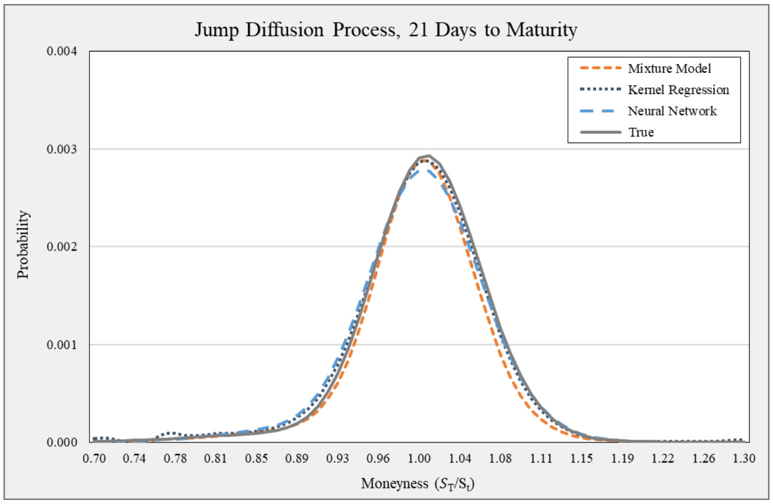
TABLE 5. SUMMARY STATISTICS JUMP DIFFUSION

Summary statistics for the mixture model, kernel regression and neural network compared to the true RND for the jump diffusion data-generation process with 21, 63 and 126 days to maturity. The mean and standard deviation is given in terms of moneyness (S_T/S_t). The absolute percentage error, calculated as the difference between the estimated and true statistic times 100, is presented in parentheses with percentage points as unit of measurement.

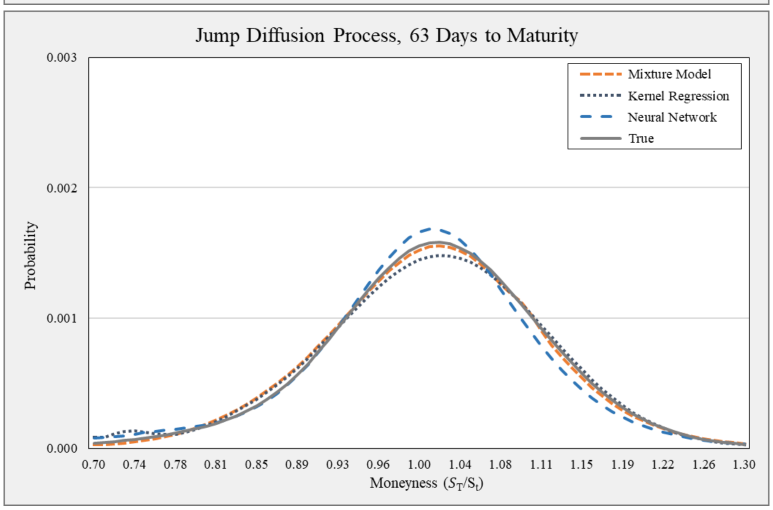
Jump Diffusion - 21 Days				
Statistics	True	Mixture Model	Kernel Regression	Neural Network
Mean	1.0010	1.0015	1.0016	1.0025
		(0.052)	(0.062)	(0.155)
Std. Dev.	0.0605	0.0598	0.0600	0.0613
		(1.128)	(0.761)	(1.392)
Skewness	-0.8174	-0.7628	-0.7816	-0.6742
		(6.680)	(4.375)	(17.521)
Kurtosis	5.8663	5.4790	5.4016	5.2911
		(6.602)	(7.921)	(9.805)
Jump Diffusion - 63 Days				
Statistics	True	Mixture Model	Kernel Regression	Neural Network
Mean	1.0014	1.0020	1.0013	1.0029
		(0.058)	(0.012)	(0.148)
Std. Dev.	0.1037	0.1014	0.1060	0.1054
		(2.161)	(2.198)	(1.691)
Skewness	-0.1702	-0.0110	-0.1897	-0.2907
		(93.564)	(11.487)	(70.791)
Kurtosis	3.3941	3.2839	3.2448	3.7063
		(3.249)	(4.399)	(9.198)
Jump Diffusion - 126 Days				
Statistics	True	Mixture Model	Kernel Regression	Neural Network
Mean	0.9871	0.9808	0.9845	0.9841
		(0.645)	(0.272)	(0.304)
Std. Dev.	0.1399	0.1393	0.1355	0.1387
		(0.437)	(3.133)	(0.874)
Skewness	0.3564	0.2435	0.2549	0.3315
		(31.693)	(28.495)	(6.994)
Kurtosis	2.7304	2.6468	2.5252	2.8605
		(3.062)	(7.514)	(4.764)

FIGURE 8. RISK-NEUTRAL DENSITIES FOR THE JUMP DIFFUSION PROCESSES

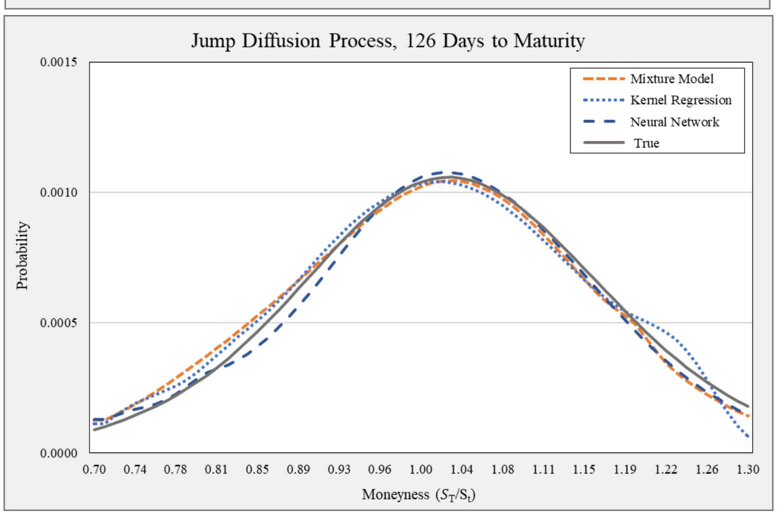
a)



b)



c)



Graphical representation of model estimated and true RNDs based on the jump diffusion process for the three maturities: 21, 63 and 126 days.

TABLE 6. SUMMARY STATISTICS STOCHASTIC VOLATILITY AND JUMP DIFFUSION

Summary statistics for the mixture model, kernel regression and neural network compared to the true RND for the stochastic volatility and jump diffusion data-generation process with 21, 63 and 126 days to maturity. The mean and standard deviation is given in terms of moneyness (S_T/S_t). The absolute percentage error, calculated as the difference between the estimated and true statistic times 100, is presented in parentheses with percentage points as unit of measurement.

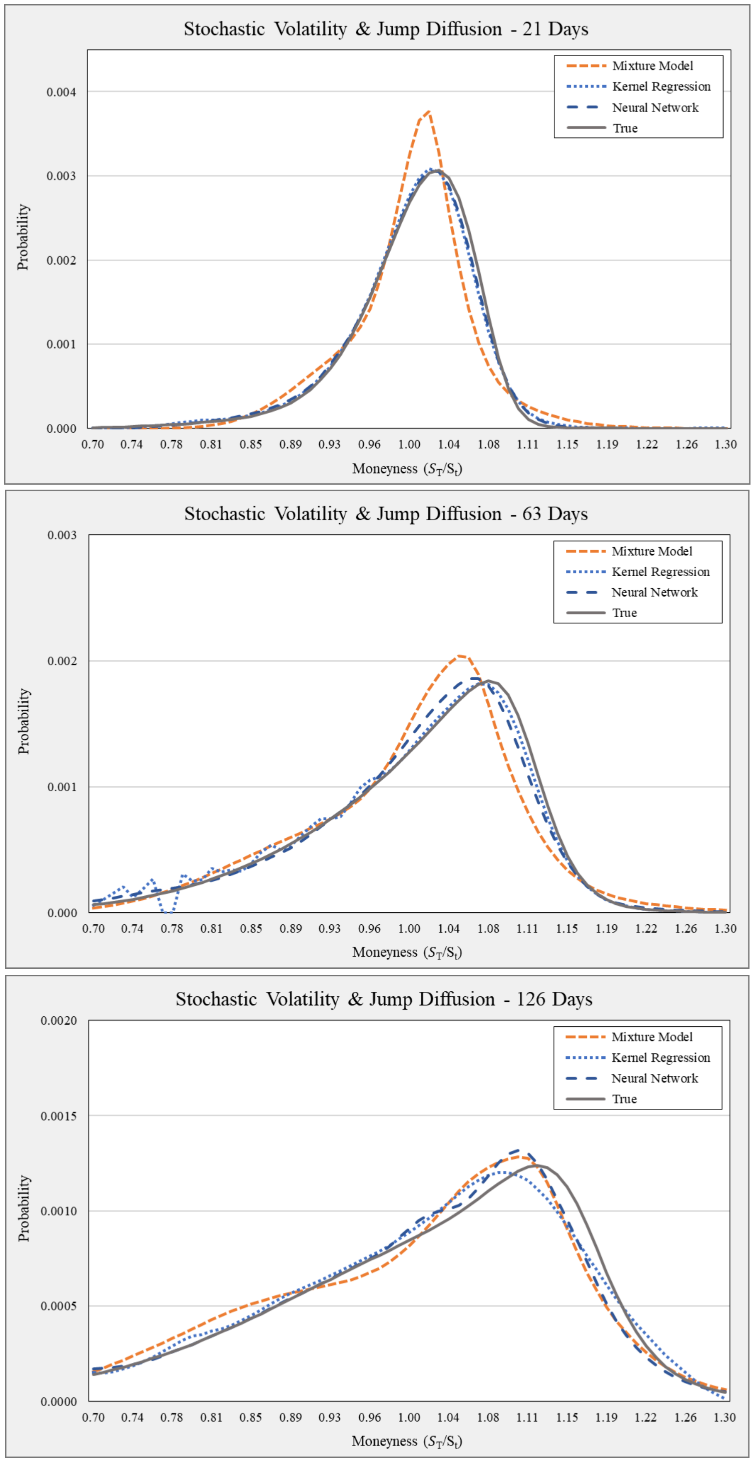
Stochastic Volatility and Jump Diffusion - 21 Days				
Statistics	True	Mixture Model	Kernel Regression	Neural Network
Mean	1.0025	1.0014	1.0018	1.0010
		(0.110)	(0.066)	(0.144)
Std. Dev.	0.0593	0.0571	0.0610	0.0597
		(3.734)	(2.907)	(0.603)
Skewness	-1.1916	-0.1623	-1.1580	-1.1366
		(86.377)	(2.814)	(4.616)
Kurtosis	5.9590	4.2822	5.9806	5.9500
		(28.140)	(0.362)	(0.150)
Stochastic Volatility and Jump Diffusion - 63 Days				
Statistics	True	Mixture Model	Kernel Regression	Neural Network
Mean	1.0059	1.0027	1.0031	1.0011
		(0.319)	(0.280)	(0.479)
Std. Dev.	0.1009	0.0988	0.1044	0.1032
		(2.013)	(3.533)	(2.272)
Skewness	-0.7177	-0.4033	-0.8626	-0.6583
		(44.378)	(20.187)	(8.286)
Kurtosis	3.4309	3.4178	3.6818	3.6064
		(0.381)	(7.314)	(5.114)
Stochastic Volatility - 126 Days				
Statistics	True	Mixture Model	Kernel Regression	Neural Network
Mean	1.0049	1.0022	0.9955	0.9949
		(0.268)	(0.941)	(0.998)
Std. Dev.	0.1362	0.1389	0.1372	0.1387
		(2.020)	(0.792)	(1.876)
Skewness	-0.1627	-0.2323	-0.1028	-0.0852
		(42.803)	(36.785)	(47.639)
Kurtosis	2.5206	2.5069	2.5342	2.5011
		(0.543)	(0.541)	(0.775)

FIGURE 9. **RISK-NEUTRAL DENSITIES FOR THE STOCHASTIC VOLATILITY AND JUMP DIFFUSION PROCESSES**

a)

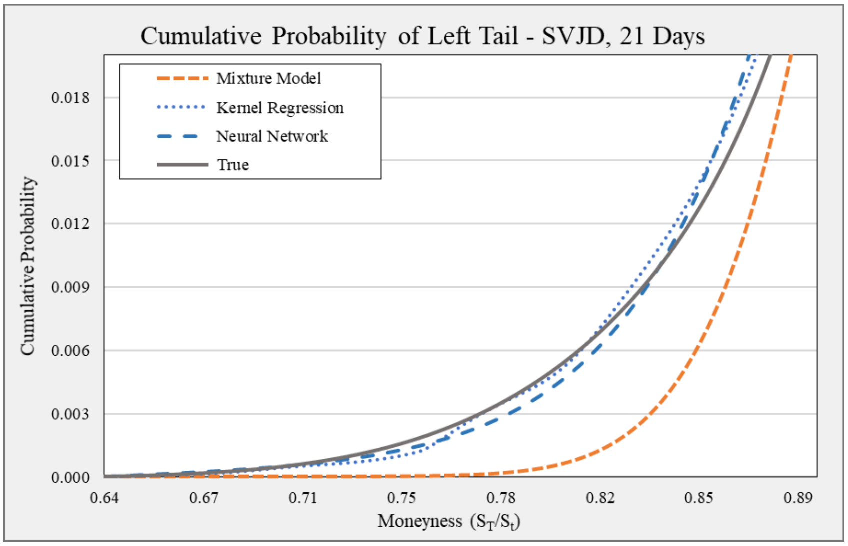
b)

c)



Graphical representation of the model estimated and true RNDs based on the stochastic volatility and jump diffusion process for the three maturities: 21, 63 and 126 days.

***FIGURE 10.* CUMULATIVE PROBABILITY, LEFT TAIL, SVJD - 21 DAYS**



The figure illustrates the cumulative probability observed in the left tail for the densities with the stochastic volatility and jump diffusion process as the underlying data-generation process, with 21 days to maturity.

***TABLE 7.* MODEL PERFORMANCE SUMMARY STATISTICS OVERVIEW**

Overview of the best performing model in terms of replicating the summary statistics of the true RND for each maturity and data-generation processes. The ranking presented in the table is based on the smallest absolute percentage error. Here, MM refers to the mixture model, KR to the kernel regression model and NN to the neural network model.

	Brownian Motion			**Stochastic Volatility**		
Statistics	21	63	126	21	63	126
Mean	NN	KR	KR	NN	MM	NN
Std. Dev.	MM	MM	KR	NN	NN	KR
Skewness	NN	NN	MM	NN	NN	KR
Kurtosis	NN	NN	MM	KR	NN	KR

	Jump Diffusion			**Stochastic Volatility & Jump Diffusion**		
Statistics	21	63	126	21	63	126
Mean	MM	KR	KR	KR	KR	MM
Std. Dev.	KR	NN	MM	NN	MM	KR
Skewness	KR	KR	NN	KR	NN	KR
Kurtosis	MM	MM	MM	NN	MM	KR

5.2 Mann-Whitney U-Test

To further quantify the estimated RNDs' goodness of fit, we conducted the nonparametric, two-sided Mann-Whitney U test introduced by Mann and Whitney (1947). The test compares two sample distributions and determines if the two independent sample distributions originate from a population with the same distribution. Thus, we want to investigate if our estimated RND originates from the same data-generation process as the true RND. The Mann-Whitney U test compares the ranks of each observation in the two distributions and verifies whether the ranks are evenly dispersed across both groups.

We can define our null hypothesis as,

$$H_0\colon Both\ distributions\ originate\ from\ the\ same\ population\ distribution.$$

The Mann-Whitney U test is conducted by ranking each observation in both distributions, or referred to as groups, from highest to lowest, where the lowest value from both data groups gets the value 1 and the highest gets the value $n = n_1 + n_2$, where n_1 is the number of observations in data group 1 and n_2 the number of observations in data group 2. For the purpose of this test, we evaluate 1000 observations across S_T for each density function and categorize the estimated RND observations as group 1 and the true RND observations as group 2. We continue by summarizing the ranks of each group, where the sum of ranks for group 1 is denoted as R_1, group 2 as R_2, and the total sum of the ranks $R = n(n+1)/n$.

According to our null hypothesis, we would expect that the values of R_1 and R_2 are similar to each other. However, R_1 could be equal to R_2 by coincidence and not because the ranks are dispersed evenly across the groups' observations. Thus, we have to compute the Mann-Whitney U test statistic, denoted as U, and compare it to a probability distribution. The test statistic U is defined as,

$$U = \begin{cases} U_1, & for\ U_1 \leq U_2 \\ U_2, & for\ U_1 > U_2 \end{cases}$$

where,

$$U_j = n_1 n_2 + \frac{n_j(n_j + 2)}{2} - R_j, \tag{5.1}$$

for $j = 1$,2. The standardized z score is given by,

$$z = \frac{U - \mu_U}{\sigma_U}, \tag{5.2}$$

where the mean μ_U and standard deviation σ_U of U is defined as,

$$\mu_U = \frac{n_1 n_2}{2},$$

$$\sigma_U = \sqrt{\frac{n_1 n_2}{12}\left[(n_1 + 1) - \sum_{i=1}^{k} \frac{t_i^3 - t_i}{n(n-1)}\right]}, \qquad (5.3)$$

where t_i is the number of observations that have the same score and subsequently share the same rank. Here, we conduct the Mann-Whitney U test as a two-sided test, thus our critical values are defined as -1.96 and 1.96 , where we reject the null hypothesis if $z < -1.96$ or $z > 1.96$, or equivalent to a p-value < 0.05. The results from our Mann-Whitney U test can be observed in *table 8* and *table 9.*

Interpreting the results of the Brownian motion data-generation process, presented in *table 8,* one can observe that we fail to reject the null hypothesis for all densities, over all three maturities. These findings are consistent with the results reported in *section 5.1*, where we concluded that all of the models can sufficiently estimate a nearly identical RND to the true RND if the asset returns are log-normal.

Observing the results from the stochastic volatility processes in *table 8*, we find we reject the null hypothesis for the mixture model estimated RNDs for 21 and 63 days to maturity. These results are coherent with the findings from the previous section, where we concluded that the mixture model neglected the excess skewness, observed in *table 4* and illustrated in *figures 7a* and *7b*. We further find that we reject the null hypothesis for the RND with 21 days to maturity estimated by the kernel regression model. However, observing the results in *table 4* and *figure 7a,* it is hard to verify why we reject the null hypothesis. Thus, we have to investigate the tails of the two densities. Interpreting the data, we found that the right tail of the density estimated by the kernel regression model converges almost instantaneously to zero at $S_T = 3100$, which is illustrated in *figure 11*. Due to the rapid convergence towards zero, the ranks for the true density are considerably higher compared to the density estimated by the kernel regression for the last 500 observations in the densities' right tails, illustrated in *figure 12*. Meanwhile, considering the estimated RNDs for the neural network, we can conclude that we fail to reject the null hypothesis for all maturities.

Continuing by interpreting the results from the jump diffusion processes, observed in *table 9*, we recognize that we fail to reject the null hypothesis for all models over all three maturities. However, we recognize that the p-values are lower for the mixture model and kernel regression compared to the neural network, but never significant.

Lastly, we can observe the results reported in *table 9* for the stochastic volatility and jump diffusion process, where we reject the null hypothesis for the mixture model at the 21 days to maturity. This is further supported by the findings from *table 6*, where we see that the mixture model underestimates the skewness of the SVJD process for the 21 days to maturity. Further, we can conclude that we fail to reject the null hypothesis for the remaining results across models and maturities.

As a conclusion, we can verify that the mixture model's estimated RND proves to be statistically different from the true RND for the SV and SVJD data-generation processes, with 21 days to maturity, as well as for the SV processes with 63 days to maturity. We further validated that the kernel regression model's estimated RND is statistically different from the true RND for the SV data-generation processes with 21 days to maturity, while we conclude that the RNDs estimated by the neural network model is statistically identical to the true RNDs for all data-generation processes and maturities.

TABLE 8. MANN-WHITNEY U TEST, BM AND SV

Reported results from the two-sided Mann-Whitney U test for the three different models: the mixture model, kernel regression model and neural network model, for the Brownian motion and stochastic volatility data-generation processes over the three maturities: 21 days, 63 days and 126 days. The results are reported in terms of U statistics and the p-values are reported in parentheses. Significance level: p* < .05.

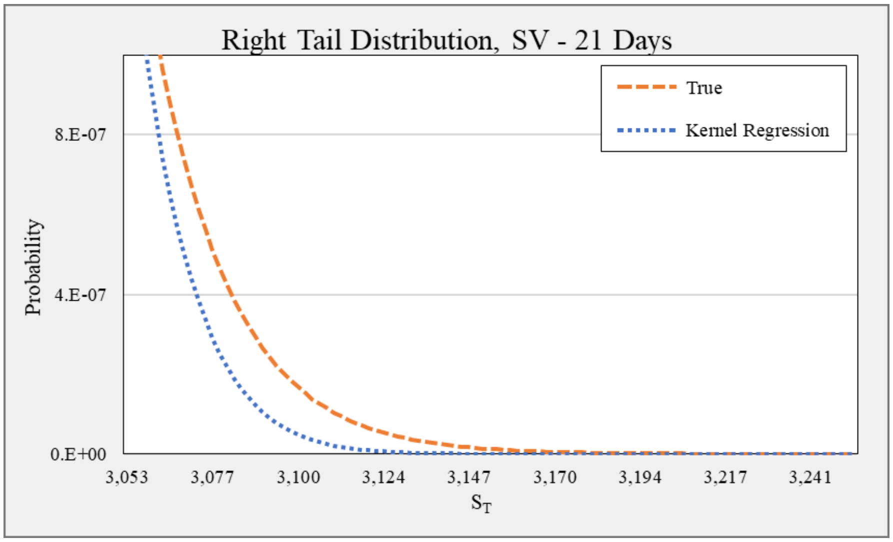
	Brownian Motion			**Stochastic Volatility**		
Time to Maturity	Mixture Model	Kernel Regression	Neural Network	Mixture Model	Kernel Regression	Neural Network
21 Days	494421.0	498291.0	496116.5	589210.0	463648.5	488386.5
	(0.666)	(0.895)	(0.764)	(0.000)*	(0.005)*	(0.369)
63 Days	496547.5	499302.5	497712	525509.5	501992	501251.5
	(0.789)	(0.957)	(0.859)	(0.048)*	(0.877)	(0.923)
126 Days	499414.5	498816.5	498557.5	500862	499224	502995
	(0.964)	(0.927)	(0.911)	499224	(0.952)	(0.817)

TABLE 9. MANN-WHITNEY U TEST, JD AND SVJD

Reported results from the two-sided Mann-Whitney U test for the three different models: the mixture model, kernel regression model and neural network model, for the jump diffusion and stochastic volatility with jump diffusion data-generation processes over the three maturities: 21 days, 63 days and 126 days. The results are reported in terms of U statistics and the p-values are reported in parentheses. Significance level: p* < .05.

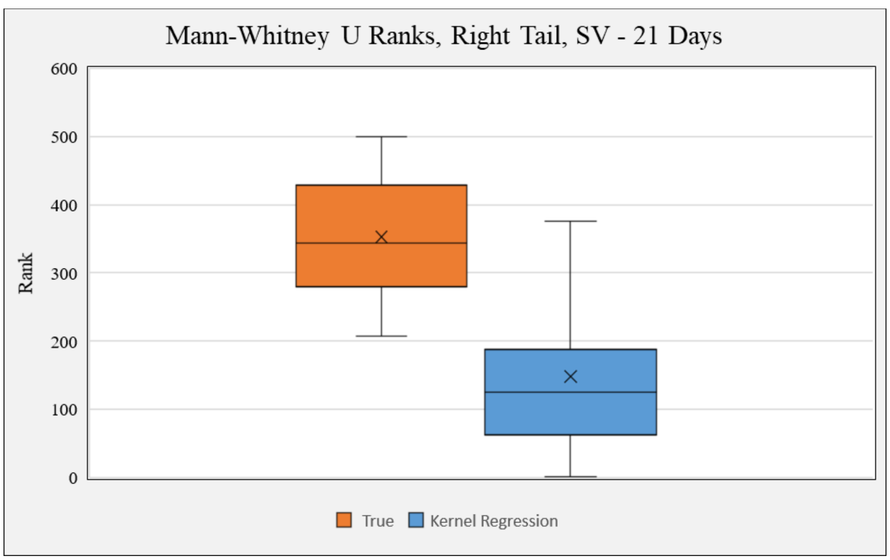
	Jump Diffusion			**Stochastic Volatility & Jump Diffusion**		
Time to Maturity	Mixture Model	Kernel Regression	Neural Network	Mixture Model	Kernel Regression	Neural Network
21 Days	483083.0	486302.0	493373.5	474682.5	519288.0	502707.0
	(0.190)	(0.289)	(0.608)	(0.0499)*	(0.974)	(0.834)
63 Days	493787.0	499966.0	503627.5	516148.0	489554.0	516551.0
	(0.616)	(0.998)	(0.779)	(0.211)	(0.419)	(0.200)
126 Days	495905.5	488854.5	501142.0	508250.0	502127.5	496747.0
	(0.751)	(0.388)	(0.930)	(0.523)	(0.869)	(0.801)

***FIGURE 11.* RIGHT TAIL PROBABILITY DISTRIBUTION, SV - 21 DAYS**



The probability distribution observed in the right tail for the true and kernel regression estimated RND, where the underlying follows a stochastic volatility process with 21 days to maturity.

***FIGURE 12.* MANN-WHITNEY U RANKS, RIGHT TAIL, SV – 21 DAYS**



Box diagram for the last 500 Mann-Whitney U ranks distributed over the right tail of the true and kernel regression estimated RND, following a stochastic volatility process with 21 days to maturity. The uneven distribution of ranks suggests that the two densities do not come from the same population data-generation process.

5.3 Empirical Pricing of Options

To evaluate the theoretical valuation capabilities of the estimated RNDs, we conducted a simple risk-neutral valuation analysis for a set of different derivatives. To price the derivatives, we utilize the fundamental risk-neutral valuation equation observed in *section 2.3, equation (2.7)*. However, for the purpose of this thesis and for the reader to be able to easily interpret the results, we will simplify the payoff structure similarly to that of a binary option, where the call option pays 100 if the terminal stock price is equal to or above the strike price and the put options pays 100 if the terminal stock price is equal to or below the price strike. We define the payoff for the call (C) and put (P) options as follows,

$$C = \begin{cases} 0, & for\ S_T < K \\ 100, & for\ S_T \geq K \end{cases}, \qquad P = \begin{cases} 0, & for\ S_T > K \\ 100, & for\ S_T \leq K \end{cases}. \tag{5.4}$$

Thus, utilizing *equation (2.10)* we can further define the price of the theoretical, binary call and put options, incorporating our payoff structure in *equation (5.4)*, simply as,

$$c(S_t, K) = e^{-rT} \int_{S_T=K}^{\infty} Cq(S_T)dS_T,$$

$$p(S_t, K) = e^{-rT} \int_{0}^{S_T=K} Pq(S_T)dS_T, \tag{5.5}$$

where $c(S_t, K)$ is the current price of the call option, $p(S_t, K)$ is the current price of the put option and $q(S_T)$ is the risk-neutral density evaluated at S_T. We want to investigate the robustness of the models' valuation efficiency; thus, we will proceed to evaluate two call and two put options. First, we define one call and one put option with strike price equal to the spot price, K=2673.61. Further, we define a call option with strike price equal to 3000, with the purpose of analyzing the densities' right tails. Lastly, we define a put option with a strike price equal to 2300, with the purpose of analyzing the densities' left tails. To further examine the robustness, we evaluate each option over three different maturities: 21, 63 and 126 trading days. The valuation results are presented in *tables 10 – 17*.

The valuation results for the options where the underlying follows a Brownian motion process is presented in *table 10* and *table 11*. As concluded in the earlier sections, we see that the models replicate the true option price reasonably well, especially for the longer maturities. The two more exceptional errors can be observed for the shorter maturity of 21 days. We see that the neural network undervalues the 21 days to maturity call option with a strike price equal to 3000, thus we can conclude that the neural network estimates lower probability for the right tail outcomes, while the mixture model undervalues the 21 days to maturity put option with a strike price equal to 2300, hence the model estimates a lower probability for the left tail outcomes. In general, we see that the true densities have fatter right tails than the estimated RNDs, where practically all models undervalue both call options for all maturities, however, not considerably.

We proceed by investigating the results given in *table 12* and *table 13*, where the underlying follows a stochastic volatility process. The most striking result is the mixture model's estimated price for the 21 days to maturity call option with a strike price of 3000, where the price percentage error is quite salient. This result suggests that the estimated RND has a noticeable fatter right tail than the true RND. Observing the results in *table 4* we see that all models and particularly the mixture models underestimate the negative skew of the true RND for the 21 days to maturity. Hence, we can confirm that all three models, and especially the mixture model, overestimate the probability of the spot price surging over 3000 at maturity T=21 days. Moreover, we observe that the kernel regression model prominently overvalues the K=3000, 63 days to maturity call option, that is justified by the kernel regression's heavier right tail observed *figure 7b* in *section 5.1*. Interpreting the results for the put options, we can conclude that the kernel regression does reasonably well for the K=2300 put options for all maturities, while the mixture and neural network models undervalues the 21 days to maturity put option. However, observing the results for the K=2673.61 call and put options, for all maturities, we recognize that the mixture model manages to estimate the prices quite sufficiently for all maturities compared to the other models.

Continuing, by observing the outcomes from the valuation of options following the jump diffusion process, in *table 14* and *table 15,* we examine similar, inconclusive results to the ones observed in the summary statistics, in *section 5.1.* We can see that all models perform quite insufficiently, with the neural network as the only exception for the 21 days to maturity call and put option with K=3000 and K=2300, respectively. For the jump diffusion process in general, we might argue that the neural network is the best performing, or at least the best performing model in terms of consistency for both the put and call options for all maturities. More strikingly, we see that for the longer maturity of 126 days, the models undervalue the call options and overvalue the put options, concluding that the models allocate higher probability to the asset returns occurring in the left tail of the density and less probability to the returns in the right tail. These findings are further confirmed when observing *figure 8c.*

Finally, we interpret the results from *table 16* and *table 17,* where the underlying security follows a stochastic volatility and jump diffusion process. Similar to the results from the stochastic volatility process, we recognize that the mixture model overvalues the shorter 21 days to maturity call option with strike K=3000. This result can clearly be verified by observing *figure 9a* and *table 6*, where we see the salient right tail estimated by the mixture model and further underestimated negative skewness. Observing the same option, we notice that both the kernel regression and neural network models overvalue the call option K=3000, indicating heavier right tails. However, more interestingly, we see that both models, the kernel regression and neural network, further overvalue the 21 days to maturity K=2300 put option, implying that the model estimated RNDs have heavier left tail and right tails compared to the true RND. However, observing the summary statistics in *table 6*, we see that the true RND and two models further have almost similar kurtosis and skewness. To fully investigate the given results, one has to closely observe the RNDs' tails, presented in *figure 13*. Observing the figure, we can infer that the true density distributes higher probabilities to events occurring at strikes lower than K=2100, compared to the kernel and neural network models. Thus, to properly capture the features of the tail and more extreme negative values, one would have to price a deeper out-of-the-money put option. Overall, we see that in the majority of cases the models overvalue the put options and undervalue the call options, indicating that the models have estimated fatter left tails and thinner right tails, similar to the results observed for the jump diffusion process. As a conclusion, one can argue that the neural network model once more outperforms, at least in terms of consistency, the two other models for all observed cases similar to the results presented for the jump diffusion process.

Table 18 presents an overview of the model with the lowest absolute percentage error for each data-generation process and maturity. Interpreting the table, we can conclude that the mixture model has on most occasions the lowest absolute percentage pricing error for the call options, and especially for the at-the-money options. However, recognizing the results for both the call and put options, we see that the neural network has the lowest absolute percentage pricing error on most occasions overall, followed by the mixture model and lastly the kernel regression model.

As a conclusion, we can see that the valuation of the binary options, where the underlying follows a Brownian motion, are coherent with our earlier results from previous tests, where all three models were able to predict the prices with reasonable accuracy for all maturities. In general, we recognize that the mixture model is able to predict the option prices efficiently well for the at-the-money options, while falling short on options based on a RND with a distinct negative skew, observed for the stochastic volatility processes. The kernel regression tends to overvalue the price of the call options with K=3000, especially for the options following a stochastic volatility process and with shorter maturities. The neural network on the other hand performs reasonably well compared to the other models, especially for the put options with strike K=2300, for options with longer maturities and for the options where the underlying follows a jump diffusion process. Lastly, we recognized that the neural network had the lowest absolute percentage pricing error on most occasion considering both the call and put options for all maturities.

TABLE 10. CALL OPTION VALUATION - BROWNIAN MOTION

Pricing of a hypothetical call option that has a payoff of a 100 if the spot price (S_T) is greater than or equal to the strike price (K), otherwise the payoff is 0, for the three different maturities: 21, 63 and 126 days. The price of the option is compared between the true RND, the mixture model, kernel regression and neural network where the underlying data-generation follows a Brownian motion process. The values are presented as option prices with corresponding percentage errors in parentheses with percentage points as unit of measurement. The percentage error is calculated as the percentage difference between the true and the model's estimated price, times 100.

	K = 2673.61				K = 3000			
Maturity	True	Model	Regression	Network	True	Model	Regression	Network
21 Days	49.7619	49.4108	49.4033	48.7093	0.8127	0.7922	0.7888	0.7242
		(-0.706)	(-0.721)	(-2.115)		(-2.516)	(-2.935)	(-10.891)
63 Days	49.6599	49.4420	49.4173	49.3559	8.2088	8.1216	8.1187	7.9478
		(-0.439)	(-0.489)	(-0.612)		(-1.062)	(-1.097)	(-3.179)
126 Days	49.0943	48.9329	48.9381	48.8760	15.7611	15.6595	15.6550	15.4744
		(-0.329)	(-0.318)	(-0.445)		(-0.645)	(-0.673)	(-1.819)

TABLE 11. PUT OPTION VALUATION - BROWNIAN MOTION

Pricing of a hypothetical put option that has a payoff of a 100 if the spot price (S_T) is less than or equal to the strike price (K), otherwise the payoff is 0, for the three different maturities: 21, 63 and 126 days. The price of the option is compared between the true RND, the mixture model, kernel regression and neural network where the underlying data-generation follows a Brownian motion process. The values are presented as option prices with corresponding percentage errors in parentheses with percentage points as unit of measurement. The percentage error is calculated as the percentage difference between the true and the model's estimated price, times 100.

	K = 2673.61				K = 2300			
Maturity	True	Model	Regression	Network	True	Model	Regression	Network
21 Days	50.1299	50.3692	50.3888	50.8833	0.0835	0.0744	0.0836	0.0831
		(0.477)	(0.516)	(1.503)		(-10.972)	(0.116)	(-0.509)
63 Days	50.0072	50.1213	50.1507	50.0136	3.4797	3.5076	3.4761	3.3773
		(0.228)	(0.287)	(0.013)		(0.802)	(-0.103)	(-2.942)
126 Days	49.8569	49.9214	49.9228	49.8059	9.9505	10.0035	10.0308	9.7630
		(0.129)	(0.132)	(-0.102)		(0.533)	(0.807)	(-1.884)

TABLE 12. CALL OPTION VALUATION - STOCHASTIC VOLATILITY

Pricing of a hypothetical call option that has a payoff of a 100 if the spot price (S_T) is greater than or equal to the strike price (K), otherwise the payoff is 0, for the three different maturities: 21, 63 and 126 days. The price of the option is compared between the true RND, the mixture model, kernel regression and neural network where the underlying data-generation follows a stochastic volatility process. The values are presented as option prices with corresponding percentage errors in parentheses with percentage points as unit of measurement. The percentage error is calculated as the percentage difference between the true and the model's estimated price, times 100.

	K = 2673.61				K = 3000			
Time to Maturity	True	Mixture Model	Kernel Regression	Neural Network	True	Mixture Model	Kernel Regression	Neural Network
21 Days	55.2289	55.5281	54.2285	54.3328	0.0361	0.9039	0.0468	0.0451
		(0.542)	(-1.811)	(-1.623)		(2,400.992)	(29.414)	(24.898)
63 Days	58.6892	58.8209	56.0840	58.2869	2.9792	3.6578	4.5768	3.2110
		(0.224)	(-4.439)	(-0.686)		(22.776)	(53.623)	(7.777)
126 Days	60.7648	60.6397	60.3106	61.9207	11.5098	12.3035	12.2141	10.6114
		(-0.206)	(-0.748)	(1.902)		(6.896)	(6.119)	(-7.806)

TABLE 13. PUT OPTION VALUATION - STOCHASTIC VOLATILITY

Pricing of a hypothetical put option that has a payoff of a 100 if the spot price (S_T) is less than or equal to the strike price (K), otherwise the payoff is 0, for the three different maturities: 21, 63 and 126 days. The price of the option is compared between the true RND, the mixture model, kernel regression and neural network where the underlying data-generation follows a stochastic volatility process. The values are presented as option prices with corresponding percentage errors in parentheses with percentage points as unit of measurement. The percentage error is calculated as the percentage difference between the true and the model's estimated price, times 100.

	K = 2673.61				K = 2300			
Time to Maturity	True	Mixture Model	Kernel Regression	Neural Network	True	Mixture Model	Kernel Regression	Neural Network
21 Days	44.6630	44.1969	45.5637	45.4593	0.7311	0.5440	0.7332	0.6447
		(-1.044)	(2.017)	(1.783)		(-25.592)	(0.279)	(-11.817)
63 Days	40.9162	40.6408	43.4919	41.0502	6.1996	6.0925	5.6937	5.7749
		(-0.673)	(6.295)	(0.328)		(-1.728)	(-8.16)	(-6.85)
126 Days	37.7023	38.2794	38.2307	36.2706	10.4894	11.7359	10.3176	9.8981
		(1.531)	(1.402)	(-3.797)		(11.884)	(-1.637)	(-5.637)

TABLE 14. CALL OPTION VALUATION - JUMP DIFFUSION

Pricing of a hypothetical call option that has a payoff of a 100 if the spot price (S_T) is greater than or equal to the strike price (K), otherwise the payoff is 0, for the three different maturities: 21, 63 and 126 days. The price of the option is compared between the true RND, the mixture model, kernel regression and neural network where the underlying data-generation follows a jump diffusion process. The values are presented as option prices with corresponding percentage errors in parentheses with percentage points as unit of measurement. The percentage error is calculated as the percentage difference between the true and the model's estimated price, times 100.

	K = 2673.61				K = 3000			
Time to Maturity	True	Mixture Model	Kernel Regression	Neural Network	True	Mixture Model	Kernel Regression	Neural Network
21 Days	55.4099	52.7766	52.9747	52.1095	1.5787	1.3262	1.3591	1.6016
		(-4.752)	(-4.395)	(-5.956)		(-15.992)	(-13.905)	(1.450)
63 Days	56.7847	56.9267	57.7848	54.3400	13.6552	13.8241	14.6971	13.3927
		(-0.19)	(0.060)	(1.568)		(1.237)	(7.630)	(-1.923)
126 Days	56.3352	53.9882	53.6013	54.6009	24.2279	22.3083	22.4365	22.3835
		(-4.166)	(-4.853)	(-3.079)		(-7.923)	(-7.394)	(-7.613)

TABLE 15. PUT OPTION VALUATION - JUMP DIFFUSION

Pricing of a hypothetical put option that has a payoff of a 100 if the spot price (S_T) is less than or equal to the strike price (K), otherwise the payoff is 0, for the three different maturities: 21, 63 and 126 days. The price of the option is compared between the true RND, the mixture model, kernel regression and neural network where the underlying data-generation follows a jump diffusion process. The values are presented as option prices with corresponding percentage errors in parentheses with percentage points as unit of measurement. The percentage error is calculated as the percentage difference between the true and the model's estimated price, times 100.

	K = 2673.61				K = 2300			
Time to Maturity	True	Mixture Model	Kernel Regression	Neural Network	True	Mixture Model	Kernel Regression	Neural Network
21 Days	44.5771	47.0653	47.2150	47.4684	2.0227	2.5724	2.5712	1.9976
		(5.582)	(5.918)	(6.486)		(27.173)	(27.117)	(-1.241)
63 Days	42.9231	45.8532	45.5862	44.5247	6.6509	6.5254	7.4682	6.8175
		(6.826)	(6.204)	(3.731)		(-1.887)	(12.288)	(2.505)
126 Days	41.7355	43.8481	43.9925	42.5929	11.7831	13.9491	13.4177	12.6625
		(5.062)	(5.408)	(2.054)		(18.382)	(13.872)	(7.462)

TABLE 16. CALL OPTION VALUATION - STOCHASTIC VOLATILITY AND JUMP DIFFUSION

Pricing of a hypothetical call option that has a payoff of a 100 if the spot price (S_T) is greater than or equal to the strike price (K), otherwise the payoff is 0, for the three different maturities: 21, 63 and 126 days. The price of the option is compared between the true RND, the mixture model, kernel regression and neural network where underlying data-generation follows a stochastic volatility and jump diffusion process. The values are presented as option prices with corresponding percentage errors in parentheses with percentage points as unit of measurement. The percentage error is calculated as the percentage difference between the true and the model's estimated price, times 100.

	K = 2673.61				K = 3000			
Time to Maturity	True	Mixture Model	Kernel Regression	Neural Network	True	Mixture Model	Kernel Regression	Neural Network
21 Days	58.7233	55.8921	57.2959	58.6401	0.3040	2.0514	0.5752	0.4815
		(-4.821)	(-2.431)	(-0.142)		(574.841)	(89.224)	(58.407)
63 Days	60.3896	58.2663	58.6229	59.2587	8.7643	8.0809	7.8616	7.7856
		(-3.516)	(-2.925)	(-1.873)		(-7.797)	(-10.299)	(-11.166)
126 Days	59.7377	58.6443	57.4401	57.9493	25.8016	22.1535	22.2551	21.9937
		(-1.83)	(-3.846)	(-2.994)		(-14.139)	(-13.745)	(-14.758)

TABLE 17. PUT OPTION VALUATION - STOCHASTIC VOLATILITY AND JUMP DIFFUSION

Pricing of a hypothetical put option that has a payoff of a 100 if the spot price (S_T) is less than or equal to the strike price (K), otherwise the payoff is 0, for the three different maturities: 21, 63 and 126 days. The price of the option is compared between the true RND, the mixture model, kernel regression and neural network where underlying data-generation follows a stochastic volatility and jump diffusion process. The values are presented as option prices with corresponding percentage errors in parentheses with percentage points as unit of measurement. The percentage error is calculated as the percentage difference between the true and the model's estimated price, times 100.

	K = 2673.61				K = 2300			
Time to Maturity	True	Mixture Model	Kernel Regression	Neural Network	True	Mixture Model	Kernel Regression	Neural Network
21 Days	41.1494	43.8570	42.5657	41.0527	2.7139	1.4974	3.0292	2.9715
		(6.580)	(3.442)	(-0.235)		(-44.826)	(11.617)	(9.492)
63 Days	38.9834	41.1838	40.6895	39.5582	9.2795	9.7655	10.0249	9.6223
		(5.644)	(4.376)	(1.474)		(5.238)	(8.033)	(3.694)
126 Days	37.9488	39.8142	40.0663	39.0479	13.6750	16.5484	14.7918	14.6343
		(4.916)	(5.580)	(2.896)		(21.012)	(8.166)	(7.015)

FIGURE 13. **PROBABILITY DISTRIBUTION, LEFT TAIL, SVJD – 21 DAYS**

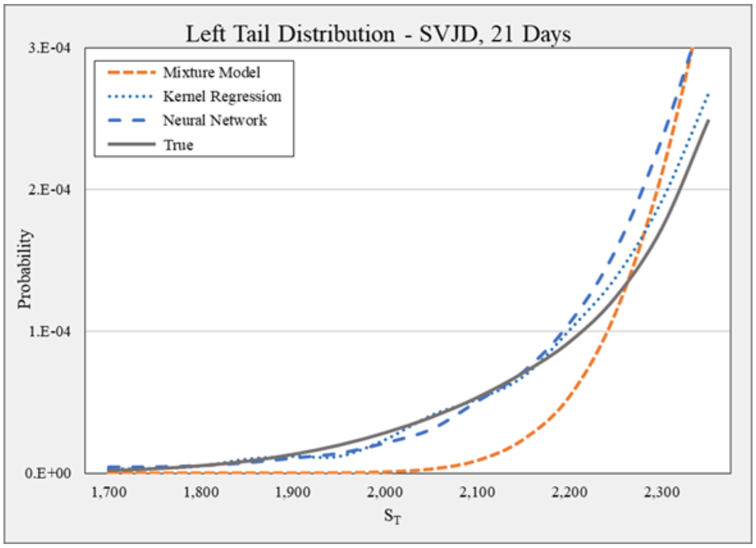


Illustration of the left tail distribution for the true and model estimated RNDs, where the underlying follows a stochastic volatility and jump diffusion processes, with 21 days to maturity. The purpose of the figure is to illustrate the heaver tail for the true RND with strikes lower than 2100 compared to the other model estimated RNDs.

TABLE 18. **OVERVIEW OF OPTION VALUATION PERFORMANCE**

The table represents the model estimated RND with smallest absolute percentage pricing error compared to prices based on the true RNDs, for each data-generation processes and maturity. The first section of the table represents the results from the binary call options, while the second part represents the results from the binary put options.

Call Option	**K = 2673.61**				**K = 3000**			
	BM	SV	JD	SVJD	BM	SV	JD	SVJD
21 Days	MM	MM	KR	NN	MM	NN	NN	NN
63 Days	MM	MM	KR	NN	MM	NN	MM	MM
126 Days	KR	MM	NN	MM	MM	KR	KR	KR
Put Option	**K = 2673.61**				**K = 2300**			
	BM	SV	JD	SVJD	BM	SV	JD	SVJD
21 Days	MM	MM	MM	NN	KR	KR	NN	NN
63 Days	NN	NN	NN	NN	KR	MM	MM	NN
126 Days	NN	KR	NN	NN	MM	KR	NN	NN

5.3.1 Summary Statistics and Absolute Pricing Error

As a last test for the pricing of the theoretical binary options, we want to investigate the relationship between the difference in option valuation and summary statistics, where we intend to study the impact of skewness, kurtosis and time to maturity on the pricing errors for the K=2300 put option. We conducted a simple multivariate regression, constituted by the absolute pricing error,

$$absolute\ precentage\ pricing\ error = \ |precentage\ pricing\ error|,$$

as the dependent variable and skewness, kurtosis observed from the true density, as well as time to maturity as the independent variables. We further divide the results into two different regressions for each model, where we also control for the specific data-generation process. The data for the independent variable skewness is converted to 'absolute' skewness for accurate regression results. For the test we define two corresponding null hypotheses for the two statistical moments as,

H_{01}: *A change in skewness does not have a statistical effect on the absolute percentage pricing error*

H_{02}: *A change in kurtosis does not have a statistical effect on the absolute percentage pricing error*

The regression results are presented in *table 19.* However, it is important to emphasize the fact that these results are only representable for the pricing error of a binary put option, with payoff defined in *equation (5.4)* and with a strike price of 2300. Further and most notably, the test is only conducted on twelve observations, thus, these results are only intended as indications of the models' pricing error given the summary statistics of the true densities.

Interpreting *table 19*, we observe that the mixture model does not show any relationship between the error produced from pricing the theoretical options and any of the summary statistics. Thus, this suggests that the mixture model is statistically indifferent of the density features such as skewness and kurtosis when pricing this particular put option and thus we fail to reject both null hypotheses. Observing the results for the kernel regression, we can verify that the kurtosis feature is statically significant at the 5% level. We can interpret the result as when the RND is associated with an increased kurtosis, we additionally examine an increase in the mispricing of the option. Thus, the kernel regression model performs worse for pricing put options with a strike price of 2300, where the underlying asset's RND is characterized by a higher kurtosis. Hence, we can reject our second null hypothesis H_{02}. Further, we notice that the kernel regression model is also statistically significant for skewness at the 10% level, and thus we reject our first null hypothesis H_{01} when controlling for the data-generation process. However, this relationship is the opposite of what one would initially assume, where the results suggest that the mispricing decreases with higher skewness.

Investigating the relationship by observing the results for *table 14* and *table 15*, we notice the kernel regression model's higher pricing errors for the jump diffusion process, accompanied by lower skewness for the two longer maturities. Thus, the results associated with the jump diffusion process, as well as the low number of observations, contribute to the negative relationship between skewness and the pricing error for the kernel model. Lastly, observing the results for the neural network model, we recognize that the skewness is statistically significant at the 5% level when we do not control for the underlying process and 10% level when we do. Thus, the results suggest that the neural network model price options more accurately when the terminal asset return distribution exhibits less skewness, a relationship that would confirm ones initial expectations and thus we can reject our first null hypothesis H_{01}. However, interpreting the results for the kurtosis, we can conclude that the variable is not statistically significant, therefore we fail to reject our second null hypothesis H_{02}.

In conclusion, we presented a simple multivariate OLS-regression to examine the relationship between mispricing of the $K = 2300$ put option and the density characteristics associated with skewness and kurtosis. The regressions showed that kurtosis and skewness have an impact on the valuation for the kernel regression and neural network model respectively, but due to only a few observations for the regressions the results should be interpreted with some skepticism.

TABLE 19. MULTIVARIATE REGRESSION, PRICING ERROR AND SUMMARY STATISTICS

The following table represents the results from the multivariate regressions, with the dependent variable *absolute percentage pricing error* in terms of percentage points from the put option with K=2300. The mean values of regression coefficients are presented as units, while the independent variable *time* is presented in terms of years and standard deviations are reported in parentheses. The significance levels are defined as ***, **, * for a statistical significance of 1%, 5% and 10% respectively.

Variables	Mixture Model		Kernel Regression		Neural network	
Skewness	5.3102	2.1132	-14.6847	-16.2905*	11.6807**	11.0091*
	(19.253)	(18.871)	(8.819)	(8.455)	(4.837)	(4.891)
Kurtosis	6.4624	4.7875	9.2461**	8.4048**	-2.7893	-3.1412
	(6.604)	(6.555)	(3.025)	(2.937)	(1.659)	(1.699)
Time to Maturity	1.3377	-7.7034	15.9229	11.3816	-1.8277	-3.7269
	(27.162)	(27.386)	(12.442)	(12.271)	(6.824)	(7.097)
No. Observations	12	12	12	12	12	12
Data-Generation Processes	No	Yes	No	Yes	No	Yes
R^2	0.3979	0.5036	0.5846	0.6723	0.4245	0.4952
Adjusted R^2	0.1721	0.2199	0.4288	0.4851	0.2087	0.2067
F-Statistics	1.7620	1.7753	3.7529	3.5904	1.9673	1.7165
Prob > F	0.2319	0.2381	0.0598	0.0675	0.1976	0.2498

5.4 Pricing of S&P 500 Options

Until this point, we have only measured the performance of the three specific models on artificially simulated data. As a last test, we want to investigate the models' performance on real world market data, containing exceptionally more noise than our generated data. However, as concluded earlier, the real market RND cannot be observed directly. Thus, instead of directly comparing the RNDs, we estimate the RNDs from market data and value the option based on the estimated RND using *equation (2.7),* similar to the pricing of the hypothetical binary options in previous segment, but now with a conventional payoff structure, similar to *equation 2.3* in *section 2.2.1*.

For this test we evaluated the RND for two different time to maturities options on two different trading dates. The data analyzed is based on the historical SPX European option prices observed between June 1st, 2017 and December 29th, 2017, where the S&P 500 index is the underlying security. The risk-free rate is the observed one-month U.S. Treasury bill on that specific trading date and the spot price is simply the closing price of the S&P 500 on that specific trading date. The option prices are calculated as the average closing price between the bid and ask price of the option contract. To minimize noise in the dataset we further discarded option contracts with less than 20 open interests. The options were selected based on open interest and settled agreements and are thus assumed to be efficiently priced, liquid options. The descriptive statistics as well as the implied volatilities for the different time to maturities can be observed in *table 20* and *figure 14,* respectively. The given summary statistics and estimated RNDs for the models can be observed in *table 21* and *figure 15,* respectively.

The most noticeable observations from the summary statistics in *table 21* are the considerably higher kurtosis compared to the kurtosis generated from our simulations covered in *section 5.1*. Thus, we notice that the volatility skew, observed in *figure 14*, is prominently reflected in the RND and thus one could argue that the jump diffusion process is more dominate as a underlying process of the spot price. This observation is aligned with the findings of Jackwerth (2004), who argues that the volatility skew generated by jump diffusion models cannot sufficiently capture the skew observed in stock indices. Further, we can observe in the *figure 15*, that the majority of probability is concentrated around the mean, while we have higher probabilities on the more extreme values on the left tail, incorporating possible stochastic factors, such as negative market jumps. Therefore, one could further make the claim that the market prices incorporate a jump premium. We further see that the skewness is negative for all options. Thus, we can confirm the existence of a negative correlation between volatility and asset returns, similarly to the findings in previous literature. Lastly, as concluded from the previous sections, we observe from *table 21* that the mixture model is underestimating the negative skewness compared to the other models.

Table 22 represents the results from pricing the SPX call options based on the model estimated RNDs on two different trading dates with two different maturities respectively. Interpreting the table, we observe that the models can more accurately replicate the price of the two options with 17 and 111 days to maturity, both of which have lower estimated kurtosis and skewness compared to the other two options. We observe once more how the mixture model is able to fairly accurately predict the valuation of the call options with lower strikes but fall short for options with higher strike prices. Similar results can be interpreted from the pricing of a binary put option in previous *section 5.3*. We further notice that all models greatly overvalue the 39 days to maturity call option, concluding that the models estimate a higher probability to the right tail events than implied by the market. Moreover, we notice the pricing errors are relatively small for all models for the 111 days to maturity options, however, inspecting *figure 15b,* we notice that the kernel regression has allocated substantially more probability on the right tail of the density compared to the other two models. Observing the interval between moneyness 1.05 and 1.14, pricing a call option with a strike price within that region, utilizing the kernel estimated RND, would most certainly lead to a substantial overvaluation. Meanwhile, for the 17 days to maturity, we recognize that the models, especially the kernel regression model, undervalues the call option. More noticeable, from *figure 16b,* we observe that the same call option is well over the strike price (K=2375) on maturity, where the S&P 500 closed on $2500.23. Lastly, one might reason that the neural network once more performed better, or at least more consistently, compared to the other two models for all options and maturities, in terms of pricing SPX call options.

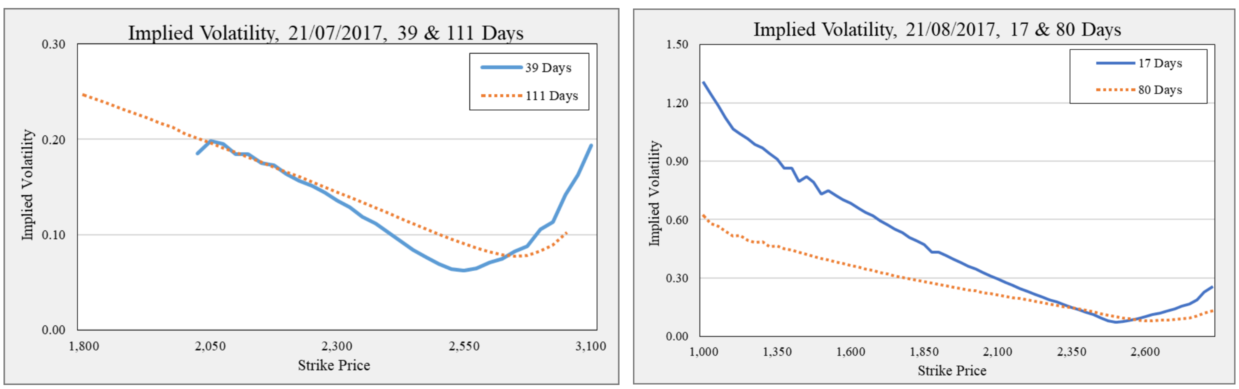
As a conclusion, real market data undoubtedly consists of more noise than our previously simulated option data. Thus, predicting well-behaved RNDs for the S&P 500 is a difficult task. Observing the results from *table 21* and *figure 15*, we notice that the estimated RNDs are characterized by additional excess kurtosis and negative skewness, indicating that the market prices may very well incorporate considerable jump diffusion processes. Pricing the call options, we notice that call options with higher strikes are, as expected, harder to adequately value, especially for the mixture model, which overvalues both options substantially. Moreover, one could make the argument that once again the neural network proved to be the most consistent model for valuating SPX call options based on the estimated RND.

TABLE 20. DESCRIPTIVE STATISTICS S&P 500 INDEX OPTIONS

Descriptive statistics for the four different call options with different time to maturities, at two different trading dates. Time to maturity is rounded to the closest date in terms of trading days, where there are 252 trading days annually. The spot price represents the closing price of S&P 500 index on the specific trading date. The risk-free rate is represented by the one-month U.S. Treasury bill.

Trading Date	21/07/2017	21/07/2017	21/08/2017	21/08/2017
Time to Maturity (Trading Days)	39	111	17	80
Spot Price	2472.54	2472.54	2428.37	2428.37
Strike Price	2520	2025	2375	2600
Option Price	7.75	448.85	66.10	3.15
Risk-Free Rate	0.012304	0.012304	0.012308	0.012308
Implied Volatility	0.065	0.2012	0.1373	0.0817

FIGURE 14. S&P 500 INDEX OPTIONS IMPLIED VOLATILITIES



Visual representation of the implied volatilities over strike prices observed on July 21st and August 21st, 2017 for the different time to maturities. The closing spot price on the 21/07/2017 was $2472.52 and $2428.37 on the 21/08/2017.

TABLE 21. SUMMARY STATISTICS ESTIMATED RNDS, S&P 500 INDEX OPTIONS

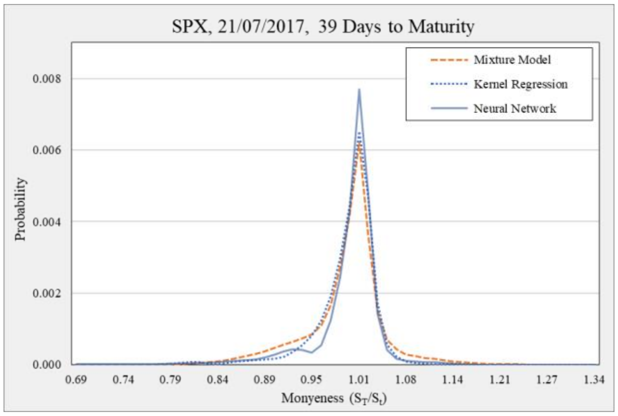
Summary statistics of the risk-neutral implied denstities estimated by the mixture model, kernel regression and neural network models on S&P 500 index options. The densities are evaluated over two different dates with two different maturities, respectively. The values for the mean and standard deviation are reported in terms of moneyness (S_T/S_t).

	21/07/2017					
	39 Days			**111 Days**		
Statistics	Mixture Model	Kernel Regression	Neural Network	Mixture Model	Kernel Regression	Neural Network
Mean	1.0005	1.0019	1.0033	0.9970	0.9999	1.0026
Std. Dev.	0.0487	0.0363	0.0396	0.0822	0.0722	0.0766
Skewness	-0.5240	-1.7712	-2.2018	-0.5339	-0.6876	-0.7950
Kurtosis	6.1989	7.4940	8.9064	5.2513	4.5755	5.4970

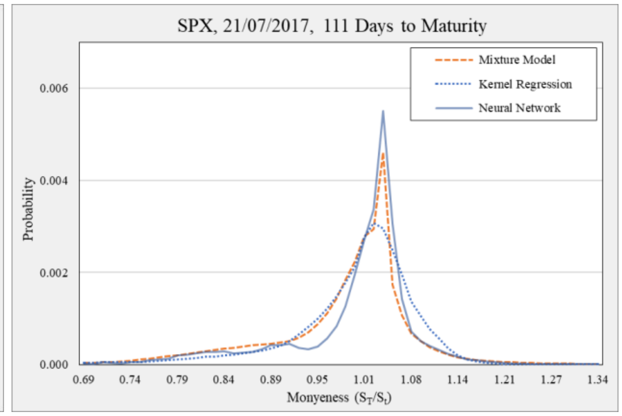
	21/08/2017					
	17 Days			**80 Days**		
Statistics	Mixture Model	Kernel Regression	Neural Network	Mixture Model	Kernel Regression	Neural Network
Mean	0.9986	1.0015	1.0197	0.9993	1.0022	1.0038
Std. Dev.	0.0305	0.0389	0.0423	0.0711	0.0745	0.0574
Skewness	-0.1769	-0.6998	-0.6096	-0.7226	-1.4984	-1.5448
Kurtosis	5.5613	5.8431	5.6744	5.0273	7.0228	6.8045

FIGURE 15. **MODEL ESTIMATED RISK-NEUTRAL DENSITIES, S&P 500 INDEX**

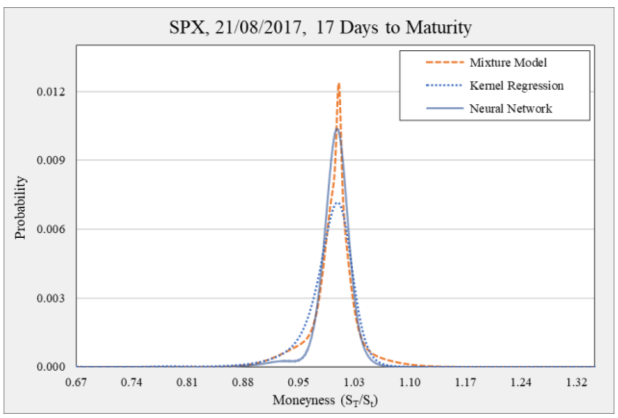
a)



b)



c)



d)

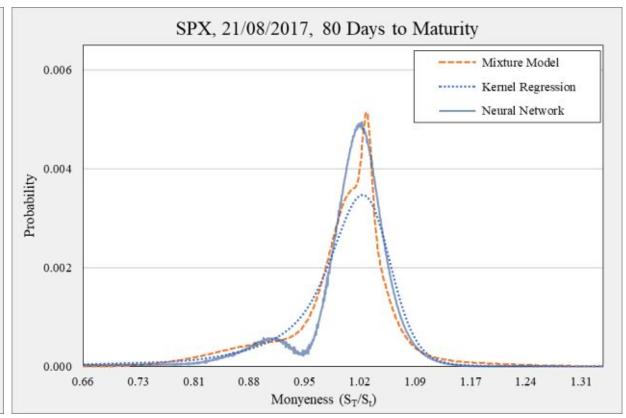


Illustration of the model estimated RNDs based on S&P 500 option data for four different options maturities, on two different trading dates. Figures a) and b) represent the estimated RNDs for the options observed on the 21/07/2017, where a) has 39 days to maturity and b) has 111 days to maturity. Figures c) and d) represent the estimated RNDs for the options observed on the 21/08/2017, where c) has 17 days to maturity and d) has 80 days to maturity.

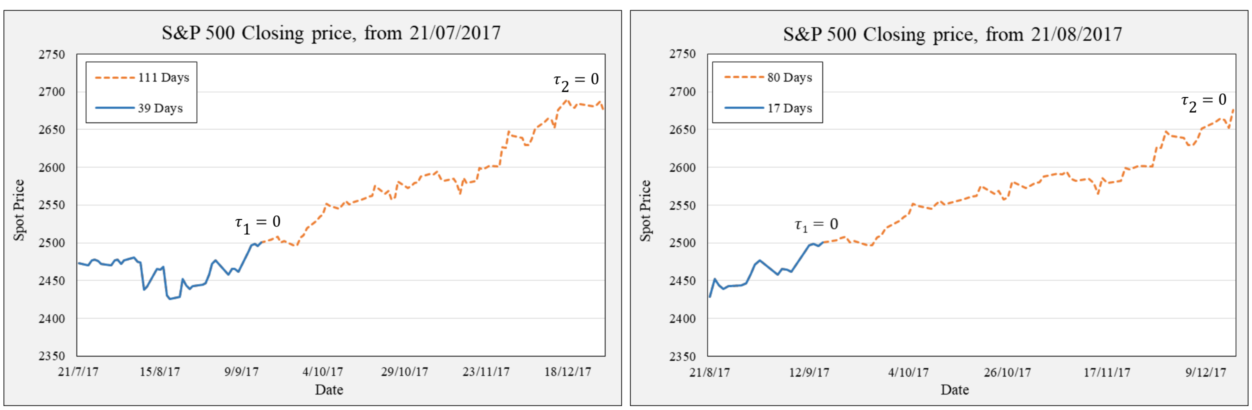
TABLE 22. S&P 500 INDEX CALL OPTION VALUATION

The following table represents the results from pricing four different S&P 500 call options on two different dates, with two different maturities each. The valuations are based on the RNDs estimated by the mixture, kernel regression and neural network models. The results are given as price in dollars at the given trading date and percentage pricing error is given in parentheses as the percentage difference from the true price, times 100, with percentage points as unit of measurement.

		21/07/2017		
Time to Maturity	True	Mixture Model	Kernel Regression	Neural Network
39 Days	7.7500	19.1970	11.8331	11.9541
		(147.704)	(52.685)	(54.246)
111 Days	448.8500	448.0497	465.1584	465.5493
		(-0.178)	(3.634)	(3.720)
		21/08/2017		
Time to Maturity	True	Mixture Model	Kernel Regression	Neural Network
17 Days	66.1000	63.8314	56.5342	62.0606
		(-3.432)	(-14.472)	(-6.111)
80 Days	3.1500	9.1159	4.5695	2.5374
		(189.392)	(45.064)	(-19.447)

FIGURE 16. S&P 500 CLOSING PRICE MOVEMENT

a) ***b)***



Visual representation of the closing prices of the S&P 500 index for two time intervals, where we further observe the maturity dates for the options as $\tau_i = 0$. Figure a) represents the time interval between 21/07/2017 to 29/12/2017, where $\tau_1 = 0$ on 15/09/2017 and $\tau_2 = 0$ is on the 29/12/2017. Figure b) represents the time interval between 21/08/2017 to 15/12/2017, where $\tau_1 = 0$ on 15/09/2017 and $\tau_2 = 0$ is on the 15/12/2017.

6. Conclusions

This thesis was conducted with the purpose of investigating how one can efficiently obtain and compare the performance of estimated option implied RNDs, and subsequently conclude which model is best suited for the purpose of obtaining the RNDs. The thesis presented the parametric mixture model, as well as the two nonparametric kernel regression and neural network models for estimating the option implied RNDs. By simulating the option prices, utilizing four different data-generation processes, the geometric Brownian motion, stochastic volatility, jump diffusion and stochastic volatility with jump diffusion, we remain in full control of the underlying stock process and can thus, by utilizing the closed-form solutions for the diffusion processes, obtain the theoretical or 'true' RND for the corresponding option prices. We continued by reporting the test results in *section 5* and further answer our three main research questions. *1)* By calculating the first four statistical moments of the RNDs, the mean, standard deviation, skewness and kurtosis, we can compare the moments of the estimated RNDs to the true RNDs and subsequently investigate which model can adequately imitate the features of the true RND. The results vary over data-generation processes and maturities; however, there is evidence identifying the flaws associated with the mixture model, where it falls short for options with shorter maturities, when the implied volatility skew is more evident. The kernel regression and neural network models' estimated RNDs are almost identical in terms of performance. However, we can conclude that the kernel regression model can replicate the moments slightly more efficiently in terms of precision with lower absolute percentage errors. *2)* Using the Mann-Whitney U test we can statistically compare the estimated RNDs to the true RND, where we reject the null hypothesis if the two distributions do not originate from the same population distribution. We find that we reject the null hypothesis for the mixture model on three occasions, when the RNDs are characterized by a distinct negative skew. We further reject the null hypothesis on one occasion for the kernel regression estimated RND following the stochastic volatility process. However, we fail to reject the null hypothesis for the neural network model for all data-generation processes and maturities. *3)* Constructing hypothetical binary call and put options, we investigate the valuation capabilities of the estimated RNDs compared to the options priced by the true RND. We recognized that the mixture model performed well on at-the-money options and the call options in general. However, we conclude that the neural network model is more consistent and further able to price both the call and put options adequately compared to the other models. In conclusion, considering all the results, we conclude that the nonparametric models are better suited for the purpose of obtaining the option implied RND for the tests conducted in this thesis, whereas the neural network is slightly more reliable in terms of overall performance. However, the mixture model has desirable aspects, where it outperforms the nonparametric models on smaller datasets and in terms of computational efficiency.

6.1 Limitations and Future Research

One of the main limitations of the conducted research is the use of a limited dataset of simulated options. For the purpose of this thesis we limited the set of simulated options to one trading date over three maturities. Thus, by adding more scenarios in terms of periods with higher volatilities, longer maturities or conduct different tests for different sets of simulation parameters, one can more efficiently test the robustness and determine the models' strengths and shortcomings. Secondly, due to time limitations, we recognize that 19,500 options per data-generation process is considered to be a relatively small training set. We further realize that the chosen number of iterations for the Monte Carlo Simulations might not be considered sufficiently for spot prices following a stochastic volatility process or jump diffusion, where further implementations of improved simulations techniques, such as variance reduction, would have been a positive contribution in terms of increased precision. Lastly, we recognize the limitation of our small sample of true RNDs, where it is difficult to fully determine the overall model performance for each data-generation process and maturity with a limited set of observations.

For future research, we propose to further investigate the performance of nonparametric methods for estimating the option implied RND, where new approaches and models are constantly being developed. Secondly, by utilizing diversified datasets, constituted by higher and lower volatilities and different underlying processes such as stochastic interest rates, one can better determine the beneficial characteristics of certain methods and approaches. Further, by analyzing the RNDs of different options, such as interest rates and macroeconomic derivatives, one can conceivably gain valuable insights into future market conditions and beliefs. Lastly, by observing real market implied RNDs for periods with low and high volatility, for instance before and during the crash of 2008, one can better investigate the differences in RND characteristics and how changes to future market beliefs impact the option implied RNDs.

The research conducted on option implied RNDs is still in its early stages and there is still a substantial amount of work to be done within the field of option implied information and return probability distributions. The information contained in RNDs in terms of market representative agent's utility, future asset returns and macroeconomic events are still an unexplored field for many researchers, financial institutions and fiscal policy makers. Meanwhile, we recognize that the choice of a state-of-the-art method for estimating the option implied RND is practically impossible, since different methods are able to capture certain aspects of the underlying factors that determine the option implied RND better than others. Thus, by recognizing those underlying processes, such as stochastic volatility and jump diffusion, one can instead determine the most reliable method for obtaining the RND for a specific purpose.

References

Aıt-Sahalia, Y., & Lo, A. W. (2000). Nonparametric risk management and implied risk aversion. *Journal of Econometrics, 94*(1-2), 9-51.

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

Albrecher, H., Mayer, P., Schoutens, W., & Tistaert, J. (2007). The little Heston trap. *Wilmott*(1), 83-92.

Applebaum, D. (2009). *Lévy processes and stochastic calculus*: Cambridge university press.

Arrow, K. J. (1974). General economic equilibrium: purpose, analytic techniques, collective choice. *The American Economic Review, 64*(3), 253-272.

Arrow, K. J., & Debreu, G. (1954). Existence of an equilibrium for a competitive economy. *Econometrica: Journal of the Econometric Society*, 265-290.

Bahra, B. (1997). Implied risk-neutral probability density functions from option prices: theory and application.

Bakshi, G., Cao, C., & Chen, Z. (1997). Empirical performance of alternative option pricing models. *The journal of finance, 52*(5), 2003-2049.

Bates, D. S. (1991). The Crash of '87: Was It Expected? The Evidence from Options Markets. *The journal of finance, 46*(3), 1009-1044.

Bates, D. S. (2008). The market for crash risk. *Journal of Economic Dynamics and Control, 32*(7), 2291-2321.

Black, F. (1976). The pricing of commodity contracts. *Journal of financial economics, 3*(1-2), 167-179.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy, 81*(3), 637-654.

Bouden, A. (2007). *Comparing Risk Neutral Density Estimation Methods using Simulated Option Data.* Paper presented at the World Congress on Engineering.

Breeden, D. T., & Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of business*, 621-651.

Brennan, M. J. (1979). The pricing of contingent claims in discrete time models. *The journal of finance, 34*(1), 53-68.

Bu, R., & Hadri, K. (2007). Estimating option implied risk-neutral densities using spline and hypergeometric functions. *The Econometrics Journal, 10*(2), 216-244.

Cauchy, A. (1847). Méthode générale pour la résolution des systemes d'équations simultanées. *Comp. Rend. Sci. Paris, 25*(1847), 536-538.

Celis, O. S., Liang, L., Lemmens, D., Tempere, J., & Cuyt, A. (2015). Determining and benchmarking risk neutral distributions implied from option prices. *Applied Mathematics and Computation, 258*, 372-387.

Chacón, J. E., & Duong, T. (2018). *Multivariate kernel smoothing and its applications*: Chapman and Hall/CRC.

Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of financial economics, 10*(4), 407-432.

Conrad, J., Dittmar, R. F., & Ghysels, E. (2013). Ex ante skewness and expected stock returns. *The journal of finance, 68*(1), 85-124.

Cont, R. (1997). Beyond implied volatility: Extracting information from options prices. *Econophysics. Dordrecht: Kluwer*.

Cont, R., & Tankov, P. (2004). Nonparametric calibration of jump-diffusion option pricing models. *Journal of computational finance, 7*, 1-49.

Cooper, N. (1999). *Testing techniques for estimating implied RNDs from the prices of European-style options.* Paper presented at the Proceedings of the BIS workshop on implied PDFs, Bank of International Settlements, Basel.

Coutant, S., Jondeau, E., & Rockinger, M. (2001). Reading PIBOR futures options smiles: The 1997 snap election. *Journal of Banking & Finance, 25*(11), 1957-1987.

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

Cox, J. C., Ross, S. A., & Rubinstein, M. (1979). Option pricing: A simplified approach. *Journal of financial economics, 7*(3), 229-263.

Debreu, G. (1959). *Theory of value: An axiomatic analysis of economic equilibrium*: Yale University Press.

Dennis, P., & Mayhew, S. (2002). Risk-neutral skewness: Evidence from stock options. *Journal of Financial and Quantitative Analysis, 37*(3), 471-493.

Efromovich, S. (2008). *Nonparametric curve estimation: methods, theory, and applications*: Springer Science & Business Media.

Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications, 14*(1), 153-158.

Feng, P., & Dang, C. (2016). Shape constrained risk-neutral density estimation by support vector regression. *Information Sciences, 333*, 1-9.

Figlewski, S. (2018). Risk-neutral densities: A review. *Annual Review of Financial Economics, 10*, 329-359.

Franke, G., Stapleton, R. C., & Subrahmanyam, M. G. (1999). When are Options Overpriced? The Black—Scholes Model and Alternative Characterisations of the Pricing Kernel. *Review of Finance, 3*(1), 79-102.

Gemmill, G., & Saflekos, A. (2000). How useful are implied distributions? Evidence from stock index options. *Journal of derivatives, 7*(3), 83-98.

Giamouridis, D., & Tamvakis, M. (2002). Asymptotic distribution expansions in option pricing. *Journal of derivatives, 9*(4), 33.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*: MIT press.

Grith, M., Härdle, W. K., & Schienle, M. (2012). Nonparametric estimation of risk-neutral densities. In *Handbook of computational finance* (pp. 277-305): Springer.

Grossman, S. J., & Zhou, Z. (1996). Equilibrium analysis of portfolio insurance. *The journal of finance, 51*(4), 1379-1403.

Han, B. (2007). Investor sentiment and option prices. *The Review of Financial Studies, 21*(1), 387-414.

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

Hilpisch, Y. (2015). *Derivatives analytics with Python: data analysis, models, simulation, calibration and hedging*: John Wiley & Sons.

Hull, J. C. (2017). *Options, Futures, and Other Derivatives*: Pearson.

Jackwerth, J. C. (1999). Option implied risk-neutral distributions and implied binomial trees: A literature review. *Journal of derivatives, 7*(2), 66-82.

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

Jackwerth, J. C. (2004). Option-implied risk-neutral distributions and risk aversion.

Jarrow, R., & Rudd, A. (1982). Approximate option valuation for arbitrary stochastic processes. *Journal of financial economics, 10*(3), 347-369.

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

Jondeau, E., & Rockinger, M. (2000). Reading the smile: the message conveyed by methods which infer risk neutral densities. *Journal of International Money and Finance, 19*(6), 885-915.

Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management science, 48*(8), 1086-1101.

Lai, W.-N. (2014). Comparison of methods to estimate option implied risk-neutral densities. *Quantitative Finance, 14*(10), 1839-1855.

Lynch, D. P., & Panigirtzoglou, N. (2002). *Summary statistics of implied probability density functions and their properties.* Paper presented at the EFMA 2002 London Meetings.

Madan, D. B., Carr, P. P., & Chang, E. C. (1998). The variance gamma process and option pricing. *Review of Finance, 2*(1), 79-105.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 50-60.

Martin, I. (2017). What is the Expected Return on the Market? *The Quarterly Journal of Economics, 132*(1), 367-433.

Martin, I., & Wagner, C. (2018). What is the Expected Return on a Stock? *Journal of Finance, forthcoming*.

Mayhew, S. (1995). Implied volatility. *Financial Analysts Journal, 51*(4), 8-20.

Mayshar, J., & Benninga, S. (1997). Heterogeneity and Option Pricing. *Available at SSRN 15136*.

Melick, W. R., & Thomas, C. P. (1997). Recovering an asset's implied PDF from option prices: an application to crude oil during the Gulf crisis. *Journal of Financial and Quantitative Analysis, 32*(1), 91-115.

Merton, R. C. (1973). Theory of rational option pricing. *Theory of Valuation*, 229-288.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The journal of finance, 29*(2), 449-470.

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

Munk, C. (2013). *Financial asset pricing theory*: OUP Oxford.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications, 9*(1), 141-142.

Pritsker, M. (1998). Nonparametric density estimation and tests of continuous time interest rate models. *The Review of Financial Studies, 11*(3), 449-487.

Ritchey, R. J. (1990). Call option valuation for discrete normal mixtures. *Journal of Financial Research, 13*(4), 285-296.

Rockinger, M., & Jondeau, E. (2002). Entropy densities with an application to autoregressive conditional skewness and kurtosis. *Journal of Econometrics, 106*(1), 119-142.

Rompolis, L. S., & 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.

Rosenberg, J. V., & Engle, R. F. (2002). Empirical pricing kernels. *Journal of financial economics, 64*(3), 341-372.

Ross, S. A. (1976). Options and efficiency. *The Quarterly Journal of Economics, 90*(1), 75-89.

Rouah, F. D. (2011). Euler and milstein discretization. *Documento de trabajo, Sapient Global Markets, Estados Unidos. Recuperado de www. frouah. com*.

Rubinstein, M. (1976). The valuation of uncertain income streams and the pricing of options. *The Bell Journal of Economics*, 407-425.

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.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1988). Learning representations by back-propagating errors. *Cognitive modeling, 5*(3), 1.

Santos, A., & Guerra, J. (2015). Implied risk neutral densities from option prices: Hypergeometric, spline, lognormal, and Edgeworth functions. *Journal of Futures Markets, 35*(7), 655-678.

Schittenkopf, C., & Dorffner, G. (2001). Risk-neutral density extraction from option prices: improved pricing with mixture density networks. *IEEE Transactions on Neural Networks*, *12*(4), 716-725.

Schwert, G. W. (1989). Why does stock market volatility change over time? *The journal of finance, 44*(5), 1115-1153.

Shimko, D. (1993). Bounds of probability. *Risk, 6*(4).

Shonkwiler, R. W. (2013). *Finance with Monte Carlo*: Springer.

Söderlind, P., & Svensson, L. (1997). New techniques to extract market expectations from financial instruments. *Journal of Monetary Economics, 40*(2), 383-429.

Steeley, J. M. (2004). Stock price distributions and news: Evidence from index options. *Review of Quantitative Finance and Accounting, 23*(3), 229-250.

Syrdal, S. A. (2002). *A study of implied risk-neutral density functions in the Norwegian option market*: Norges Bank.

Tankov, P., & Voltchkova, E. (2009). Jump-diffusion models: a practitioner's guide. *Banque et Marchés, 99*(1), 24.

Toft, K. B., & Prucyk, B. (1997). Options on leveraged equity: Theory and empirical tests. *The journal of finance, 52*(3), 1151-1180.

Tompkins, R. G. (2001). Implied volatility surfaces: uncovering regularities for options on financial futures. *The European Journal of Finance*, *7*(3), 198-230.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 359-372.

Yatchew, A., & Härdle, W. (2006). Nonparametric state price density estimation using constrained least squares and the bootstrap. *Journal of Econometrics*, *133*(2), 579-599.

Appendix

a. Derivation of the Breeden-Litzenberger equation

Breeden and Litzenberger (1978) stated that in a market with a continuum of quoted strikes, the risk-neutral forward transition density equals the price of an infinitely small butterfly spread portfolio strategy, such that,

$$\frac{\partial^2 c(S,K)}{\partial K^2} = e^{-rT} q(S_T)_{K=S_T} = \lim_{\varepsilon \to \infty} \frac{c_1 + c_3 - 2c_2}{\varepsilon^2}.$$

If c_i is the value of the i-th call option in a cross-sectional set of option prices, on the underlying asset S, where $t = 0$, we can define the option price under risk-neutral measure as,

$$c(S,K) = e^{-rT} E^Q\{\max(0, S_T - K)\},$$

which we can express as follows,

$$c(S,K) = e^{-rT} \int_K^\infty (S_T - K) q(S_T) dS_T,$$

where $q(S_T)$ is the risk-neutral distribution at S_T. By differentiating the call price once with respect to K,

$$\frac{\partial}{\partial K} c(S,K) = \frac{\partial}{\partial K} e^{-rT} \left[-(K-K)q(K) + \int_\infty^K (-1) q(S_T) dS_T \right],$$

which results in the simplified solution,

$$= -e^{-rT} \int_K^\infty q(S_T) dS_T.$$

Taking the partial derivative of $c(S,K)$ with respect to K a second time,

$$\frac{\partial^2}{\partial K^2} c(S,K) = \frac{\partial^2}{\partial K^2} e^{-rT} \left[-\int_K^\infty q(S_T) dS_T \right],$$

$$= e^{-rT} \int_{\infty}^{K} q(S_T) dS_T,$$

where $\int_{\infty}^{K} dS_T = 1$, thus we get the final derivation, provided by Breeden and Litzenberger (1978),

$$\frac{\partial^2 c(S,K)}{\partial K^2} = e^{-rT} q(S_T),$$

or more commonly expressed as,

$$q(S_T) = e^{rT} \frac{\partial^2 c(S,K)}{\partial K^2}.$$

b. Computation of Moments

Given random variable X with its density function $g(x)$, we can define the first four moments, mean, variance, skewness and kurtosis. The first moment, the mean, is the weighted average of random variable X and defined as,

$$m = \int_{-\infty}^{\infty} x g(x) dx.$$

The second moment, the variance, is the weighted average of squared deviation from the mean and measures the dispersion of a density. We can define the variance as,

$$v = \sqrt{\int_{-\infty}^{\infty} (x-m)^2 g(x) dx}\,.$$

The third moment, skewness, is a measurement of asymmetry in the density and defines the existence and magnitude of distinct tails in the density, where a Gaussian density, or normal density has a skewness of 0. Positive skewness indicates that the density has a larger right tail or positively skewed, while negative skewness indicates a larger left tail and is thus left skewed. We can define skewness as,

$$s = \frac{1}{v^3} \int_{-\infty}^{\infty} (x-m)^3 g(x) dx,$$

The fourth moment, the kurtosis, defines the "tailedness" of a density, where a density with higher kurtosis has fatter tails and subsequently assigns a higher probability to more extreme events. Kurtosis of a certain density is measured relative to a Gaussian distribution that has a kurtosis equal to 3. We can define kurtosis as,

$$k = \frac{1}{v^4} \int_{-\infty}^{\infty} (x - m)^4 g(x) dx$$

Given that we know the density function $g(x)$ for the random variable X, we can compute the four first moments by numerically computing the integrals defined earlier. Calculations of the defined moments are performed numerically utilizing the trapezoidal integration rule (Syrdal, 2002).