OPTION PRICING AND MACHINE LEARNING: A COMPARISON OF BLACK-SCHOLES, BACHELIER, AND ARTIFICIAL NEURAL NETWORKS

Research submitted by

EDEN GROSS

in partial fulfilment of the requirements for the degree of

MASTER OF COMMERCE

specialising in

ACTUARIAL SCIENCE

School of Management Studies

Section of Actuarial Science

University of Cape Town

July 2021

Supervisor:

Associate Professor Ryan Kruger

Abstract

Practitioners and academics alike have applied the Black-Scholes model (or derivatives thereof) when pricing options practically since the introduction of the model in 1973. The recent coronavirus pandemic and the oil futures price crash of April 2020 have caused major markets to briefly switch to the less widely-known Bachelier model to price derivatives, as the model allows for negative strikes on the underlying. This study evaluates the predictive ability and accuracy of both the Bachelier model and the Black-Scholes model when pricing European call options on the Standard & Poor's (S&P) 500 Index using five different volatility estimation methods. Moreover, it then compares the forecasts of the two parametrised models to a deep feed-forward artificial neural network which is also used to price such options. Overall, the artificial neural network is statistically superior in its predictive ability relative to both of the parameterised models, and the Black-Scholes model is statistically superior in its predictive ability to the Bachelier model.

UNIVERSITY OF CAPE TOWN

FACULTY OF COMMERCE

Igniting Knowledge and Opportunity

Declaration

COMPULSORY DECLARATION:

1. This dissertation has been submitted to Turnitin (or equivalent similarity and originality checking software) and I confirm that my supervisor has seen my report and any concerns revealed by such have been resolved with my supervisor.
2. I certify that I have received Ethics approval (if applicable) from the Commerce Ethics Committee.
3. This work has not been previously submitted in whole, or in part, for the award of any degree in this or any other university. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works of other people has been attributed, and has been cited and referenced.

Student number	GRSEDE001
Student name	Eden Gross
Signature of Student	Signed by candidate
Date:	23 July 2021

Acknowledgements

Once again, I'd like to thank my family first and foremost. I could not have done any of this without you and your endless love and support.

Second, I'd like to thank Associate Professor Ryan Kruger, who has once again agreed to supervise my work. It is unlikely that this dissertation would have been completed without your endless help, support, and guidance.

Last, to the Section of Actuarial Science, for allowing me to undertake this research.

Table of Contents

List of Tables

List of Figures

1. Introduction

The pricing of options on global exchanges such as the Chicago Mercantile Exchange (CME) has followed the option pricing model developed by Black and Scholes (1973) since as early as 1975 (Merton, 1998). However, for a brief period during the first six months of 2020, the Black-Scholes model was replaced by a less familiar model that pre-dated it by over seven decades - the Bachelier model. The Bachelier model allows for the price of the underlying asset of an option to take on negative values, something that is not possible when using the former model.

The recent pandemic of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus and its accompanying disease, the coronavirus disease 2019 (COVID-19), caused some severe damage in global markets in the first half of 2020. One of the most noticeable disruptions was the emergence of government-sanctioned lockdowns in many countries around the world, many of which overlapped. These overlapping lockdowns have restricted movement not only between countries, but within them too. This meant that the transportation industry, be it via aeroplanes, ships, or motor vehicles, has suffered tremendously due to the sharp decline in usage, a decline whose reversal's date or time frame was unknown at the time.

This sharp decline in movement between and within countries has caused the demand for oil to plummet (Ma & Wang, 2020). The lack of demand had resulted in an unprecedented low price for reference oil futures in the United States (U.S.), the West Texas Intermediate (WTI). The WTI saw its May 2020 futures prices plummet into the negative due to the combined effect of severely reduced demand and the high volatility of market sentiment regarding financial markets, in general, and oil, in particular (Huang & Zheng, 2020).

There were several reasons for the severely volatile oil market prices for the WTI around the end of April 2020. First, oil reserves around the world were not depleted as quickly as they would have been under normal circumstances, which led to excess supply relative to depleted demand. In addition, the failure to reach an agreement for a reduction in the supply of crude oil a month earlier by oil-producing countries, coupled with the process's usual petropolitics, aided in the price crash as the market saw no real indication of changes (decreases) in the supply of crude oil (Ma & Wang, 2020).

Second, traders entered the futures market to profit off of the petropolitical issues identified above, in the belief that the price of crude oil will revert to its appropriate equilibrium price (Ma & Wang, 2020). The inevitable exit of traders from the market has put pressure on the already depressed price, which was simultaneously coupled with the emergence of lockdowns. Many traders in oil futures are not qualified to make or take possession of the underlying (Ma & Wang, 2020), often resulting in them closing out their positions before contract expiry, further depressing the price. However, given the emergence of lockdowns and depressed retail demand for oil, many of these traders would have struggled to exit their positions. Those left who were qualified to take possession of the underlying found themselves with severely high storage costs due to a shortage in storage capacity for oil in the U.S. (Ma & Wang, 2020).

Hence, several conditions, which collectively are required to allow for the negative prices of futures contracts, were met. These included aspects such as the announcement by the CME stating that energy contracts can trade at negative prices on its New York Mercantile Exchange (NYMEX) just five days before the crash (Sui & Yang, 2020); the events taking place within a week of the May contracts' expiry date; the illiquidity issues created in the market due to an influx of traders incapable of taking and making delivery on the underlying; and the petropolitical issues mentioned above.

The negative futures price phenomenon led some groups and financial marketplaces (such as the CME Group, Inc.) to temporarily[1] shift to the Bachelier option pricing model (CME Group, Inc., 2020). As mentioned above, the Bachelier model allows for negative prices when considering the price of the underlying.

This shift to using the Bachelier model to price options, even temporarily, raises several questions, one of which is the question surrounding the superiority in the predictive ability of the Bachelier model over the Black-Scholes model, the latter of which is widely considered to be the industry standard when it comes to option valuation (Nunes, et al., 2018). This industry standard has been widely used by practitioners and academics alike in their assessment of option prices around the world. Hence, the move by the CME Group,

[1] The CME Group, Inc., shifted to the Bachelier model on 21 April 2020 and shifted back to its prior model on 13 August 2020 (Yuan, 2020).

Inc., to use the Bachelier model may have resulted in significant changes to option markets and their members around the world.

Since the Black-Scholes model has been the popular choice to value options, it is unsurprising that academics and practitioners alike have attempted to dethrone the parametric model with some other model. This is evident by attempts to improve the model parametrically, for example, by trying to relax some of its constraining assumptions (see, for example, Scott (1987); see Section 2.1 for a discussion of such studies) or by the introduction of a superior non-parametric model (an early example of which can be seen in the study by Hutchison, Lo, and Poggio (1994); see Section 4.2 for a discussion of such studies).

The attempt to identify a non-parametric model that is superior to the Black-Scholes model has advanced significantly with advancements in machine learning methods such as artificial neural networks (ANNs), and the introduction of these in the financial literature. This kind of application allowed those constructing an ANN to learn the inherent relationships between various market variables, forces, and the resultant price, without the use of an explicit formula (as used in parametric models such as Black-Scholes and Bachelier). This has led to many studies comparing the performance and predictive ability of the two kinds of models, usually considering different construction methods of the ANN and different data on which options are priced.

As the U.S. market has played a major role in the global financial market, it is unsurprising that many studies applying ANNs to option pricing focus on this market. Studies examining option pricing in the U.S. often make use of the Standard & Poor's (S&P) 500 Index. Examples of such studies, discussed further in Chapter 4, are those of Bakshi, Cao, and Chen (1997); Garcia and Gençay (2000); Baruníkova and Baruník (2011); and, more recently, Vach (2019).

Hence, while the negative price of oil futures experienced in 2020 may be considered an anomaly (Sui & Yang, 2020), the introduction of the Bachelier model to industry raises a new avenue in research, one which can focus on the comparison of the predictive accuracy of the Bachelier model to (i) the Black-Scholes model, and (ii) an ANN designed to predict the price of options. This study aims to provide exactly that – a comparison of the predictive accuracy of the two 'traditional' (parametric) models used in industry, as well as a comparison of predictive accuracy between these traditional models and ANNs.

Moreover, several tools are available to estimate implied volatility when pricing options and this, too, will be considered in this study.

Therefore, this study aims to compare the predictive ability of the re-introduced Bachelier model to the more accomplished Black-Scholes and ANN models when pricing European call options, using several volatility measures. It is important to note that this study does not aim to find the optimal structure of an ANN when it comes to pricing options. Therefore, this study will rely on the conclusions made by previous studies which examine the optimal construction of an ANN when employed in pricing options, specifically that of Vach (2019). Moreover, several other principles established in the literature (such as those presented in Baruníkova and Baruník, 2011) will be employed when selecting the data and constructing the ANN used in this study.

While the use of the Bachelier model was brief, its re-introduction has sparked renewed interest in the model and its usefulness in practice. Practitioners who use some form of an in-house model may incorporate certain aspects of the Bachelier model when pricing options in the future, along with those of other models. Hence, the comparison undertaken in this study is of both academic and practical use in the investment and trading world.

The remainder of this study is structured as follows: Chapter 2 introduces the two traditional options pricing models: the Bachelier model and the Black-Scholes model; Chapter 3 then introduces some machine learning concepts before discussing the mathematics behind the artificial neural network; Chapter 4 provides a literature review of the studies comparing the predictive nature of the traditional models described above to that of an artificial neural network. Chapter 5 discusses the data and methodology employed in this study, followed by the results discussed and analysed in Chapter 6. Finally, this study is concluded in Chapter 0.

2. Option Pricing Models

This chapter introduces the two ‘traditional’ models which are examined in this study, namely the Black-Scholes model and the Bachelier model. It begins by examining the intuition surrounding each model before looking at each model’s assumptions and technical definitions.

This chapter also contains a literature review of the models, whereas Chapter 4 provides a literature review comparing the predictive ability of these ‘traditional’ models to the predictive ability of the ANN.

2.1. The Black-Scholes Model

The Black-Scholes model was introduced by Fischer Black and Myron Scholes in their 1973 study. In their study, the authors present a closed-form option valuation model with no unknown parameters. The authors claim that this model is different to those which preceded it due to the sole use of known parameters, i.e., all option valuation models which preceded the Black-Scholes model involved some parameters whose values were unknown. However, as will be described in Section 2.2, this is not true. The Bachelier model, presented by Louis Bachelier in 1900, did not include any unknown parameters, yet it failed to be mentioned in the study presented by Black and Scholes (1973).

Black and Scholes (1973) base their work on two main studies. First, they note the work of Thorp and Kassouf (1967), in which the authors introduce the notion of a hedged position using warrants involving both shares and options. Black and Scholes note, however, that Thorp and Kassouf fail to notice that, in equilibrium, the expected return earned by an investor on the hedged position must equal the risk-free rate earned in the market. Second, the authors note the introduction of the concept that the value of an option is dependent on the share price of the underlying stock, as presented by Samuelson and Merton (1969). Black and Scholes build on these two concepts, the known expected return in equilibrium and the dependency of option values on share prices, to build their own option valuation model. Black and Scholes recognise, however, that the volatility of the option is a crucial input into their model, as it is not constant.

Black and Scholes (1973) base their option pricing model on the assumption that the underlying's price can be modelled using the following stochastic model:

$$S_t^{Black-Scholes} := S_0 \exp\left(\sigma^{Black-Scholes} W_t + \left(r - \frac{(\sigma^{Black-Scholes})^2}{2}\right) t\right) \quad (1)$$

where $S_t^{Black-Scholes}$ is the price of the underlying at time $t \in [0, T]$ as modelled by the Black-Scholes model, denoted simply by S_t in the remainder of this subsection; S_0 is the price of the underlying at time 0; $\exp(\upsilon)$ is the natural exponential function of the input variable υ; r is the risk-free rate of interest earned for a period of length t; $\sigma^{Black-Scholes}$ is the model's standard deviation, denoted simply by σ in the remainder of this subsection; and $(W_t)_{t \in [0,T]}$ is a random variable that follows a standard Brownian motion process at time t (Schachermayer & Teichmann, 2008).

The Black-Scholes model is based on a set of conditions that the authors refer to as 'ideal'. These include the following:

i. a known and constant short-term interest rate
ii. the share price is log-normal in distribution with a constant rate of return variance
iii. no dividends are paid to the shareholder
iv. the option taken out is European[2] in nature
v. transaction costs are zero for both shares and options
vi. money to buy the share can be borrowed in any quantity at the known and constant short-term interest rate
vii. short selling is possible in any quantity without penalties

Keeping the above assumptions in mind, the value of an option is not dependent on any unknown variables besides for the share's price and the contract's term, together with some constants outlined above (Black & Scholes, 1973). Black and Scholes (1973) propose a hedged position which includes a long position in the share coupled with a short position in the option. This position's value will then only be dependent on the term of the option and the constant values outlined above.

[2] A European option means that an option-holder can only exercise the option at the end of its term but not before such time.

Keeping with the notation introduced for Equation (1), Black and Scholes (1973) define the value of a call option to be $C_{S_t,T-t} \coloneqq C(S_t, T-t)$, where S_t is the value of the underlying share at time $t \in [0,T]$ and $T-t$ denotes the term of the option, often simplified to be written as t^*, and where the price of the underlying share and the term of the option contract are independent variables. The reciprocal of the partial derivative of C_{S_t,t^*} with respect to the price of the underlying is then equal to the number of options that forms an investor's short position per one share in the long position, i.e., if the number of options that forms an investor's short position per one share in the long position is labelled by $C_1^{-1}(S_t, t^*)$, then $C_1^{-1}(S_t, t^*) = \frac{1}{C_1(S_t,t^*)}$, where $C_1(S_t, t^*)$ represents the first derivative of C_{S_t,t^*} with respect to S_t. For an explanation as to why this hedged position is not dependent on S_t, see Black and Scholes (1973).

If such a hedged position's number of options (which are sold short) is continuously adjusted, then there is no risk in the position itself (Black & Scholes, 1973), leading to the conclusion that the return earned on such a hedged position is equal to the risk-free rate, r. Since this continuous hedging is performed in the short term, it follows that the short-term interest rate is equal to the risk-free rate of return, which corresponds to the model's assumptions listed above.

Using the assumptions above and stochastic calculus, Black and Scholes (1973) develop the following stochastic differential equation to determine the value of an option.

$$C_2(S_t, t^*) = rC_{S_t,t^*} - rS_tC_1(S_t, t^*) - \frac{1}{2}\sigma^2 S_t^2 C_{11}(S_t, t^*) \tag{2}$$

where C_{S_t,t^*} and $C_1(S_t, t^*)$ are as defined above; $C_2(S_t, t^*)$ is equal to the partial derivative of C_{S_t,t^*} with respect to t^*; and $C_{11}(S_t, t^*)$ is equal to the partial derivative of $C_1(S_t, t^*)$ with respect to S_t. Moreover, σ^2 is the variance of the share's return rate; and r is the risk-free rate as identified above.

Equation (2) can then be coupled with the payout pattern of a call option as follows, using T to denote the date at the end of the term of the option, i.e., the maturity date.

$$C(S_t, T) = \begin{cases} S_T - X, & S_T \geq X \\ 0, & S_T < X \end{cases} \tag{3}$$

where X is the option's strike price.

When combining the differential equation depicted in Equation (2) with the payoff condition depicted in Equation (3), the only solution for the stochastic differential equation which satisfies the boundary condition is then:

$$C_{S_t,t^*} = S_t\Phi(d_1) - Xe^{-rt^*}\Phi(d_2) \tag{4}$$

$$d_1 = \frac{\ln\frac{S_t}{X} + \left(r + \frac{1}{2}\sigma^2\right)(t^*)}{\sigma\sqrt{t^*}}$$

$$d_2 = \frac{\ln\frac{S_t}{X} + \left(r - \frac{1}{2}\sigma^2\right)(t^*)}{\sigma\sqrt{t^*}}$$

where $\Phi(d)$ is the cumulative distribution function of the normal distribution at point d.

Using Equation (4), it can be seen that the expected return on the share underpinning the option is irrelevant to the value of the option. It is only relevant when the expected return on the option is of interest (Black & Scholes, 1973).

Now that the model itself has been described, attention can be turned to a review of the literature surrounding the Black-Scholes model for option pricing. Besides the degree of accuracy offered by the model when pricing options, two of the major areas of debate in the literature are the assumptions upon which the model is based[3], and the method employed when calculating the volatility input parameter of the model. Hence, the remainder of this section discusses these in more detail.

Whenever a mathematical formula is presented as a blueprint to a process that is believed to be non-parametric, scepticism surrounding the parameterisation's completeness is often encountered. Essentially, how can users of such a parameterisation of a process (or, equivalently, a model) be certain that the model's creators have encompassed and considered every variable that affects the output of such a process? Early studies considering this issue in the context of the Black-Scholes model are those presented by Long (1974) and Smith (1976).

Long (1974) makes an explicit remark of the completeness assumption of input variables by stating that the model does not take into account aspects such as the price of beer. He then explains that, while beer is obviously not a factor, he could have easily said 'the value of the market portfolio' or 'the rate of inflation' in its place, highlighting that

[3] Recall the seven assumptions made by Black and Scholes (1973), as on page 6.

some of the variables that are not considered by the Black-Scholes model may, perhaps, seem to exhibit a likely effect on the output of such a model. A reason for accepting such a simplification of the input variables of the model was given to maintain the linearity of the solution of the differential equation used to solve the model and, thereby, provide a closed-form solution for the Black-Scholes model (Long, 1974).

Another input that the model assumes to be unrelated is an investor's degree of risk appetite (Smith, 1976). Smith points out that many of the assumptions surrounding the model have been relaxed by various authors[4] and seem to have no meaningful impact on the analysis and results (Smith, 1976). On the other hand, Jagannathan (1984) shows that the investor's risk appetite may affect the value that the investor places on an option. Smith's discussions and conclusions of no meaningful impact have relaxed much of the criticism of the model's completeness when it comes to input variable considerations, while Jagannathan's conclusions, coupled with the extensive use of the model in practice, suggest that such a compromise is accepted by the market.

Other compromises of the Black-Scholes model which seem to be accepted by the market at some stage include the assumptions surrounding the log-normal distribution of share returns and the existence of frictionless markets, which do not hold and exist, respectively, as shown by Gastineau and Madansky (1979). A further compromise is the effects of transaction costs on the model's output, leading to biased predictions (Leland, 1985).

The inconsistencies noted above between the Black-Scholes model's predictions and market prices for options are discussed further by Rubinstein (1985). In his study, Rubinstein suggests that these differences are the results of the model's specification. He points out that two key assumptions are made by Black and Scholes (1973) when developing their model. First, they assume that the path of prices of the underlying is a continuous path when plotted against time. Second, they assume constant volatility. The latter assumption is paramount, with many studies attempting to tackle its shortfalls and consequences.

An early study examining the volatility input of the model is that offered by Scott (1987). In his study, Scott tries to relax the assumptions surrounding the variance of a

[4] For further detail, see Smith (1976).

stock's returns (assumed by Black and Scholes (1973) to be constant, as mentioned) by letting it change randomly, taking into account the dependence of a share's returns over time and the possible mean reversion that the returns may exhibit. While the study only examines options surrounding a single share and Scott fails to specify a model explicitly, it does find that Monte Carlo simulations of the proposed process used to update the variance of a share's returns offer marginally better pricing for said options. Therefore, Scott's findings illustrate the benefit of allowing the variance input to vary, i.e., relaxing the constant variance assumptions of the Black-Scholes model.

Supporting the work presented by Scott (1987), and published around the same time, Hull and White (1987) and Wiggins (1987) point out that the constant variance assumption of the Black-Scholes model is not realistic. However, the alternative models presented by the authors do not have closed-form solutions readily available for them to be widely applied (Heston, 1993), while the Black-Scholes model does offer a closed-form solution.

Heston (1993) also highlights the failure of the Black-Scholes model to be applied to options that use an underlying other than a share (such as bonds or currencies). Heston points out that the Black-Scholes model is based on a fundamental assumption of return normality with known and constant mean and variance, a recognition that is in agreement with the studies of Scott (1987), Hull and White (1987), and Wiggins (1987). However, Scott offers a stochastic volatility model as an alternative to the Black-Scholes model, where Scott's model assumes a spot price that is correlated to volatility and also offers a closed-form solution (Heston, 1993) – an improvement relative to the models presented in the three 1987 studies. The model proposed by Heston does yield a price for a European call option with an underlying which may be a bond or a currency, i.e., the underlying is not restricted to shares alone.

Other studies attempt to tackle the normality assumption. One such study is that presented by Kou (2002), in which the author builds on past studies (such as that by Hull and White, 1987) and proposes a double exponential jump-diffusion process to model the volatility of share prices when pricing options. Kou's model recognises the leptokurtic nature of returns (i.e., their left skewness and fatter tails relative to the normal distribution) and the convexity of the strike prices when the implied volatility of the option is plotted against its strike price (resulting in what has been termed a 'volatility smile' in the literature), yielding a model that incorporates both of these phenomena in its construction.

The model proposed by Kou is capable of pricing both call and put options, together with interest rate derivatives and more exotic options (Kou, 2002). Kou recognises that the model proposed has several limitations, a key one being the incompatibility of the model with hedging – a result of the jump process yielding incomplete markets.

As mentioned earlier, a key avenue in the literature is concerned with the method used to calculate the volatility needed as a model parameter. This is especially crucial if the assumption of constant volatility is foregone and stochastic volatility is utilised. The use of stochastic volatility is supported by the literature, with an early claim to that effect offered by Scott (1987), as discussed above.

Another such study is that of Figlewsky (1997). In his discussion of the Black-Scholes model, Figlewski points out that, of the five input parameters of the model, volatility is the only one that is not readily available. Moreover, the volatility that is expected to be experienced over the lifetime of the option has to be forecasted if such an option is to be priced (Figlewski, 1997). Figlewski further points out that, while using historical volatility values as these aforementioned forecasted volatility values is a possible and somewhat common approach when pricing options, it is not necessarily the most accurate of methods available, especially when considering the variability in measuring methodology (e.g., annualised monthly volatility versus annual volatility). Moreover, Figlewski also points out that using historical volatility values as forecast values is flawed as this assumes that the volatility experienced in the market is destined to remain at its historical levels. The author suggests that this fact, together with the absence of normality of returns in the market, urges the use of stochastic (or time-varying) volatility models, such as the autoregressive conditional heteroscedasticity (ARCH) family of models.

An early study invoking the use of a generalised ARCH (GARCH) model, specifically the GARCH(1,1)[5] model, is that by Heston and Nandi (2000). In their study, the authors provide a closed-form model to price options using the GARCH(1,1) model as a variance forecasting tool. The authors note that the results obtained from the discrete-time option pricing model utilising the single lag forecasting tool are sufficiently close to the results obtained using a continuous-time model, but the discrete-time process offers the advantage

[5] When specifying a GARCH(p, q) model, the variable p represents the time lag related to the variance while the variable q represents the order of the squared error terms of the relevant ARCH process.

of ease when used to forecast volatility by using historical prices for the underlying (Heston & Nandi, 2000). The authors conclude by stating that, due to the model having a closed-form solution, the model can be applied when coupled with the input variables' data.

Building on the work produced by Heston and Nandi (2000), Christoffersen and Jacobs (2004) examine the performance of various GARCH models in producing volatility forecasts for option pricing. While the authors do acknowledge that some departures from the traditional assumptions used in the GARCH model parameterisation process (e.g., the normality of returns) offer higher accuracy of forecasts, Christoffersen and Jacobs do suggest that the frequent updating of model parameters (including volatility using a GARCH-type model) yields good performance when such an option pricing model is employed (Christoffersen & Jacobs, 2004).

More recent studies (see, for example, Yavuz and Özdemir, 2018; and Fall, Ndiaye, and Sene, 2019) have focused on the application of fractional calculus to improve the predictive ability of the Black-Scholes model. While the mathematics of the avenues proposed by these and similar recent studies are fascinating, they are beyond the scope of this study.

The fundamental arguments presented in the literature suggest that, of the various input variables required for the Black-Scholes option pricing model, volatility is the only one that is to be forecasted for the period until the maturity of the option. Moreover, the literature suggests that constant volatility is inappropriate, especially if historical figures are to be used for the forecasting period. The applications of stochastic variance (or volatility) forecasting techniques, such as the use of the GARCH(1,1) model, are, therefore, best suited for option pricing in the context of the application of the Black-Scholes model. Hence, when pricing options, this study will employ such stochastic variance forecasting methods to price European call options when employing the Black-Scholes model, together with methods such as the moving average historical volatility measure.

Moreover, it cannot be ruled out that there are more efficient methods to forecast the variance needed to price options or the intricate relationships between the input variables and the option price output. As pointed out by Rubinstein (1985), the output obtained from an application of the Black-Scholes model is not necessarily accurate, mainly due to the various assumptions made and the parameterisation constructed by its developers. Hence, a non-parametric approach may be superior to learning the inherent relationships between the input and output variables. As is discussed in Chapter 3, an ANN is one such non-parameterised approach.

2.2. The Bachelier Model

The Bachelier model was introduced by Louis Bachelier in his 1900 doctoral thesis. In his thesis, Bachelier had remarkable breakthroughs of originality when developing and applying many mathematical approaches to financial markets independently of other works which were later produced with great rigour, such as stochastic analysis, martingales, Brownian motion, the Markov property, and option pricing, to name a few (Sullivan & Weithers, 1991).

While Bachelier wrote his thesis in French, a translation of his work is provided in *Louis Bachelier's Theory of Speculation* by Mark Davis and Alison Etheridge, published in 2006. In their book, Davis and Etheridge credit Bachelier with developing the field known today as option pricing theory, while acknowledging that Bachelier's contribution was not isolated from previous research in the field, even though a full reference list is not available for Bachelier's thesis.

The option pricing model developed by Bachelier (1900) shares similarities with the Black-Scholes model (see Section 2.1). For one, Bachelier's model was built upon the attempt to price an option by examining the underlying asset of such an option using a stochastic process (Davis & Etheridge, 2006), by modelling the underlying's price using the following stochastic model:

$$S_t^{Bachelier} := S_0 + \sigma^{Bachelier} W_t \tag{5}$$

where $S_t^{Bachelier}$ is the price of the underlying at time $t \in [0, T]$ as modelled by the Bachelier model, denoted simply by S_t in the remainder of this subsection; S_0 is the price of the underlying at time 0; $\sigma^{Bachelier}$ is the model's standard deviation, denoted simply by σ in the remainder of this subsection; and $(W_t)_{t\in[0,T]}$ is a random variable that follows a standard Brownian motion process at time t (Schachermayer & Teichmann, 2008).

Further, by applying this process (Equation (5)), Bachelier (1900) assumes that the price of the underlying asset continuously changes and deduces that an essential consequence of such, coupled with homogeneity in both time and space, is the normal distribution of the underlying asset's price when time is a fixed input (Davis & Etheridge, 2006). This assumption corresponds to assumption ii. above in the Black-Scholes model, where Black and Scholes (1973) assume that the price of the underlying asset (a share, in the authors' study) follows a log-normal distribution. Bachelier's model also attempts to price an option by examining the underlying asset's price, variance, and the price at which the option would be exercised (Sullivan & Weithers, 1991), the same inputs required by the Black-Scholes model.

Bachelier (1900) constructs an option pricing model based on the principle that the expectation of any payoff from the option in the view of the buyer is zero, the expectation being understood in its normal mathematical sense. While understanding that the market may display random movement around price S_t of the underlying, Bachelier defines the probability of the underlying's price being around S_t, i.e., $\pm S_t$, to be equal to π. Bachelier then defines the call premium to equal to C (equivalent to C_{S_t,t^*} in the Black-Scholes model) and the strike price to equal to X (equivalent to X in the Black-Scholes model), thereby stating that the true spread[6] of the option at the call premium C is, then, the quantity $X + C$ (i.e., $X + C_{S_t,t*}$).

Hence, the value of an option is then evaluated using the expectation assumption in three scenarios:

i. for a price $S_t \in (-\infty, X]$;
ii. for a price $S_t \in (X, X + C]$; and
iii. for a price $S_t \in (X + C, \infty)$.

In Case i, i.e., for a price $S_t \in (-\infty, X]$, the call option's underlying's price is now less than its strike price, X. Hence, the option will not be exercised. Therefore, the buyer of the option can expect to forfeit the premium paid for the option, that is, the loss incurred by the buyer of the call option is equal to C. Hence, the expectation of this case is equal to $-\int_{-\infty}^{X} \pi C \, dS_t$.

[6] The true spread is defined by Bachelier (1900) to be the difference between the option's price and the option's true price as measured on the option's call date.

In Case ii, i.e., for a price $S_t \in (X, X + C]$, the call option's underlying's price is now greater than its strike price, X, but still less than its true spread, $X + C$. This, too, suggests that the buyer of the call option will make a loss equal to the true spread less the price of the underlying, i.e., a loss equal to $X + C - S_t$. The expectation of this case corresponds to the expression $-\int_X^{X+C} \pi(X + C - S_t)\, dS_t$.

Lastly, in Case iii, i.e., for a price $S_t \in (X + C, \infty)$, the call option's underlying's price now exceeds its true spread, $X + C$, and is not bound from above. Hence, the buyer of the call option will now make a profit equal to the underlying's price less the true spread, i.e., a profit equal to $S_t - (X + C) = S_t - X - C$. The expectation of this final case corresponds to the expression $\int_{X+C}^{\infty} \pi(S_t - X - C)\, dS_t$.

By applying the principle of total expectation, Bachelier (1900) derives the relationship between the call option's premium, C, and the probabilities attached to the expected profit (or loss) attached to a call option whose underlying is priced at S_t with an exercise price X. The relationship is as follows.

$$\int_{X+C}^{\infty} \pi(S_t - X - C)\, dS_t - \int_X^{X+C} \pi(X + C - S_t)\, dS_t - \int_{-\infty}^{X} \pi C\, dS_t = 0 \tag{6}$$

By simplifying Equation (6), the following equation is obtained.

$$C + X \int_X^{\infty} \pi\, dS_t = \int_X^{\infty} \pi S_t\, dS_t \tag{7}$$

A further simplification can be applied to represent the premium of the call price as a function of the probabilities attached to the expected profit (or loss) of a call option with an underlying priced at S_t with a strike price X. This simplification is as follows.

$$\begin{aligned} C + X \int_X^{\infty} \pi\, dS_t &= \int_X^{\infty} \pi S_t\, dS_t \\ \Rightarrow C &= \int_X^{\infty} \pi S_t\, dS_t - \int_X^{\infty} \pi X\, dS_t \\ &= \int_X^{\infty} \pi S_t - \pi X\, dS_t \\ &= \int_X^{\infty} (S_t - X)\pi\, dS_t \end{aligned} \tag{8}$$

Hence, the call option's premium (i.e., the call option's price), C, is the expectation of the difference between the price of the underlying and the option's strike price.

Finally, Bachelier (1900) combines the conclusions above with stochastic calculus to conclude that the price of the underlying share follows the following stochastic differential equation.

$$dS_t = rS_t dt + \sigma dW_t \tag{9}$$

where S_t, r, σ, and W_t are as defined above.

When combining the differential equation depicted in Equation (9) with the payoff condition of a European call option (as depicted in Equation (3)), the only solution to the stochastic differential equation which satisfies the boundary condition and non-zero risk-free rates is then:

$$\begin{gathered} C(S_t, t^*) \coloneqq C = \left(S_t - Xe^{-rt^*}\right)\Phi(d) + \sigma\sqrt{\frac{1 - e^{-2rt^*}}{2r}}\,\phi(d) \\ d = \frac{S_t - Xe^{-rt^*}}{\sigma\sqrt{\frac{1 - e^{-2rt^*}}{2r}}} \end{gathered} \tag{10}$$

where all of the variables used correspond in meanings to those used in the Black-Scholes equation, as depicted in Equation (4); and $\phi(d)$ is the probability density function of the normal distribution at point d.

Notable differences between Bachelier's model in Equation (10) and the Black-Scholes model in Equation (4) include Bachelier's lack of incorporation of a minimum value which the underlying must take, i.e., the price of the underlying can fall below zero in Bachelier's model. Bachelier's work is, however, heavily based in mathematics and applied physics, rather than in economics (Davis & Etheridge, 2006), the latter field being the focus of Black and Scholes (1973). Hence, while the Bachelier model is not as complete as that of Black and Scholes, it serves as a remarkable advancement, pioneering the field of option pricing, especially by a mathematician with a non-economics focus, pre-dating the work of Black and Scholes by over 70 years.

Further highlights of Bachelier's non-economic background are found in his assumption that the expectation of price change is zero, i.e., the market price is correct, while any fluctuations around it are random in nature (Sullivan & Weithers, 1991). While one who is familiar with economic and financial theories may realise that this is not necessarily true, Sullivan and Weithers (1991) point out that this necessarily implies that the average investor in the market displays neutrality when it comes to risk, a characteristic that is both uncommon as an assumption in the literature as well as untrue in practice. This suggests that Bachelier (1900) assumes that the average market participant displays constant absolute risk aversion decades before such a notion was even developed.

Regardless of deviation from standard financial theory (which was not so standard at the time, but was developed about half a century later), Bachelier's work did exactly what it was set out to do. Cootner (1964) points out that the rational methodology applied by Bachelier (1900) has resulted in options priced using Bachelier's model serving as relatively accurate proxies to the prices of the option in the market.

While the literature may criticise Bachelier's understanding of economics, his training is, in fact, in mathematics. However, this does not mean that the literature has spared him from criticism in that field, either.

Davis and Etheridge (2006) not only present a translation of Bachelier's (1900) work, but also a collection of comments and footnotes, linking his work to later literature. In their analysis of his work, the authors point out the mistakes that Bachelier makes in several sections of his thesis, from writing incorrect bounds on integrals to the lack of rigour in his proofs. In fact, Davis and Etheridge point out that well-known mathematicians have incorporated Bachelier's ideas in their work, as a basis for their lack of rigour statement. One such author is Kolmogorov. In his 1931 paper, Kolmogorov voices the opinion that, while Bachelier's work is on the right track with regards to his conclusions, the method in which such conclusions are arrived at lack any rigour. In fact, Davis and Etheridge report that Bachelier's treatment of the differential equations solved using the payoff constraint inherent in a call option lacks the consideration for the uniqueness of the solution.

Even if Bachelier's applications lack rigour, his development of the arithmetic Brownian motion in his thesis is one of the various signs of intellectual contributions to the field of mathematics. Smith (1976) points out that Bachelier's application of an arithmetic Brownian motion process, as opposed to a geometric one, allows for the possibility of non-zero probabilities being attached to negative prices of the underlying asset, something which is viewed as a non-existent phenomenon. However, this exact property is the reason that the CME Group, Inc., has moved to use the Bachelier model as the pricing model for options on its exchanges after the negative price of oil futures for May experienced in late April 2020.

The literature covering the Bachelier model is scarce relative to the large volume of studies examining and discussing the Black-Scholes model. Below are a few studies that discuss the assumptions made by Bachelier, similar to the arguments made above for the Black-Scholes model. However, since the work presented by Bachelier (1900) was submitted in French, many authors use a translated form of his thesis as their basis, and a discussion of some of his careless mistakes and more general mathematical omissions and errors was provided above.

An early study examining the fundamental assumption of the Bachelier model is that of Murphy (1990). In his study, Murphy acknowledges that Bachelier's model's assumption of normality of share returns (as opposed to log-normality) implies that, in the Bachelier model, share price changes follow an arithmetic Brownian motion process as opposed to a geometric one (the latter eliminated the possibility of negative share prices). In response, Murphy proposes the use of a truncated normal distribution (at zero) for share price changes to account for the limited liability as experienced by companies whose shares are traded on exchanges (Murphy, 1990). While Murphy's insight into the Bachelier model is important when a literature review is compiled, it is important to note that the reason that the CME Group, Inc. adopted the Bachelier model following April 2020's events is precisely due to its ability to produce negative values. Hence, a truncated normal distribution is not directly applicable to this study.

Musiela and Rutkowski (2005) claim that the Bachelier model can, in fact, be a better tool for (short-lived) derivatives in some markets, particularly when considering short-lived options. This notion is then further supported by Schachermayer and Teichmann (2008), who note that, in developing his model, Bachelier (1900) considers the relatively popular two-month options, meaning that his model was developed for short-term predictions. Schachermayer and Teichmann also note that, due to this reason, the Bachelier model actually provides relatively good approximations to market prices and market volatilities (Schachermayer & Teichmann, 2008) (relative to actual prices and volatilities, as well as those achieved when employing the Black-Scholes model – the latter comparison is discussed in more detail in Section 4.1).

A relatively recent attempt to build on the Bachelier model is that presented in a study by Magdziarz, Orzeł, and Weron (2011). In their study, the authors attempt to build on Bachelier's (1900) observation of temporary stagnation of the financial process and its similarity to trapping events of sub-diffusive particles (Magdziraz, et al., 2011). The authors introduce a sub-diffusive process (the Fractional Fokker-Planck equation) which avails the authors of a range of distributions to use for the periods in which the price stagnates. The authors conclude that their proposed expanded model is arbitrage-free, but incomplete (meaning that the model is not defined by a unique martingale measure or, equivalently, that not all contingent claims can be replicated via a self-financing strategy) (Magdziraz, et al., 2011).

3. Artificial Neural Networks

An ANN is a specific application of machine learning. Hence, this chapter begins with an introduction of machine learning fundamentals before the complexities of ANNs are explored.

A basic overview of an ANN is provided before this chapter discusses the construction of an ANN and its uses, its relative advantages and disadvantages, together with the ANN's basic building blocks required for the construction of such structures. Concepts such as hidden layers and activation functions are also discussed, together with some examples of and discussions surrounding the latter.

3.1. Machine Learning Overview

Machine learning can be aptly defined as the field of auto-learning computer systems which are structured to self-improve their auto-learning processes from their own experiences (Mitchell, 2006). In essence, machine learning processes are algorithms that are designed to learn from their own experiences with the central theme of using such lessons to improve their own processes. When it comes to output-targeted processes, machine learning algorithms are tasked with enhancing their own predictive ability by learning relationships that exist in the data fed to them.

This makes machine learning an interesting tool to be applied to naturally complex concepts. These complex concepts may often be non-linear or non-parametric in nature. The ability of an algorithm to handle a non-linear problem in a non-parametric approach to solving a task is exceptionally useful when coupled with the task of auto-enhancement of the process by which the algorithm improves its own ability to solve the task at hand. Therefore, machine learning algorithms are well-suited to the task of pricing options – a task that is non-linear in nature – when employing a non-parametric approach.

As mentioned, machine learning algorithms rely on data fed into them in their process of self-learning. The data used is often separated into a training dataset and a validation dataset (Suthaharan, 2014). The former, the training dataset, is fed into the algorithm with both input and the desired output, allowing the algorithm to learn the implicit relationship that exists between the two variables. The latter, the validation dataset, is used by only letting the algorithm access the input set. The algorithm then runs its course, outputting an output set, which is compared to the output set of the validation dataset. The idea is to compare the algorithm's output to the actual output, and to judge the algorithm's predictive ability. Should the algorithm's predictive ability not be sufficient, further training may take place.

Machine learning algorithms can be further broken down into algorithms that employ supervised learning versus those that employ unsupervised learning. Supervised learning algorithms are supplied both input and output data in the training process, allowing the algorithm to learn any implicit relationships between the two sets, as described above (Barber, 2012). Unsupervised learning algorithms, on the other hand, provide the machine learning algorithm with the input data alone, allowing the algorithm to group the input data according to the attributes it can learn from the data (Barber, 2012).

Lastly, supervised learning can be further broken down into classification and regression tasks. If the machine learning algorithm's output is discrete in nature and, therefore, can be grouped into clusters (say, sections of the number line for numeric outputs), then the algorithm deals with a classification task (Barber, 2012). Conversely, if the machine learning algorithm's output is continuous in nature, then the algorithm deals with a regression task (Barber, 2012).

3.2. Artificial Neural Networks Construction

The ANN is a supervised learning application of machine learning. An ANN receives data through a network of neurons which, in turn, decodes the implicit relationship between the input data and output data of the dataset used. The ANN then uses further input data to forecast output data based on the inherent relationships learned.

Mathematically, an ANN is classified as a directed graph. The graph's nodes are analogous to the brain's neurons. These neurons are, in turn, connected to each other, often in layers, by means of the graph's edges, following the directionality of the network. This directionality begins with the input layer and ends with the output layer (Jain & Mao, 1996). In between, there may be any number of hidden layers. Any ANN which includes more than one hidden layer is said to be a deep ANN.

As a technical classification example, as presented in Basheer and Hajmeer (2000), consider the following. In a simple classification problem, an ANN is used to solve a classification task. The system is fed information by making use of an input vector, $\vec{x}$. This input vector $\vec{x}$ is then used to calculate the net input vector, $\vec{\xi}$, by calculating the dot product of the input values $\vec{x}$ and their respective weights, $\vec{w}$, i.e., $\vec{\xi} = \sum_{i=1}^{n} w_i \cdot x_i$ for an input vector $\vec{x}$ of length n.

For this net input vector $\vec{\xi}$ to pass from the first neuron and serve as the network's output, its value must exceed the current neuron's bias level (also known as the threshold limit), b. Since this is a classification ANN, this is done by assigning the Boolean variable y (the output variable) the value of either 0 or 1, where 0 implies that the neuron is not 'activated' due to the net input vector $\vec{\xi}$ not exceeding the bias level, b, and 1 implies that the neuron is activated due to the net input vector $\vec{\xi}$ exceeding the bias level, b (Basheer & Hajmeer, 2000).

The ANN is an attractive structure to apply to complex problems due to the structure's learning and generalisation capabilities coupled with its adaptability and noise insensitivity, amongst other desirable qualities (Jain & Mao, 1996). These qualities are desirable since they accommodate the non-linearity of data, increased prediction accuracy due to noise tolerance, and the ability to update to changing informational circumstances and new, unseen, data points. Moreover, the ANN is well-suited to handle tasks whose solutions may be achieved regardless of the degree of precision available in the formulation of the task (Basheer & Hajmeer, 2000).

On the other hand, a serious downfall of ANNs is their need for large datasets (Hutchinson, et al., 1994). This may pose a problem when utilising an ANN to value options if said options are either thinly traded or, alternatively, their available history is insufficient to allow for correct pricing. Moreover, there exists the potential issue of new options which display pricing history in the market which is abnormal, leading to future predictions based on the ANN's predictions being incorrect. However, in both cases, the ability of an ANN to be re-trained will allow such distortions to be corrected as more data become available.

The versatility of ANNs, specifically multilayer perceptron (MLP) networks (see Section 3.2.2 below), has been shown to be conclusive in the results of Hornik, Stinchcombe, and White (1989) and Hornik (1991). Since then, ANNs have been applied to various applications throughout the literature, from stock trading (Vanstone & Finnie, 2010) and optimising solar power generation (Marquez, et al., 2013) to nuclear engineering (Cong, et al., 2013) and food science (Huang, et al., 2007). The variety of applications in the different fields testifies to the versatility of the ANN as a tool to solve complex problems. Moreover, although the applications in the various fields are vast, the field of finance, in general, and option pricing, in particular, is not excluded from the vast application of ANNs, as discussed further in Section 4.2.

The remainder of this chapter will discuss the key elements of the construction of an ANN. While these details are important, it is worthwhile to remember that the construction of the ANN is not the primary focus of this study. The primary focus of this study is the comparison of the predictive ability of the Bachelier model relative to the already-established usefulness of both ANNs and the Black-Scholes model, as discussed in the literature (see Chapter 4).

3.2.1. The Perceptron

In their most basic form, ANNs can be used for classification tasks, as discussed earlier. This will entail constructing an ANN whose output is a Boolean variable y, as presented in Basheer and Hajmeer (2000). This describes the 'network'[7] called the perceptron, which makes use of the following function.

$$y(\vec{x}) = \begin{cases} 1, & \vec{w} \cdot \vec{x} + b > 0 \\ 0, & otherwise \end{cases} \tag{11}$$

where $\vec{w}$, $\vec{x}$, and b are as defined earlier; $\vec{w} \cdot \vec{x}$ is the dot product of the two vectors; and $y = y(\vec{x})$.

3.2.2. The Multilayer Perceptron

When layering several neurons together, the MLP, also known as a feed-forward neural network, is formed, as presented in Fukushima (1980). Each of the layers receives an input from the previous layer, and passes on its output to the next layer, thereby forming a network of neurons. Notice that the directionality of the network is fixed (mathematically, due to its directed graph construction, as described earlier), hence the name 'feed-forward'.

The name also suggests the existence of at least two layers. While two-layer (input layer and output layer) ANNs are useful, ANNs are also constructed using several hidden layers,

[7] The perceptron, strictly speaking, is not a network as it is constructed using a single neuron, activated by the Heaviside function. For more detail, see Rosenblatt (1958).

i.e., the layers whose placement is between the input layer and the output layer. As mentioned earlier, an ANN is known as 'deep' if it includes more than one hidden layer. Each layer in the network can have any number of neurons[8].

While the activation function of the perceptron accommodates for a binary output (see Equation (11)), MLPs make use of activation functions that allow them to be used in regression applications, i.e., applications which require output from a continuous input set. The input vector $\vec{x}$ is again used as an input to the activation function, which may be a logistic sigmoid function, as depicted in Equation (12); a hyperbolic tangent function, as depicted in Equation (13); or some rectifier linear unit (ReLU) function, a basic example of which is shown in Equation (14), below. A specific derivative of the ReLU function that is of interest is the Softplus function, depicted in Equation (15) below[9]. Finally, Equation (16) depicts the exponential linear unit (ELU), which is also of interest in this study. The use of any of these activation functions results in each neuron taking on a real number as opposed to simply a Boolean output.

$$f(\vec{x}) = \frac{e^{\vec{x}}}{1 + e^{\vec{x}}} = \frac{1}{1 + e^{-\vec{x}}} \tag{12}$$

$$g(\vec{x}) = \frac{2}{1 + e^{-2\vec{x}}} - 1 \tag{13}$$

$$h(\vec{x}) = \max(0, \vec{x}) \tag{14}$$

$$j(\vec{x}) = \ln\left(1 + e^{\vec{x}}\right) \tag{15}$$

$$k(\vec{x}) = \begin{cases} \vec{x}, & x > 0 \\ \alpha\left(e^{\vec{x}} - 1\right), & x \leq 0 \end{cases} \tag{16}$$

[8] There is ample literature that discusses the construction of ANNs to aid the constructor of such networks in deciding both the number of layers and the number of neurons in each layer. See, for example, Rivals and Personnaz (2003).

[9] While there are other activation functions that may be used in the context of a MLP, such as the Leaky ReLU (LReLU) and Parameterised ReLUs (PReLUs), it is worth noting that the activation functions listed are those often used when the output of the ANN is required to be positive, as is the case in this study.

where $\vec{x}$ is the input vector as described above and $\ln(v)$ is the natural logarithm of the input value v; and the ELU parameter α dictates the inclusion rate of the network when it encounters negative input values.

The logistic sigmoid activation function is one of the most commonly used activation functions in the ANN literature (Sibi, et al., 2013), especially in the literature dealing with pricing options using ANNs (see, for example, Hutchinson, et al., 1994; and Garcia and Gençay, 2000). The logistic sigmoid function's output is bound between 0 and 1, i.e., $f(\vec{x}) \in (0,1)$. Therefore, it performs best when the input data is also bound within this output range (Sibi, et al., 2013), although the function's domain is technically unrestricted.

Both the logistic sigmoid and the hyperbolic tangent functions result in ANNs which are both well-behaved and efficient (Baughman & Liu, 2014). This means that these functions lend themselves to solve modelling tasks of increasing complexity.

The ReLU function (and, consequently, its derivative functions, such as the Softplus function) has increased in popularity in recent years due to its more efficient learning process when compared to the more traditional functions discussed above (Hara, et al., 2015). Another advantage of the use of ReLU as the activation function is its ability to train ANNs with up to 1,000 hidden layers (He, et al., 2016). An obvious disadvantage of the use of the ReLU (and, indeed, the Softplus) activation function is that the function's output is bound from below by 0. This means that, in tasks where a negative output is possible, the ReLU function will fail to train the network adequately.

Considering the last statement, it is important to take a step back and consider if such a function is appropriate for the task at hand. After all, the recent introduction of the use of the Bachelier model in practice was the result of the prices of futures on the WTI plummeting to values below zero. Hence, the question of whether the ReLU (and the Softplus) function is appropriate is crucial. Since this study is examining options, and the predictive ability of various parametric and non-parametric models to value such options, it is straightforward to conclude that an option whose value is negative will be forfeited by the holder of such an option and let it expire worthless. Hence, even though the Bachelier model allows for the input of negative prices for the underlying, the resulting values of options are still bound from below by zero. Therefore, the ReLU activation function and its derivatives are appropriate for this study.

Finally, the ELU function, as presented by Clevert, Unterthiner, and Hochreiter (2016), is similar to the ReLU activation function in the sense that is has a minimum value calculated for negative input values and it makes use of the input value as the output value for positive input values. Unlike the ReLU activation function, its value assigned to those negative input values is transformed exponentially as opposed to assigned to be zero. The use of the exponential function allows the ELU activation function to avoid neuronal death[10] while the use of positive input values allows the ELU activation function to avoid any possibility of gradient disappearance, just as is the case for the ReLU activation function (Qiumei, et al., 2019).

3.2.3. Backpropagation

The backpropagation algorithm, as introduced by Rumelhart, Hinton, and Williams (1986), is a commonly-used algorithm employed to train ANNs. This training (or the 'learning' of the network, as sometimes referred to in the literature) involves the re-balancing of weights assigned to the various parameters in the network to minimise the overall prediction error of the network (Baruníkova & Baruník, 2011). The basic notions of the backpropagation algorithm, in general terms, are as follows.

The input layer takes, as input, the data fed to the network, together with a bias term. These are processed in the input layer and passed on to the next layer (a process known as the forward pass), being either the first hidden layer or the output layer. If the layer is a hidden layer, a similar process is repeated to that which took place in the previous layer, where the data passed forward by the previous layer are processed together with a bias term. In each step of the algorithm, i.e., at each layer of the ANN, the parameters being processed by the network are updated with the current parameter values calculated via the activation function and the bias of the layer in question. This process repeats until the process concludes at the output layer, concluding with some kind of steady-state value(s). If the output layer consists of a single output neuron, the processes completed by the MLP ANN is methodically comparable to non-linear regression (Stark, 2017).

As mentioned, the backpropagation algorithm also calculates the output error, being the difference between the value predicted by the network and the true value. The aim of the

[10] This follows a gradient update that is too large and results in the activation function no longer activating the neuron in question.

algorithm, coupled with an appropriate activation function, is to minimise this output error parameter (Baruníkova & Baruník, 2011).

4. Literature Review

This chapter begins with a review of the similarities and differences of the two parametric models discussed in this study, namely the Black-Scholes model and the Bachelier model, as discussed in the literature. While much of the information discussed below may have already been discussed in other parts of this study for each model in isolation, specifically in Chapter 2, this chapter aims to compare and contrast the two models against one another, and repetition of some information may be inevitable. However, the comparison of the two models will highlight their (dis)similarities.

Later in this chapter, a review of the literature surrounding the predictive ability of the parameterised models (again, the Black-Scholes and Bachelier models) relative to that of ANNs is also provided. While the comparison of the Black-Scholes model's output is often compared in the literature to that of an ANN, the same cannot be said of a comparison of the output of the Bachelier model to that of an ANN, mainly due to its lack of commonplace application (relative to the Black-Scholes model). Hence, the literature surrounding the latter is relatively scarce.

4.1. The Bachelier and Black-Scholes Models

To start with the obvious, the Bachelier and the Black-Scholes models are both parameterised stochastic option pricing models, each of which has a closed-form solution. Both models make assumptions surrounding their input variables, and the completeness of such collection of input variables has been contemplated in the literature (see, for example, Long, 1974). Both models also disregard the investor's risk appetite (the Black-Scholes model excludes this by working with risk-neutral probabilities, see Smith (1976); the Bachelier model was developed with this notion assumed implicitly decades before the notion of risk appetite was developed). The models, evidently, have many similarities. It is, however, their dissimilarities that are of interest; the most notable of which resulted in the CME Group Inc. changing from the use of the Black-Scholes model to the use of the Bachelier model in 2020 (CME Group, Inc., 2020).

A notable difference between the two models is that of the fundamental principles underlying the models. The Black-Scholes model is based on the no-arbitrage principle, which is based on the principles of hedging and portfolio/position replication, while the Bachelier model is an equilibrium model (Schachermayer & Teichmann, 2008).

Another noteworthy difference is that the notions of a ‘risk-free’ rate of return and ‘risk-neutral’ probabilities are absent from the Bachelier model, while they are vital to the Black-Scholes model. When the Bachelier model was developed, bourse de Paris’s settlement rules were that settlement of options contracts (together with all payments attached to entering such option contracts) were payable at the option’s maturity. This means that, when developing his model, Bachelier (1900) uses forward price (Schachermayer & Teichmann, 2008). The need to discount payments at some ‘risk-free’ rate of return and attach some probability to this value (to obtain an expected present value) was unnecessary. These 1900 trading rules (and the resultant implication on the option pricing model of that time) are in clear contrast to today’s trading rules, whereby an option premium is paid upfront by the long party to an option contract (to minimise counterparty/credit risk). This modern notion of receiving payment prior to option maturity is heightened for exchange-traded options, where a mark-to-market system will be enforced by the exchange on a daily basis to further reduce its exposure, as the exchange takes on the role of the counterparty to each original option trade party via the process of novation.

When looking at how the two models modelled the underlying’s prices, the natural filtrations of the stochastic variables in the two models are, in fact, equivalent (Musiela & Rutkowski, 2005). Moreover, Schachermayer and Teichmann (2008) point out that the models’ forms and primary differences can be compared to the forms and primary differences between calculating simple interest and compound interest, where the Bachelier model is comparable to the former and the Black-Scholes model is comparable to the latter. Schachermayer and Teichmann elaborate that, by looking at the models’ corresponding Itô stochastic differential equations, this conclusion is more easily apparent. Further, it is an appropriate conclusion when considering that Bachelier (1900) modelled the prices of options using an arithmetic Brownian motion process while Black and Scholes (1973) did so by using an exponential Brownian motion process. Furthermore, the Bachelier model assumes a normal distribution for the underlying’s returns, while the Black-Scholes model assumes a log-normal distribution for the same. All of these fundamental property differences do lend themselves to the appropriate analogy of interest rate calculations, as pointed out by Schachermayer and Teichmann.

Another similarity is that, when modelling the prices of the underlying (see Equations (1) and (5)), the two models do not yield a simple inversion of the model to calculate the implied volatility, i.e., it is not trivial to obtain a value for the volatility experienced if the underlying's price and the corresponding premium for an at-the-money option are possessed (Schachermayer & Teichmann, 2008).

Lastly, while the two models are remarkably different in their fundamentals (e.g., the distribution of the underlying's returns, see Murphy, 1990), they are remarkably close in their predictive output for short durations (Schachermayer & Teichmann, 2008). This conclusion is especially useful in practice as options' maturities are frequently short-term (e.g., three months), with longer-term positions often consisting of rolling over the position (i.e., stringing together multiple three-month-long options back-to-back to achieve a longer-term position). The primary difference which resulted in the Bachelier model being taken on recently by key market participants (e.g., the CME Group, Inc.) is its ability to incorporate negative values of the underlying for the derivatives in question. This, coupled with the Bachelier model's outputs' proximity to those of the Black-Scholes model in short term options, makes the model an appropriate tool to price options, without significant downsides. Moreover, the Bachelier model may, in fact, be a more natural approach from a statistical standpoint (Schachermayer & Teichmann, 2008). However, the Bachelier model is still a parameterised model, and its predictive accuracy should be measured relative to a non-parameterised approach when pricing options.

4.2. Artificial Neural Networks and Options Pricing

The primary difference between an ANN used to price options and either of the above models is that the ANN is referred to as a non-parametric approach to pricing options. This means that the ANN learns the relationships between the input variables and the network's output (the option's price) without assigning an explicit formula to describe relationships, i.e., it does not explicitly parameterise the relationship (Hutchinson, et al., 1994).

In their study comparing the efficacy of using a non-parametric ANN to price both American and European options, Hutchinson, et al., (1994) use the same input variables as those used by Black and Scholes (1973) for their ANN, and define the option's price as the output variable. The ANN then maps the input variables and the desired output to learn those inherent relationships between the output and each of the inputs, together with the input variables' interactions with each other (Hutchinson, et al., 1994). While Hutchinson, et al., do admit that (at the time) the ANN is not a substitute to the traditional Black-

Scholes parametric model, they assert that the use of an ANN allows for a non-constrained methodology to determine the price of options, one that is not negatively affected by any specification and model errors inherent in the parametric model's approach. The authors take this statement a step further by stating that ANNs may be a flexible alternative when it comes to dealing with the range of derivatives available on the market, and each of their specific price dynamics, to better price such derivatives non-parametrically (Hutchinson, et al., 1994). Hutchinson, et al., conclude, however, that the ANN method requires training, a process that is data-intensive and, therefore, is only appropriate to well-established option markets and those options that are widely traded within them.

An early study examining the various models proposed by the literature to price options is that of Bakshi, Cao, and Chen (1997). In their study, the authors open with listing the variations of the model proposed by Black and Scholes (1973) in the literature, and the overwhelming number of such variations. While the study does not consider ANNs, the authors propose a model that incorporates some stochastic variables (volatility, interest rates, and random jumps), and concludes that many of the variations offered in the literature are special cases of this proposed general model (Bakshi, et al., 1997). The study's relevance to the implementation of ANNs to price options may not be readily apparent until the usefulness of such abstraction employed is considered. The usefulness of the model presented by Bakshi, et al., is that in developing a more general model, the authors allowed for the model to be suitable in various cases (evident by the various models presented in the literature being special cases of such a model). Hence, it can be argued that an ANN, a model that is non-parameterised at all, may be the most general of such option pricing models and, therefore, is suitable for almost any application. This can be further supported by the logical notion that any proposed parametric model is a special case of a non-parametric model, as the latter learns those inherent relationships of the former.

In an attempt to model an ANN similarly to the Black-Scholes model, Garcia and Gençay (2000) develop a feed-forward ANN model. The authors separate the pricing function into two parts: one being a function of the ratio of the spot price to the strike of the option, while the other a function of the time to maturity (Garcia & Gençay, 2000). The authors find that the use of a feed-forward ANN produces delta-hedging errors which are superior (i.e., lower) relative to those incurred when implementing the Black-Scholes model (Garcia & Gençay, 2000), concluding that an ANN model is superior to the Black-

Scholes model when implementing hedging strategies using options. This is similar to the conclusion provided in Hutchinson, et al., (1994). The notion of basing the ANN's general structure on that of the Black-Scholes model is further supported by Dougas, Bengio, Bélisle, Nadeau, and Garcia (2001), where the authors conclude that such modelling aids in reducing the complexity of the learning algorithms employed to learn the inherent relationships between the variables at the construction stage of the model. This, in turn, improves the efficiency of the algorithm's performance (Dougas, et al., 2001).

The importance of the studies presented by Garcia and Gençay (2000) and Dougas, et al., (2001) is that these studies incorporate the economic axioms upon which the Black-Scholes model is based into their respective models. Intriguingly, these studies conclude that the ANNs perform better when the economic axioms are accounted for. This suggests that it is not always the most general of models that yields the most accurate and efficient results – the slightly less general ANN (one that employs the economic axioms in its construction) produces superior performance relative to the more general ANN (one that does not employ the economic axioms in its construction) when pricing options.

On the other hand, several studies have alluded to the superior performance of the Black-Scholes model relative to that of an ANN in certain cases. Two early studies are, too, from the same decade (1990s to 2000), namely the study by Malliaris and Salchenberger (1993) and that by Geigle (2000).

In their study, Malliaris and Salchenberger (1993) state that the outperformance of one model over another seems to depend on the moneyness of the option examined, with the best performance of both models exhibited for at-the-money options (Malliaris & Salchenberger, 1993). The Black-Scholes model seems to over-price deep-out-of-the-money options, while both models under-price in-the-money options, with the Black-Scholes model being slightly more accurate (Malliaris & Salchenberger, 1993). This is further supported by Geigle (2000), who finds that the Black-Scholes model is the most biased in its predictions when dealing with in-the-money and out-of-the-money options.

While neither of these two studies outright claims that the Black-Scholes model is ultimately the superior model, they do find that, under certain conditions (e.g., the moneyness of the option), the Black-Scholes model's performance is superior to that of the ANN. Moreover, both studies find that, again, in certain cases, the performances of the two models are comparable. However, it is important to note that these early applications of ANNs to price options did not base their ANNs on the Black-Scholes model and its accompanying economic axioms (which, as discussed above, would have improved the ANN's performance).

Looking at the performance of ANNs when pricing European options in the United Kingdom (U.K.), Bennell and Sutcliffe (2004) compare said performance to that of the Black-Scholes model using FTSE 100 options. While the authors do generally conform with the literature regarding the superiority of the ANN relative to the Black-Scholes model when pricing options, the results of the study contradict the findings of Malliaris and Salchenberger (1993) and Geigle (2000), finding that the ANN model performs poorly relative to the Black-Scholes model when ITM options or those with terms to maturity in excess of 200 days are considered (Bennell & Sutcliffe, 2004).

While the studies presented by Bakshi, et al., (1997), Garcia and Gençay (2000), and Dougas, et al., (2001) present findings based on the U.S. data, a study by Lajbcygier (2002) using Australian data concludes that an ANN not only performs on par with the parameterised models considered (which include the Black-Scholes model), but demonstrates superior performance for short expiry near-the-money options, which the author considers to be the most important set of options in their study. Lajbcygier further concludes that the ANN model should be utilised in parallel with the Black-Scholes model to enhance the latter's performance (which is already reasonably accurate in the Australian market) (Lajbcygier, 2002). Moreover, Lajbcygier suggests that if the ANN was to be trained based on the errors made by the Black-Scholes model, the potential performance of the ANN might be even more accurate.

Still examining the Australian market and building on the study presented by Lajbcygier (2002), Hahn (2013) uses a GARCH(1,1) model to incorporate varying volatility levels in the Australian market when pricing options using ANNs, focusing on the change in dynamics brought upon the market by the 2008 global financial crisis. Hahn finds that non-parametric models such as the ANN do offer superior results when using stochastic methods to estimate volatility, specifically relative to more traditional models such as the Black-Scholes model (Hahn, 2013).

Using U.S. data once more, but now applying a new technique to the construction of ANNs, Gradojevic, Gençay, and Kukolj (2009) use S&P 500 data from 1987 to 1993 to price European call options using a modular neural network (MNN), with the modules of the network based on the moneyness of the option as well as its term to maturity. The authors find that a MNN based on moneyness modules offers superior performance when compared to both a feed-forward ANN and the Black-Scholes model. On the other hand, the MNN based on time to maturity modules performs poorly (Gradojevic, et al., 2009). Gradojevic, et al., conclude, however, that three-module MNNs are less accurate in their predictions when compared to the Black-Scholes model for the first year of their study (1987). The contribution of the study is, however, that the modularity of the ANN (i.e., allowing different ANNs to specialise in different levels of moneyness of the options as opposed to utilising a single ANN to attempt modelling all possible moneyness positions) improves the efficiency and predictive ability of the network.

The implementation of a MNN was also examined in the study conducted by Baruníkova and Baruník (2011). The study makes use of daily S&P 500 data from 2004 to 2007 and it, too, makes use of moneyness and time to maturity modules to price European call and put options. The MNN's results are compared to those achieved by an application of the Black-Scholes model and the study finds that the MNN resulted in lower pricing errors relative to the parameterised model, re-affirming the findings presented by Gradojevic, et al., (2009). Moreover, Baruníkova and Baruník find that the Black-Scholes model's bias increases proportionally to both time to maturity and the moneyness of the option in question. On the other hand, the authors claim that the MNN employed corrects for these biases (Baruníkova & Baruník, 2011).

A recent study by Yang, Zheng, and Hospedales (2017) builds on the study presented by Gradojevic, et al., (2009) by allowing for modularity to be automatically applied based on market conditions by implementing the MNN using gated neural networks. The authors find that the increased automation and generalisation of the MNN allows for better performance of the MNN when pricing European call options (Yang, et al., 2017).

Regardless of differences, the brain's ability to process information received in varying modes and interpret the information in a useful manner, often resulting in some output, lends its processes and inner workings to the non-trivial process of option pricing. The ANN's ability to mimic the flexible paths of information processing of the brain in a structured function serves itself to the task of option pricing sensibly. The blurriness in which the ANN may predict the price of an option mirrors how the brain processes and interprets the information that it absorbs from its surroundings. The literature demonstrates that, when comparing the accuracy of the forecasts made by ANNs and the Black-Scholes model, the ANNs are generally superior, especially if fed the economic axioms on which the Black-Scholes model is based. While useful, the modularity of the network is beyond the scope of this study, as the vast majority of the literature ultimately uses an ANN to price option as opposed to incorporating modularity, and this study aims to focus on the comparison of the Black-Scholes and ANN models with that of the Bachelier model, as opposed to focusing on the ANN employed to achieve such a task.

A recent study building on the study of Baruníkova and Baruník (2011) is that of Vach (2019). In his study, the author investigates various structures of ANNs to find the optimal structure to such an ANN when used to price European options whose underlying is the S&P 500 Index. Vach uses daily S&P 500 Index closing values from June 2004 to June 2007, the same dataset used by Baruníkova and Baruník. The author's study looks at various maturities (ranging from 1 day to 3 years) and examines both European put and call options.

Vach (2019) examines several structures as candidates for the optimal structure of an option pricing ANN. The author incorporates various combinations of numbers of hidden layers, sizes of each layer, activation functions, the use of modularity, as well as the use of virtual options[11]. Vach also examines several volatility estimation methods to be used as input into the Black-Scholes model, including a GARCH(1,1) model and moving averages of historical volatility at various lengths (ranging from 5 days to 100 days). Finally, Vach uses the yield on monthly U.S. Treasury bills as a proxy to the risk-free rate and the average of the best bid and the best ask prices on the S&P 500 Index for any given day as the underlying input into the Black-Scholes model.

Overall, Vach (2019) concludes that the optimal structure of an ANN tasked with pricing European options based on the S&P 500 Index is a deep ANN with two hidden layers and 30 neurons per hidden layer and one output neuron in the output layer. The activation functions utilised by the ANN depends on the layer in question, with the sigmoid function employed in the hidden layers and the Softplus function employed in the output layer. Vach concludes that a non-zero dropout rate and modularity do not add to the predictive ability of the ANN when pricing options.

[11] For more detail, see Vach (2019).

5. Data and Methodology

This chapter provides information surrounding the data used in this study. Moreover, this chapter also details the methodology employed to prepare the data for the analysis discussed in Chapter 6, primarily focusing on the construction and development of the ANN used to price options using the data (see Chapter 3).

This study follows several others, specifically that of Baruníkova and Baruník (2011), by using the S&P 500 Index as a proxy to the U.S. market, and evaluating European call options based on the Index. The dataset used in this study consists of the daily closing prices of the S&P 500 Index over the period 1 March 2005 to 31 December 2019, as obtained from the Bloomberg database. Hence, the period in question covers almost 15 years (3,671 trading days[12]), a period which is believed to cover approximately 1.5 business cycles of the U.S. economy. This period covers 230,571 options in the dataset. While it is not necessary to cover such an extensive period (there are studies that use much shorter periods, considering the frequency of data collected), it is believed that a longer period will allow for better calibration of the ANN. Moreover, the period will further need to be split into training, validation, and testing sets (see below for details).

Besides for the daily closing prices of the S&P 500, other data collected from the Bloomberg database include daily rates for the 6-month London Inter-bank Offer Rate (LIBOR) to be used as the risk-free rate (as commonly used in practice when valuing options whose underlying is the S&P 500 Index), and maturities, prices, and strikes for European call options based on the S&P 500 Index for various values of moneyness[13] (ranging from 80% to 120%). Maturities for the data range from one month to two years.

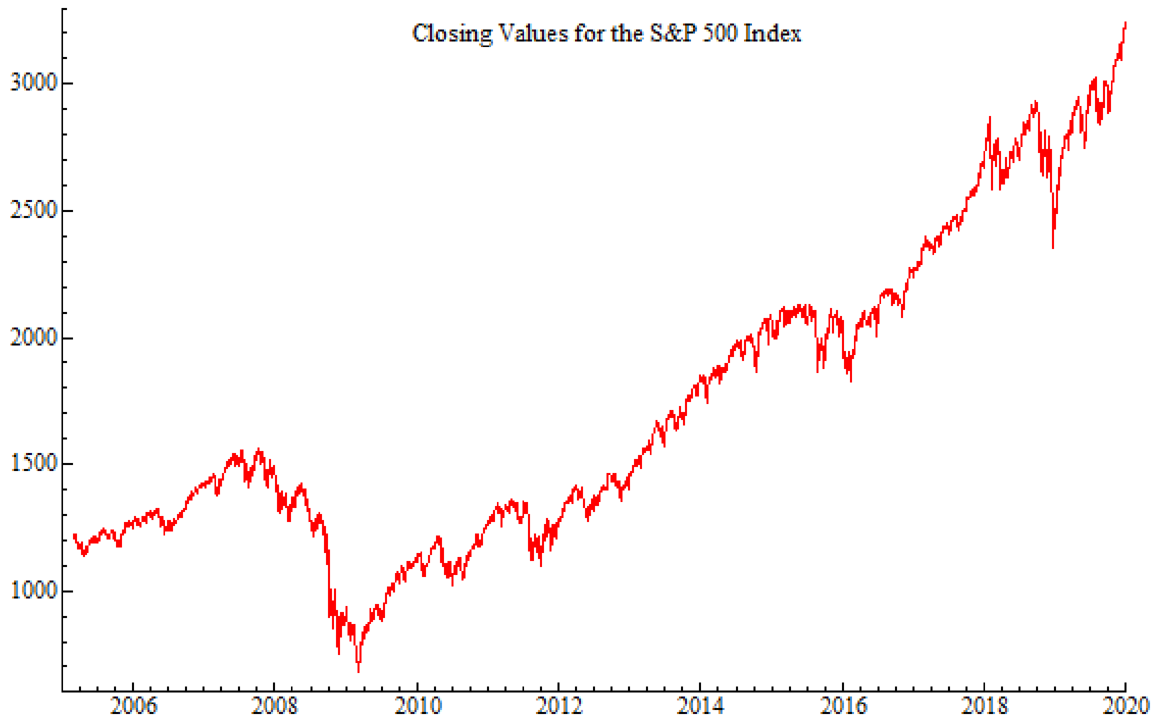
Following Baruníkova and Baruník (2011), several exclusions to the data were considered. In their study, the authors applied filters to their database. To avoid the liquidity biases identified by Bakshi, et al., (1997), the authors excluded options with less than six days to maturity. However, the minimum term to maturity considered in this study is one month, rendering this filter redundant. Next, the authors exclude any options whose price quote is lower than $0.375 to remove any potential adverse effects of the discreteness

[12] The dates used (and, therefore, the number of trading days examined) were chosen such that the time series of 6-month LIBOR and S&P 500 Index closing values overlapped, i.e., dates for which both values were available.

[13] An option’s moneyness is defined as the ratio of the price of the underlying to the strike price of the option.

of an option's price on the valuation. This condition is examined in this study for option's price based on both that based on the Black-Scholes model and that based on the Bachelier model. This filter excludes 4,401 options, leaving 226,170 options. Lastly, Baruníkova and Baruník exclude any call option whose call price (as calculated by the model(s) used) exceeds the maximum of zero and the difference between the price of the underlying and the option's strike, i.e., any option which does not satisfy the no-arbitrage argument with respect to the payoff function of a call option (as stated in Equation (3)). This, too, leads to no exclusion of options in the data of this study based on both models. Hence, the total number of options left in the dataset is 226,170.

Figure 1: Daily closing values of the S&P 500 Index



Note: This figure graphically depicts the daily closing values of the S&P 500 Index for the period 1 March 2005 to 31 December 2019. The data used were obtained from the Bloomberg database.

This study's primary aim is to introduce the Bachelier model to the body of literature that discusses the relative predictive ability of parameterised models when compared to non-parameterised models. As seen in Chapter 3, the literature discussing the relative merits of each method, with specific reference to the Black-Scholes model and ANNs, is vast. Hence, this study focuses on the comparison of the predictive ability of the three models in question, and will not focus on the different ways in which the ANNs used to price options can be structured to optimise their performances.

This study builds on that presented by Vach (2019), which establishes the optimal ANN structure to value European options whose underlying is the S&P 500 Index. To that effect, the ANN employed in this study is a deep ANN with seven inputs (equal to the number of parameters required by the parameterised model, together with the volatility inputs outlined below[14]). The ANN is constructed of two hidden layers with 30 neurons each. Vach has chosen the activation function of hidden layers arbitrarily to be the sigmoid function (see Equation (12)). However, this study employs the ELU activation function (see Equation (16)), as the use of the sigmoid function resulted in a dead gradient[15] while training the data. The hidden layers do not make use of a dropout[16]. The output layer is constructed of a single output neuron, utilising the Softplus activation function (see Equation (15)) as it ensures a positive output (as is the case for option prices in general), as used by Vach. The ratio of the data used as the training set, validation set, and test set is approximately 60:20:20, also as used by Vach.

Vach (2019) follows Baruníkova and Baruník (2011) and Bakshi, et al., (1997) once again by scaling the dataset's strikes by the options' exercise prices to obtain options which are described by their moneyness, and then sub-divides the dataset into nine sub-datasets based on moneyness and time to maturity. While this study does adopt the convention of scaling strikes by exercise prices, it does not sub-divide the network into nine different ANNs but, rather, it uses a single ANN to price across all moneyness and maturities. However, this study does sub-divide the database based on moneyness when applying the parameterised models in order to achieve further granularity for the volatility measures described further below. This results in a classification method for call options, whereby a call option can be classified as in-the-money (ITM) if its moneyness value exceeds 1.05 and out-of-the-money (OTM) if its moneyness is less than 0.97. A call option whose moneyness is in between these two values (inclusive) is then said to be at-the-money (ATM). It is worthwhile to mention that Baruníkova and Baruník used 1.03 as their

[14] The volatility measures detailed below are all used as inputs to the ANN simultaneously.

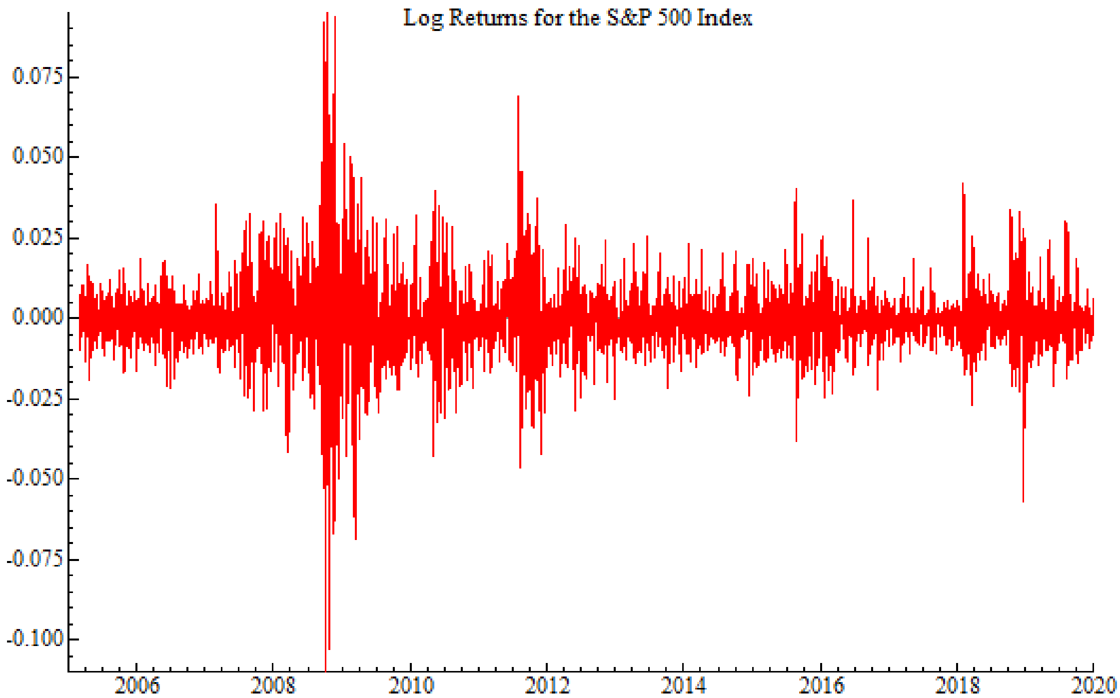
[15] A dead gradient refers to a situation whereby the learning of the network achieves some steady-state that is not in line with the outcome predicted due to the learning being ineffective.

[16] The dropout of an ANN is often employed to increase the accuracy of a backpropagation algorithm, where it represents the number of nodes dropped from each layer of an ANN (Vach, 2019).

boundary value to distinguish between ATM and ITM call options. Nonetheless, this study will adopt the more recent values proposed by Vach.

Another important variable that requires estimation is that of market volatility. This study also uses several volatility estimation methods to determine the volatility experienced in the markets. As is the case in Vach (2019), this study will use both a GARCH(1,1) model and several measures of historical volatility as a moving average (of lengths 5, 20, 60, and 100 days). Several studies find that a GARCH(1,1) model performs well when it is tasked with estimating volatility as an input to the Black-Scholes model (see, for example, Heston and Nandi, 2000), while historical volatility is not necessarily the best choice in such applications (Figlewski, 1997). However, Vach finds that certain moving average volatility measures outperform the GARCH(1,1) volatility measure when valuing options, for example, for put options. While this study does not focus on put options, but rather on European call options, it will nonetheless follow the approach of Vach by utilising both moving average and GARCH measures to calculate volatility. This will allow this study to offer more robust results surrounding the predictive accuracy of the models in question.

Figure 2: Daily log returns for the S&P 500 Index



Note: This figure graphically depicts the daily log returns of the S&P 500 Index for the period 1 March 2005 to 31 December 2019. The data used were obtained from the Bloomberg database.

In order to calculate volatility, returns earned on the underlying must first be calculated. Returns earned on the S&P 500 Index are calculated using the index's closing values as follows.

$$r_t = \ln\left(\frac{S_t}{S_{t-1}}\right) \tag{17}$$

where r_t is the daily return earned on the S&P 500 Index for day t; S_t is the level of the index at time t (denoted equivalently to the underlying at time t); and S_{t-1} is the previous day's level of the index. Therefore, Equation (17) represents a time series of log returns earned on the S&P 500 Index over the period examined in this study which, in turn, can be used to calculate volatility, a measure that is unobservable in the markets. These are depicted in Figure 2, above.

Some descriptive statistics of the log returns of the S&P 500 Index, r_t, the 6-month LIBOR rate, r_f, and the European call option prices on the S&P 500 Index, C, are captured in Table 1 below.

Table 1: Descriptive Statistics

Measure	***Number of Observations***	***Mean***	***Standard Deviation***	***Minimum***	***Maximum***
$\boldsymbol{r_t}$	*3,671*	*0.03%*	*1.18%*	*-9.47%*	*10.96%*
$\boldsymbol{r_f}$	*3,671*	*1.91%*	*1.69%*	*0.32%*	*5.64%*
$\boldsymbol{C}$	*226,170*	*\$161.65*	*\$119.40*	*\$0.54*	*\$759.10*

Note: This table depicts the descriptive statistics for each of the three time series collected for this study, namely the log return earned on the S&P 500 Index (denoted as r_t in the table), the 6-month London Inter-bank Offer Rate (LIBOR, denoted as r_f in the table), and call option premium (captured in U.S. dollars, denoted as C in the table). The data were collected from the Bloomberg database over the period 1 March 2005 to 31 December 2019. The number of observations for each line item was calculated by counting the number of dates available for each time series. The dates used were chosen such that the time series of 6-month LIBOR and S&P 500 Index closing values overlapped.

Turning back to volatility, the moving average volatility for a period of length τ is calculated as follows.

$$\sigma_{HV(\tau)} = \sqrt{\frac{1}{\tau - 1}\sum_{t=1}^{\tau}\left(r_t - \bar{r}_{(t.\tau)}\right)^2} \tag{18}$$

where $\sigma_{HV(\tau)}$ is the historical volatility for a historical period of length τ; r_t is the log return (as described in Equation (17)); and $\bar{r}_{(t,\tau)}$ is the average return over the preceding period of length τ up to and including time t. Hence, for example, $\sigma_{HV(100)}$ is the moving average historical volatility experienced over the preceding 100 days. The measure calculated in Equation (18) measures a τ-day daily volatility, which must be annualised. This is achieved by calculating the product of the square root of 252 and the τ-day daily volatility[17].

The GARCH(1,1) model applied to financial data, as presented in Engle (2001), is applied to estimate the variance of the error terms of the time series of returns of the underlying asset, stated as follows.

$$r_t = \bar{r}_t + \sqrt{\sigma_t}\varepsilon_t \tag{19}$$

where r_t is as described in Equation (17); $\bar{r}_t$ is the average return earned on the underlying up to and including time t; σ_t is the standard deviation of returns up to and including time t; and ε_t is the error term of the regression.

Hence, the GARCH(1,1) model calculates the standard deviation (i.e., volatility) of the return earned on the financial instrument in question at time $t + 1$ as follows.

$$\sigma_{t+1} = \sqrt{\omega + \alpha\sigma_t^2\varepsilon_t^2 + \beta\sigma_t^2} \tag{20}$$

where ω, α, and β are known as the GARCH model parameters and must be estimated from the time series data of the underlying's returns[18].

As can be seen in

Table *2*, the standard deviations of the moving average volatility measures decrease as the length of the period examined increases (the standard deviations decrease from 11.9% for the 5-day moving average volatility measure to 9.5% for the 100-day moving average volatility measure), as a longer period allows for further smoothing of the measure. Another thing to note is that the minimum volatility observed increases with the length of the period examined (the minima increase from 0.9% for the 5-day moving average volatility measure to 5.8% for the 100-day moving average volatility measure). The

[17] This is using the square-root-of-time rule, with 252 trading days per annum.

[18] There are some conditions to which the GARCH model parameters must adhere. For more detail, see Engle (2001).

converse is also true for the maximum volatility experienced (the maxima decrease from 120.0% for the 5-day moving average volatility measure to 61.2% for the 100-day moving average volatility measure). The averages experienced also increase with the rolling period's length. The GARCH-derived volatility measure was calculated using a rolling five-year calibration period with the underlying distribution being the normal distribution, and its statistics sit somewhere in between the values of the moving average measures, as is expected for such a conditional measure.

Table 2: Key Statistics of Volatility Measures

Measure	*Mean*	*Standard Deviation*	*Minimum*	*Maximum*
$\sigma_{HV(5)}$	*14.6%*	*11.9%*	*0.9%*	*120.0%*
$\sigma_{HV(20)}$	*15.3%*	*10.7%*	*3.3%*	*84.6%*
$\sigma_{HV(60)}$	*15.8%*	*9.9%*	*5.0%*	*73.8%*
$\sigma_{HV(100)}$	*16.1%*	*9.5%*	*5.8%*	*61.2%*
σ_{GARCH}	*15.8%*	*9.6%*	*7.0%*	*80.1%*

Note: This table depicts some key statistics of the various moving averages and GARCH volatilities calculated, namely the mean, standard deviation, minimum, and maximum. The annualised moving average volatility measures are calculated as sample standard deviations for the log returns of the S&P 500 Index over the period 1 March 2005 to 31 December 2019. The GARCH volatility measure is calculated using a GARCH(1,1) model and the log returns of the S&P 500 Index over the same period.

When it comes to determining which model is superior in its predictive ability, this study follows the methodology presented in Baruníkova and Baruník (2011), namely the use of the following metrics: The mean absolute error (MAE); the root mean square error (RMSE); and the mean absolute percentage error (MAPE). These measures are represented by Equations (21), (22), and (23), respectively. In general, the lower each of the three measures is, the better, as a lower value for each measure indicates a lower level of prediction error. This lower value indicates that either the absolute or squared difference (model-dependent) between the actual call option value and the model-predicted call option value are fairly close to one another for each datapoint. The MAE, the RMSE, and the MAPE measures are formulated as follows.

$$MAE = \frac{1}{N}\sum_{n=1}^{N}|y_n - \hat{y}_n| \tag{21}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2} \tag{22}$$

$$MAPE = \frac{1}{N}\sum_{n=1}^{N}\left|\frac{y_n - \hat{y}_n}{y_n}\right| \tag{23}$$

where N is the number of European call options in the dataset; y_n is the observed price of European call option number n; and $\hat{y}_n$ is the calculated price for that same option.

Since this study employs three measures of model prediction power and three models, it is possible that each measure will indicate that a different model is superior with respect to its predictive ability. Hence, following Baruníkova and Baruník (2011), the Diebold-Mariano statistic is used.

The Diebold-Mariano statistic, introduced by Francis Diebold and Roberto Mariano in 1995, is a statistic whose aim is to statistically compare the accuracy of forecasts made by different models. The statistic's null hypothesis is that the two models compared have equally accurate forecasts. Therefore, the Diebold-Mariano statistic aids in concluding which model produces superior forecasts by comparing two different models.

By considering the error of each forecast (defined by Dieblod and Mariano (1995) as the difference between the actual value and the predicted value), and then considering the difference of the squares of the error terms of different prediction models[19], the resulting difference is then termed the loss differential, denoted by Diebold and Mariano as d. The average loss differential for the two models' time series is then denoted by $\bar{d}$. The authors then define the autocovariance of the loss differential at lag k, denoted γ_k, as follows.

$$\gamma_k = \frac{1}{N}\sum_{i=k+1}^{n}\left(d_i - \bar{d}\right)\left(d_{i-k} - \bar{d}\right) \tag{24}$$

[19] It is also common to use the difference of the absolute values of the error terms of different prediction models.

where N is the size of the time series, equal to the number of European call options in the dataset in this study.

The Diebold-Mariano statistic is then formulated as follows.

$$DM = \frac{\bar{d}}{\sqrt{\widehat{Var}\left[\bar{d}\right]}} \tag{25}$$

where DM is the Diebold-Mariano statistic; and $\widehat{Var}\left[\bar{d}\right]$ is a consistent estimator of the sampling distribution of the average loss differential measure, derived using the autocovariance function in Equation (24) (Diebold & Mariano, 1995).

If the Diebold-Mariano statistic, DM, indicates that the null hypothesis is to be rejected, then the first model considered produces more accurate forecasts relative to the second model considered, where the rank of each model corresponds to the order of its respective error term in the stipulation of the loss differential measure. When applying the central limit theorem, the Diebold-Mariano test statistic is asymptotically distributed as a standard normal random variable.

It is important to note that the Diebold-Mariano statistic is often confused for a statistic that captures model superiority as opposed to one that captures the relative superior accuracy of forecasts (Diebold, 2015). Hence, while the Diebold-Mariano statistic may be useful in this study, its usefulness is limited to highlighting which models produce superior forecasts, as opposed to which models are inherently superior. A different calibration of a model such as the ANN model used in this study may likely result in a significantly different Diebold-Mariano statistic which, in turn, may lead to different conclusions surrounding the model's predictive ability. This notion will be revisited when the results of the models used are examined and discussed in the next chapter.

6. Results and Analysis

This chapter discusses the results obtained using the data and methodology discussed in Chapter 5, and provides the accompanying analyses of such results. More specifically, each of the parametric models introduced in Chapter 2 will be analysed for each of the volatility measures described in Chapter 5, and the superior volatility measure will be determined based on the forecast measurement techniques introduced in Chapter 5. In addition, the ANN, whose structure was also introduced in Chapter 5, will be discussed in terms of its predictive ability. Finally, the Diebold-Mariano test, as introduced in Chapter 5, will be applied to conclude which model offers superior accuracy in its forecasts among the three models considered in this study.

In this study, three models were considered when pricing European call options: Two parametric models and one non-parametric model. The two parametric models, namely the Bachelier and the Black-Scholes models, were each used to forecast the prices of European call options under five different volatility measures. The ANN model, on the other hand, used all of the volatility inputs simultaneously to learn the inherent relationships between the model's inputs and the desired output in its training set. Hence, overall, this study examined and tested eleven models; ten parametric models, and one non-parametric model.

The ten parametric models all used either a moving average volatility measure (of length 5, 20, 60, or 100 days) or a GARCH-based volatility measure, using the normal distribution as the underlying distribution and a five-year rolling calibration period.

Each model is discussed individually, beginning with the parametric models and ending with the non-parametric ANN model. The predictive ability and accuracy of each model are discussed with reference to the three error measures discussed in Chapter 5, namely the MAE, the RMSE, and the MAPE measures. This is particularly important for the parametric models as each of them uses five different volatility measures and, therefore, the three error measures will assist in concluding which of the volatility measures is a better overall estimate for implied volatility. Finally, this chapter concludes with a comparison of the best of each of the parametric models and the ANN when it comes to pricing European call options using the Diebold-Mariano test.

6.1. The Black-Scholes Model

This section analyses and discusses the results obtained using the data and methodology described in Chapter 5, as applied to the Black-Scholes model described in Section 2.1. The Black-Scholes model was used to calculate the prices of European call options over the period 1 March 2005 to 31 December 2019 over several maturities and moneyness levels. Five different volatility measures were used to price the options, and this section discusses which of the five measures yielded the most accurate forecasts, as measured by the forecasting error measures discussed in Chapter 5.

Table 3 below depicts the MAE, the RMSE, and the MAPE values for the Black-Scholes model for each of the five volatility measures used in this study across all moneyness levels. As can be seen, when evaluating which of the five volatility measures produces forecasts that are closest in values to the actual prices of European call options based on the S&P 500 Index, it is the 100-day moving average historical volatility measure that produces the most accurate forecasts across all three measures, as indicated by the lowest value for each of the three measures used. This may be a surprising conclusion based on the simplicity of the measure relative to the GARCH-based volatility measure.

This conclusion is in line with that of Vach (2019), who finds the GARCH-based volatility measure and the 100-day moving average historical volatility measure both perform well when no virtual options are included. It is important to note that the period examined in this study far exceeds that used by Vach. This may play a factor in the fact that the 100-day moving average historical volatility measure can be concluded as superior to the GARCH-based volatility measure.

Table 3: Forecasting Error Measures for the Black-Scholes Model for Various Volatility Estimators (All Moneyness Levels)

Measure	*MAE*	*RMSE*	*MAPE*
$\sigma_{HV(5)}$	*52.096*	*70.940*	*45.040%*
$\sigma_{HV(20)}$	*48.370*	*64.371*	*42.484%*
$\sigma_{HV(60)}$	*45.244*	*59.163*	*40.443%*
$\sigma_{HV(100)}$	*44.014*	*56.836*	*39.570%*
σ_{GARCH}	*45.775*	*59.660*	*40.475%*

Note: This table depicts three forecasting error measures, namely the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE) for the Black-Scholes model, as used to calculate the prices of European call options. The model uses the daily log returns of the S&P 500 Index over the period 1 March 2005 to 31

December 2019 to calculate four moving average historical daily volatility measures of length 5, 20, 60, and 100 days, as well as a daily GARCH volatility measure for the period. The total number of observations for the period was 226,170.

Interestingly, each of the forecasting error measures ranked each of the five volatility measures identically for the Black-Scholes model when moneyness is not used to break the dataset down further. The volatility measure that offers the next best forecasts is the 60-day moving average historical volatility measure, as it offers the second-lowest value for each of the three measures. The third-best measure is the GARCH-based volatility measure, followed by the 20-day moving average historical volatility measure and, finally, the 5-day moving average historical volatility measure.

While it is not strictly within the scope of this study, it is also possible to evaluate which of the five volatility measures produces superior forecasts when the data are split by moneyness levels. Tables 7, 8, and 9 in Appendix 1 depict the forecasting error measures' figures achieved by the Black-Scholes model when pricing European call options for ITM, ATM, and OTM options, respectively.

Similar to the analysis performed on all moneyness levels, the 100-day moving average historical volatility measure is the best performing volatility measure across all three forecasting error measures for ITM European call options using the Black-Scholes model (Table 7). In fact, the ranking of the measures follows that seen for all moneyness levels, as discussed above. The same conclusions are also reached for OTM European call options using the Black-Scholes model (Table 9).

On the other hand, when examining the predictive ability of ATM European call options using the Black-Scholes model, Table 8 shows that, while the best predictor is still the 100-day moving average volatility measure across all three forecasting error measures, the second-best volatility measure depends on which forecasting error measure is chosen. If either the MAE or the RMSE forecasting error measure is chosen, then the ranking of volatility measures when valuing ATM European call options using the Black-Scholes model follows those of all moneyness levels, too. However, if the MAPE measure is chosen, then the second-best volatility measure is the GARCH-based measure, the third-best volatility measure is then the 60-day moving average volatility measure, and the remaining measures are ranked as per the rest.

6.2. The Bachelier Model

This section analyses and discusses the results obtained using the data and methodology described in Chapter 5, as applied to the Bachelier model described in Section 2.2. The Bachelier model was used to calculate the prices of European call options over the period 1 March 2005 to 31 December 2019 over several maturities and moneyness levels. Five different volatility measures were used to price the options, and this section will discuss which of the five measures yielded the most accurate forecasts, as measured by the forecasting error measures discussed in Chapter 5.

Table 4 below depicts the MAE, the RMSE, and the MAPE values for the Bachelier model for each of the five volatility measures used in this study across all moneyness levels. As can be seen, when evaluating which of the five volatility measures produces forecasts which are closest in values to the actual prices of European call options based on the S&P 500 Index over all moneyness levels, the Bachelier model's results for the various volatility measures correspond to the results obtained for the Black-Scholes model, as discussed in Section 6.1. The 100-day moving average historical volatility measure produces the lowest values across all three forecasting error measures, followed by the 60-day moving average historical volatility measure, then the GARCH-based volatility measure, the 20-day moving average historical volatility measure and, finally, the 5-day moving average historical volatility measure.

Table 4: Forecasting Error Measures for the Bachelier Model for Various Volatility Estimators (All Moneyness Levels)

Measure	*MAE*	*RMSE*	*MAPE*
$\boldsymbol{\sigma_{HV(5)}}$	*52.249*	*71.258*	*45.789%*
$\boldsymbol{\sigma_{HV(20)}}$	*48.490*	*64.535*	*43.157%*
$\boldsymbol{\sigma_{HV(60)}}$	*45.343*	*59.271*	*41.104%*
$\boldsymbol{\sigma_{HV(100)}}$	*44.104*	*56.918*	*40.232%*
$\boldsymbol{\sigma_{GARCH}}$	*45.885*	*59.799*	*41.158%*

Note: This table depicts three forecasting error measures, namely the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE) for the Bachelier model, as used to calculate the prices of European call options. The model uses the daily log returns of the S&P 500 Index over the period 1 March 2005 to 31 December 2019 to calculate four moving average historical daily volatility measures of length 5, 20, 60, and 100 days, as well as a daily GARCH volatility measure for the period. The total number of observations for the period was 226,170.

As pointed out for the Black-Scholes model in Section 6.1, it is interesting that all three forecasting error measures are in agreement with each other. This is seen by the identical rank of each of the five volatility measures for each of the three forecasting error measures when examining the forecasts made by the Bachelier model when moneyness is not used to break the dataset down further.

Again, while it is not strictly within the scope of this study, it is also possible to evaluate which of the five volatility measures produces superior forecasts when the data are split by moneyness levels. Tables 10, 11, and 12 in Appendix 2 depict the forecasting error measures' figures achieved by the Bachelier model when pricing European call options for ITM, ATM, and OTM options, respectively. The analysis and conclusions for ITM, ATM, and OTM European call options when it comes to which volatility measure produces more accurate forecasts for the Bachelier model are identical to those given for the Black-Scholes model in Section 6.1 above.

6.3. The Artificial Neural Network

This section analyses and discusses the results obtained using the data and methodology described in Chapter 5, as applied to the ANN model described in Chapter 5. The ANN's performance is evaluated using the forecasting error measures discussed in Chapter 5.

The ANN used was a deep ANN with two hidden layers, each consisting of 30 neurons. It used sets of maturities, moneyness levels of the options, as well as the five volatility measures as inputs. Its output was the set of prices of the European call options on the S&P 500 Index for the period 1 March 2005 to 31 December 2019, and the data were split into a training set, a validation set, and a testing set in a 60:20:20 ratio. The network used the ELU function (see Equation (16)) as the activation function for the hidden layers and the Softplus function (see Equation (15)) as the activation function for the output function due to its positive output (corresponding to a non-negative price for European call options).

Since the ANN had all of the volatility measures calculated in Chapter 5 used as inputs simultaneously, it is impossible to compare the performance of the network per volatility measure to determine which volatility measure is superior. Moreover, as a single ANN was calibrated, regardless of the moneyness levels of the options tested, it is also impossible to compare the performance of the network per volatility measure by further breaking it down into different subsets based on moneyness levels.

Table 5 below depicts the MAE, the RMSE, and the MAPE values for the ANN used. As can be seen, the ANN performed exceptionally well during the training and validation periods and performed marginally worse during the testing period, achieving values close to zero for all three of the forecasting error measures.

Table 5: Forecasting Error Measures for the Artificial Neural Network

Period/Measure	*MAE*	*RMSE*	*MAPE*
Training and Validation	*0.0001839*	*0.0002253*	*0.0185%*
Testing	*0.0001854*	*0.0002268*	*0.0186%*

Note: This table depicts three forecasting error measures, namely the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE) for the Artificial Neural Network (ANN) model, as used to calculate the prices of European call options. The model uses sets of four moving average historical daily volatility measures of length 5, 20, 60, and 100 days, as well as a set of daily GARCH volatility values, all of which were calculated using daily log returns of the S&P 500 Index over the period 1 March 2005 to 31 December 2019, as inputs, together with a set of maturities and one of moneyness values. The output set was the set of prices of European call options on the S&P 500 Index for the same period. The total number of observations for the full period was 226,170, split between the training and validation periods (180,936) and the testing period (45,234).

6.4. Overall

This section combines the discussions and analyses of this chapter to allow for the comparison of the best volatility measure identified for each of the three option pricing models discussed. This entails a comparison of the Black-Scholes model, calibrated using the 100-day moving average historical volatility measure (see Section 6.1), the Bachelier model, also calibrated using the 100-day moving average historical volatility measure (see Section 6.2), and the ANN model.

Table 6 below depicts the MAE, the RMSE, and the MAPE values for each of the three models, as described above. As can be seen by each of the three measures, the ANN achieved forecasts which are significantly closer to the actual prices of European call options on the S&P 500 Index for the period in question. The results achieved with the ANN are truly remarkable, even more so considering that a single ANN was used for all maturity and moneyness levels combinations. Between the Black-Scholes model and the Bachelier model, the Black-Scholes is marginally superior in forecasting the prices of European call options based on each of the three forecast error measures.

Table 6: Forecasting Error Measures for the Black-Scholes Model, the Bachelier Model, and the Artificial Neural Network

Model/Measure	*MAE*	*RMSE*	*MAPE*
Black-Scholes	*44.014*	*56.836*	*39.570%*
Bachelier	*44.104*	*56.918*	*40.232%*
ANN	*0.0001854*	*0.0002268*	*0.0186%*

Note: This table depicts three forecasting error measures, namely the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE) for the three models used, namely the Black-Scholes model, the Bachelier model, and the Artificial Neural Network (ANN) model, as used to calculate the prices of European call options. The Black-Scholes model and the Bachelier model both use a 100-day moving average historical volatility as daily volatility, calculated using the daily log returns of the S&P 500 Index over the period 1 March 2005 to 31 December 2019. The ANN model used the same 100-day moving average historical volatility measure, together with 5-, 20-, and 60-day measures, and a GARCH-based volatility measure, along with other inputs. The total number of observations for the period was 226,170 for the Black-Scholes and Bachelier models, while it was 45,234 for the testing period of the ANN, and 180,936 for the calibration and validation period.

The conclusions reached above are all confirmed by the Diebold-Mariano test at the 1% significance level. The Diebold-Mariano test statistics were calculated for each combination of models (Black-Scholes versus Bachelier; Black-Scholes versus ANN; and Bachelier versus ANN) and all test statistics showed that the ANN produces significantly more accurate forecasts relative to each of the parameterised models, and the Black-Scholes model produces significantly more accurate forecasts relative to the Bachelier model.

The incredibly low values for the three forecasting measures obtained for the ANN in this study, which did not incorporate any modularity, are worth repeating. It is important to note that many previous studies implemented some form of modularity in the network based on either term to maturity or moneyness levels to achieve lower forecasting errors (as measures by the MAE, the RMSE, and the MAPE). Hence, the degree of accuracy achieved by the ANN employed in this study is remarkable. This may be the result of using a significantly longer period in this study relative to that used by Baruníkova and Baruník (2011) and Vach (2019).

The conclusion surrounding the superiority of the ANN is mostly in line with the literature, particularly the results presented by Vach (2019). Moreover, as noted by Vach, there are situations in which the moving average historical volatility measures are superior in their forecasting ability relative to GARCH-based models. This is true in this study as well, as the 100-day moving average historical volatility measure proved to produce superior forecasts in both of the parametric models used.

The conclusions surrounding the superiority of both the parametrised Black-Scholes model and the non-parameterised ANN model relative to the Bachelier model are the novel contribution of this study. These show that, while the Bachelier model allows for the use of negative strike prices on the underlying, the overall usability of such a model in practice is relatively poor when considered against its rivals with reference to predictive accuracy, as considered in this study. Hence, the exclusive use of the Bachelier model is not necessarily advised. However, the incorporation of the ability to incorporate and trade on negative strikes may prove useful if built into an in-house model by practitioners.

7. Conclusion

In conclusion, this study was conducted using data from 1 March 2005 to 31 December 2019 with the primary aim of testing the predictive ability of the Bachelier model relative to those of the Black-Scholes model and an ANN model when valuing European call options on the S&P 500 Index for the period. The study aimed to investigate whether the brief introduction of the Bachelier model by the CME Group, Inc., introduced the market to an option pricing model that is superior to other models which are commonly discussed in the literature.

The study examined all options available during the period for various maturities (ranging from one month to two years) and moneyness levels (ranging from 80% to 120%). Several options which violated certain conditions as detailed in Baruníkova and Baruník (2011) and Bakshi, et al., (1997) were excluded from the dataset, leaving a total number of 226,170 European call options in the dataset.

Several volatility measures were then calculated based on daily log returns of the S&P 500 Index, as done by Vach (2019). These were four moving average historical volatility measures (of lengths of 5, 20, 60, and 100 days), as well as a GARCH(1,1) volatility measure with an underlying normal distribution and a five-year rolling calibration period. These volatility inputs were used when calculating European call option prices using the two parameterised models considered in this study, the Black-Scholes model and the Bachelier model, thereby resulting in ten different sets of forecasts. The superior model for each of the parameterised models across the different volatility measures was determined by the application of three different forecasting error estimation measures, namely the MAE, the RMSE, and the MAPE measures. For both the Black-Scholes model and the Bachelier model, the 100-day moving average historical volatility measure proved superior over all other volatility measures across all three forecast error measures.

The ANN used in this study was a single network calibrated for all maturities and all levels of moneyness available in the dataset. The network had two hidden layers, each with 30 neurons. Its input layer had seven neurons, made up of the five volatility measures, the time to maturity of the option, and the moneyness levels of the option. Its output node was the price. The data was split between the training set, validation set, and testing set according to a 60:20:20 ratio, as used by Vach (2019). The accuracy of the ANN model was also evaluated across the three forecast error measures mentioned above, and the ANN performed remarkably well relative to both the Black-Scholes model and the Bachelier model.

To conclude which model was overall statistically better at producing more accurate forecasts, the Diebold-Mariano test was used. The statistical test revealed that the ANN model was statistically superior to both the Black-Scholes model and the Bachelier model when pricing European call options on the S&P 500 Index over the period examined at the 1% significance level. Moreover, the test showed that the Black-Scholes model was statistically superior to the Bachelier model when pricing European call options on the S&P 500 Index over the period examined at the 1% significance level.

Therefore, this study shows that, while the brief re-introduction of the Bachelier model to the markets was novel, its pricing ability when it comes to European call options would not have yielded better results for its users, as their use of the commonplace Black-Scholes model would have provided them with more accurate pricing for European call options on the S&P 500 Index which is statistically significant. Therefore, it makes sense why the Bachelier model's use by the CME Group, Inc., ceased fairly quickly after its re-introduction, and the Group reverted to the use of models such as the Black-Scholes model.

The limitations encountered when performing this study primarily surrounded the access to data. The data available for options on the S&P 500 Index on the database used were limited in both moneyness levels and time to maturity. Only very few values for each were available, which limited the scope of the study. Further research into the effects of more ARCH-type models, such as exponential GARCH (EGARCH) and fractionally integrated GARCH (FIGARCH), when predicting volatility values may add to the robustness of the results. Moreover, the use of the Harvey-Leybourne-Newbold test, introduced by Harvey, Leybourne, and Newbold (1997), would further add robustness to the results. This study can also be extended to include put options, although a more interesting avenue would be including American options. Finally, it would be interesting to see the Bachelier model applied to an underlying which commonly sees its strike being negative, as is the case with interest rates in several parts of the world.

References

Bachelier, L., 1900. *Théorie de la Spéculation.* Translated from French by M. Davis and A. Etheridge, 2006. New Jersey: Princeton University Press.

Barber, D., 2012. Machine Learning Concepts. In: *Bayesian Reasoning and Machine Learning.* New York: Cambridge University Press, pp. 305-321.

Baruníkova, M. & Baruník, J., 2011. *Neural Networks as a Semiparameteric Option Pricing Tool,* Prague: s.n.

Basheer, I. A. & Hajmeer, M., 2000. Artificial Neural Networks: Fundamentals, Computing, Design, and Application. *Journal of Microbiological Methods,* Volume 43, pp. 3-31.

Bakshi, G., Cao, C. & Chen, Z., 1997. Empirical Performance of Alternative Option Pricing Models. *Journal of Finance,* 52(5), pp. 2003-2049.

Baughman, D. R. & Liu, Y. A., 2014. *Neural Networks in Bioprocessing and Chemical Engineering,* Blacksburg: Virginia Polytechnic Institute and State University.

Bennell, J. & Sutcliffe, C., 2004. Black-Scholes versus Artificial Neural Networks in Pricing FTSE 100 Options. *Intelligent Systems in Accounting, Finance and Management,* Volume 12, pp. 243-260.

Black, F. & Scholes, M., 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy,* 81(3), pp. 637-654.

Christoffersen, P. & Jacobs, K., 2004. Which GARCH Model for Option Valuation?. *Management Science,* 50(9), pp. 1204-1221.

Clevert, D.-A., Unterthiner, T. & Hochreiter, S., 2016. *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs).* San Juan, International Conference on Learning Representations.

CME Group, Inc., 2020. *Switch to Bachelier Options Pricing Model - Effective April 22, 2020,* Chicago: CME Group, Inc..

Cong, T., Su, G., Qiu, S. & Tian, W., 2013. Applications of ANNs in Flow and Heat Transfer Problems in Nuclear Engineering: A Review Work. *Progress in Nuclear Energy,* Volume 62, pp. 54-71.

Cootner, P. H., 1964. *The Random Character of Stock Market Prices.* Cambridge: MIT Press.

Davis, M. & Etheridge, A., 2006. *Louis Bachelier's Theory of Speculation.* 1st Edition ed. New Jersey: Princeton University Press.

Diebold, F. X. & Mariano, R. S., 1995. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics,* 13(3), pp. 253-263.

Diebold, F. X., 2015. Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests. *Journal of Business & Economic Statistics,* 33(1).

Dougas, C., Bengio, Y., Bélisle, F., Nadeau, C. & Garcia, R., 2001. Incorporating Second-Order Functional Knowledge for Better Option Pricing. *Advances in Neural Information Systems,* pp. 472-478.

Engle, R., 2001. GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics. *Journal of Economic Perspectives,* 15(4), pp. 157-168.

Fall, A. N., Ndiaye, S. N. & Ndolane, S., 2019. Black-Scholes Option Pricing Equations Described by the Caputo Generalized Fractional Derivative. *Chaos, Solitons & Fractals,* Volume 125, pp. 108-118.

Figlewski, S., 1997. Forecasting Volatility. *Financial Markets, Institutions & Instruments,* 6(1), pp. 1-88.

Fukushima, K., 1980. Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Bilogical Cybernetics,* Volume 36, pp. 193-202.

Garcia, R. & Gençay, R., 2000. Pricing and Hedging Derivative Securities with Neural Networks and a Homogeneity Hint. *Journal of Econometrics,* Volume 94, pp. 93-115.

Gastineau, G. L. & Madansky, A., 1979. Why Simulations are an Unreliable Test of Option Strategies. *Financial Analysts Journal,* 35(5), pp. 61-76.

Geigle, D. S., 2000. *An Artificial Neural Network Approach to the Valuation of Options and Forecasting Volatility,* Fort Lauderdale, Florida: Nova Southeastern University.

Gradojevic, N., Gençay, R. & Kukolj, D., 2009. Option Pricing with Modular Neural Networks. *IEEE Transactions on Neural Networks,* 20(4), pp. 626-637.

Hahn, J. T., 2013. *Option Pricing Using Artificial Neural Networks: An Australian Perspective,* Robina, Australia: Bond University.

Hara, K., Saito, D. & Shouno, H., 2015. *Analysis of Function of Rectified Linear Unit Used in Deep Learning.* Killarney, International Joint Conference on Neural Networks (IJCNN), pp. 1-8.

Harvey, D., Leybourne, S. & Newbold, P., 1997. Testing the Equality of Prediction Mean Squared Errors. *International Journal of Forecasting,* Volume 13, pp. 281-291.

He, K., Zhang, X., Ren, S. & Sun, J., 2016. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition,* pp. 770-778.

Heston, S. L. & Nandi, S., 2000. A Closed-Form GARCH Option Valuation Model. *The Review of Financial Studies,* 13(3), pp. 585-625.

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), pp. 327-343.

Hornik, K., 1991. Approximation Capabilities of Multilayer Feedforward Networks. *Neural Networks,* Volume 4, pp. 251-257.

Hornik, K., Stinchcombe, M. & White, H., 1989. Multilayer Feedforward Networks are Universal Approximators. *Neural Networks,* Volume 2, pp. 359-366.

Huang, W. & Zheng, Y., 2020. COVID-19: Structural Changes in the Relationship between Investor Sentiment and Crude Oil Futures Price. *Energy Research Letters,* 1(2).

Huang, Y., Kangas, L. J. & Rasco, B. A., 2007. Applications of Artificial Neural Networks (ANNs) in Food Science. *Critical Reviews in Food Science and Nutrition,* 47(2), pp. 113-126.

Hull, J. & White, A., 1987. The Pricing of Options on Assets with Stochastic Volatilities. *Journal of Finance,* 42(2), pp. 281-300.

Hutchinson, J. M., Lo, A. W. & Poggio, T., 1994. A Nonparametric Approach to Pricing and Hedging Derivative Securities Via Learning Networks. *Journal of Finance,* 49(3), pp. 851-889.

Jagannathan, R., 1984. Call Options and the Risk of Underlying Securities. *Journal of Financial Economics,* Volume 13, pp. 425-434.

Jain, A. K. & Mao, J., 1996. Artificial Neural Networks: A Tutorial. *IEEE Computer,* Volume 29, pp. 31-44.

Kolmogorov, A. N., 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer.

Kou, S. G., 2002. A Jump-Diffusion Model for Option Pricing. *Management Science,* 48(8), pp. 1086-1101.

Lajbcygier, P., 2002. Comparing Conventional and Artificial Neural Network Models for the Pricing of Options. In: *Neural Networks in Business: Techniques and Applications.* s.l.:IGI Global, pp. 220-235.

Leland, H. E., 1985. Option Pricing and Replication with Transaction Costs. *Journal of Finance,* 40(5), pp. 1283-1301.

Long, J. B., 1974. Session Topic: Risk, Information and Capital Budgeting: Discussion. *Journal of Finance,* 29(2), pp. 485-488.

Ma, C. & Wang, X., 2020. Why Oil Prices Plunged and Settled Negative: A Game-Theoretical Perspective. In: G. X. Yuan, ed. *The CME Vulnerability: The Impact of Negative Oil Futures Trading.* Guangzhou: World Scientific Publishing, pp. 105-130.

Magdziraz, M., Orzeł, S. & Aleksander, W., 2011. Option Pricing in Subdiffusive Bachelier Model. *Journal of Statistical Physics,* Volume 145, pp. 187-203.

Malliaris, M. & Salchenberger, L., 1993. A Neural Network Model for Estimating Option Prices. *Journal of Applied Intelligence 3,* Volume 3, pp. 193-206.

Marquez, R., Pedro, H. T. C. & Coimbra, C. F. M., 2013. Hybrid Solar Forecasting Method Uses Satellite Imaging and Ground Telemetry as Inputs to ANNs. *Solar Energy,* Volume 92, pp. 176-188.

Merton, R. C., 1998. Applications of Option-Pricing Theory: Twenty-Five Years Later. *The American Economic Review,* 88(3), pp. 323-349.

Mitchell, T. M., 2006. *The Discipline of Machine Learning.* [Online]
Available at: http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf
[Accessed 26 August 2019].

Murphy, J. A., 1990. A Modification and Re-Examination of the Bachelier Option Pricing Model. *The American Economist,* 34(2), pp. 34-41.

Musiela, M. & Rutkowski, M., 2005. *Martingale Methods in Financial Modelling.* 2nd Edition ed. Berlin: Springer.

Nunes, M., Gerding, E., McGroarty, F. & Niranjan, M., 2018. *Artificial Neural Networks in Fixed Income Markets for Yield Curve Forecasting,* s.l.: SSRN.

Pritsker, M., 2006. The Hidden Dangers of Historical Simulation. *Journal of Banking & Finance,* Volume 30, pp. 561-582.

Qiumei, Z., Dan, T. & Fenghua, W., 2019. Improved Convolutional Neural Network Based on Fast Exponentially Linear Unit Activation. *IEEE Access,* Volume 7, pp. 151359-151367.

Rivals, I. & Personnaz, L., 2003. Neural-Network Construction and Selection in Nonlinear Modeling. *IEEE Transactions on Neural Networks,* 14(4), pp. 804-819.

Rosenblatt, F., 1958. A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review,* Volume 65, pp. 386-408.

Rubinstein, M., 1985. Nonparameteric 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 Aufust 31, 1978. *Journal of Finance,* 40(2), pp. 455-480.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J., 1986. Learning Representations by Back-Propagating Errors. *Nature,* Volume 323, p. 533–536.

Samuelson, P. & Merton, R. C., 1969. A Complete Model of Warrant Pricing that Maximizes Utility. *Industrial Management Review,* 10(2), pp. 17-46.

Schachermayer, W. & Teichmann, J., 2008. How Close are the Option Pricing Formulas of Bachelier and Black-Merton-Scholes?. *Mathematical Finance,* 18(1), pp. 155-170.

Scott, L. O., 1987. Option Pricing when the Variance Changes Randomly: Theory, Estimation, and an Application. *The Journal of Financial and Quantitative Analysis,* 22(4), pp. 419-438.

Sibi, P., Jones, S. A. & Siddarth, P., 2013. Analysis of Different Activation Functions Using Back Propagation Neural Networks. *Journal of Theoretical and Applied Information Technology,* 47(3), pp. 1264-1268.

Smith, C. W., 1976. Option Pricing: A Review. *Journal of Financial Economics,* Volume 3, pp. 3-51.

Stark, L., 2017. *Machine Learning and Options Pricing: A Comparison of Black-Scholes and a Deep Neural Network in Pricing and Hedging DAX 30 Index Options,* Espoo: Aalto University.

Sui, C. & Yang, M., 2020. Tanker Shipping and Negative Oil Prices: More than Just the Freight Rates. In: G. X. Yuan, ed. *The CME Vulnerability: The Impact of Negative Oil Futures Trading.* Guangzhou: World Scientific Publishing, pp. 131-146.

Sullivan, E. J. & Weithers, T. M., 1991. Louis Bachelier: The Father of Modern Option Pricing Theory. *The Journal of Economic Education,* 22(2), pp. 165-171.

Suthaharan, S., 2014. Big Data Classification: Problems and Challenges in Network Intrusion Prediction with Machine Learning. *Performance Evaluation Review,* 41(4), pp. 70-73.

Thorp, E. O. & Kassouf, S. T., 1967. *Beat the Market: A Scientific Stock Market System.* New York: Random House.

Vach, D., 2019. *Artificial Neural Networks in Option Pricing,* Prague: Charles University.

Vanstone, B. & Finnie, G., 2010. Enhancing Stockmarket Trading Performance with ANNs. *Expert Systems with Applications,* Volume 37, pp. 6602-6610.

Wiggins, J. B., 1987. Options Values under Stochastic Volatility: Theory and Empirical Estimates. *Journal of Financial Economics,* 19(2), pp. 351-372.

Yang, Y., Zheng, Y. & Hospedales, T. M., 2017. Gated Neural Networks for Option Pricing: Rationality by Design. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence,* 31(1).

Yavuz, M. & Özdemir, N., 2018. A Different Approach to the European Option Pricing Model with New Fractional Operator. *Mathematical Modelling of Natural Phenomena,* 13(1), pp. 12-23.

Yuan, G. X., 2020. *The CME Vulnerability: The Impact of Negative Oil Futures Trading.* Guangzhou: World Scientific Publishing.

Appendices

Appendix 1: Forecasting Error Measures for the Black-Scholes Model (ITM, ATM, OTM)

Table 7: Forecasting Error Measures for the Black-Scholes Model for Various Volatility Estimators (ITM)

Measure	*MAE*	*RMSE*	*MAPE*
$\sigma_{HV(5)}$	*33.433*	*49.160*	*11.379%*
$\sigma_{HV(20)}$	*31.561*	*45.084*	*10.748%*
$\sigma_{HV(60)}$	*29.913*	*42.039*	*10.176%*
$\sigma_{HV(100)}$	*29.292*	*40.653*	*9.962%*
σ_{GARCH}	*30.544*	*42.861*	*10.337%*

Note: This table depicts three forecasting error measures, namely the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE) for the Black-Scholes model, as used to calculate the prices of in-the-money (ITM) European call options. The model uses the daily log returns of the S&P 500 Index over the period 1 March 2005 to 31 December 2019 to calculate four moving average historical daily volatility measures of length 5, 20, 60, and 100 days, as well as a daily GARCH volatility measure for the period. The total number of observations for the period was 51,238.

Table 8: Forecasting Error Measures for the Black-Scholes Model for Various Volatility Estimators (ATM)

Measure	*MAE*	*RMSE*	*MAPE*
$\sigma_{HV(5)}$	*62.479*	*79.830*	*41.486%*
$\sigma_{HV(20)}$	*57.630*	*72.251*	*38.503%*
$\sigma_{HV(60)}$	*53.635*	*66.212*	*36.093%*
$\sigma_{HV(100)}$	*52.033*	*63.524*	*35.065%*
σ_{GARCH}	*53.939*	*66.392*	*35.915%*

Note: This table depicts three forecasting error measures, namely the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE) for the Black-Scholes model, as used to calculate the prices of at-the-money (ATM) European call options. The model uses the daily log returns of the S&P 500 Index over the period 1 March 2005 to 31 December 2019 to calculate four moving average historical daily volatility measures of length 5, 20, 60, and 100 days, as well as a daily GARCH volatility measure for the period. The total number of observations for the period was 102,476.

Table 9: Forecasting Error Measures for the Black-Scholes Model for Various Volatility Estimators (OTM)

Measure	*MAE*	*RMSE*	*MAPE*
$\sigma_{HV(5)}$	*50.607*	*70.616*	*73.869%*
$\sigma_{HV(20)}$	*47.160*	*64.141*	*70.558%*
$\sigma_{HV(60)}$	*44.216*	*58.956*	*68.000%*
$\sigma_{HV(100)}$	*43.084*	*56.634*	*66.880%*
σ_{GARCH}	*44.937*	*59.809*	*68.237%*

Note: This table depicts three forecasting error measures, namely the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE) for the Black-Scholes model, as used to calculate the prices of out-of-the-money (OTM) European call options. The model uses the daily log returns of the S&P 500 Index over the period 1 March 2005 to 31 December 2019 to calculate four moving average historical daily volatility measures of length 5, 20, 60, and 100 days, as well as a daily GARCH volatility measure for the period. The total number of observations for the period was 72,456.

Appendix 2: Forecasting Error Measures for the Bachelier Model (ITM, ATM, OTM)

Table 10: Forecasting Error Measures for the Bachelier Model for Various Volatility Estimators (ITM)

Measure	*MAE*	*RMSE*	*MAPE*
$\sigma_{HV(5)}$	*38.001*	*55.094*	*12.481%*
$\sigma_{HV(20)}$	*35.831*	*50.388*	*11.777%*
$\sigma_{HV(60)}$	*33.925*	*46.894*	*11.144%*
$\sigma_{HV(100)}$	*33.198*	*45.302*	*10.906%*
σ_{GARCH}	*34.621*	*47.819*	*11.315%*

Note: This table depicts three forecasting error measures, namely the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE) for the Bachelier model, as used to calculate the prices of in-the-money (ITM) European call options. The model uses the daily log returns of the S&P 500 Index over the period 1 March 2005 to 31 December 2019 to calculate four moving average historical daily volatility measures of length 5, 20, 60, and 100 days, as well as a daily GARCH volatility measure for the period. The total number of observations for the period was 51,238.

Table 11: Forecasting Error Measures for the Bachelier Model for Various Volatility Estimators (ATM)

Measure	*MAE*	*RMSE*	*MAPE*
$\sigma_{HV(5)}$	*63.286*	*80.947*	*41.764%*
$\sigma_{HV(20)}$	*58.352*	*73.153*	*38.752%*
$\sigma_{HV(60)}$	*54.294*	*67.000*	*36.321%*
$\sigma_{HV(100)}$	*52.665*	*64.260*	*35.286%*
σ_{GARCH}	*54.657*	*67.223*	*36.146%*

Note: This table depicts three forecasting error measures, namely the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE) for the Bachelier model, as used to calculate the prices of at-the-money (ATM) European call options. The model uses the daily log returns of the S&P 500 Index over the period 1 March 2005 to 31 December 2019 to calculate four moving average historical daily volatility measures of length 5, 20, 60, and 100 days, as well as a daily GARCH volatility measure for the period. The total number of observations for the period was 102,476.

Table 12: Forecasting Error Measures for the Bachelier Model for Various Volatility Estimators (OTM)

Measure	*MAE*	*RMSE*	*MAPE*
$\sigma_{HV(5)}$	*46.715*	*66.606*	*75.035%*
$\sigma_{HV(20)}$	*43.493*	*60.301*	*71.578%*
$\sigma_{HV(60)}$	*40.758*	*55.334*	*69.056%*
$\sigma_{HV(100)}$	*39.709*	*53.114*	*67.966%*
σ_{GARCH}	*41.444*	*56.161*	*69.350%*

Note: This table depicts three forecasting error measures, namely the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE) for the Bachelier model, as used to calculate the prices of out-of-the-money (OTM) European call options. The model uses the daily log returns of the S&P 500 Index over the period 1 March 2005 to 31 December 2019 to calculate four moving average historical daily volatility measures of length 5, 20, 60, and 100 days, as well as a daily GARCH volatility measure for the period. The total number of observations for the period was 72,456.