Risk Management (2025) 27:8
https://doi.org/10.1057/s41283-025-00160-0

ORIGINAL ARTICLE

A comparative analysis of option pricing models: Black–Scholes, Bachelier, and artificial neural networks

Eden Gross[1] · **Ryan Kruger**[2] · **Francois Toerien**[2]

Accepted: 20 March 2025

Abstract
Practitioners and academics alike have applied the Black–Scholes model when pricing options practically since its introduction in 1973. The COVID-19 pandemic and the oil futures price crash of April 2020 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 the Bachelier model and the Black–Scholes model when pricing European call options on the Standard & Poor's 500 index using five volatility estimation methods. Moreover, it compares the forecasts of the two parameterized models to a deep feed-forward artificial neural network (ANN) which is also used to price such options. Overall, the ANN is statistically superior in its predictive ability relative to both parameterized models, and there is no statistical difference in predictive ability between the Black–Scholes and Bachelier models. These results are of relevance to both academics and participants in the options markets.

Keywords Option pricing · Bachelier · Black–Scholes · Artificial neural network · Machine learning

✉ Eden Gross
eden.gross@wits.ac.za

Ryan Kruger
ryan.kruger@uct.ac.za

Francois Toerien
francois.toerien@uct.ac.za

1 School of Statistics and Actuarial Science, University of the Witwatersrand, Johannesburg, South Africa

2 Department of Finance and Tax, University of Cape Town, Cape Town, South Africa

Published online: 09 April 2025

Introduction

The option pricing model of Black and Scholes (1973) revolutionized the use of derivatives in the financial industry by allowing for a relatively straightforward and theoretically sound valuation of options. Today, the Black–Scholes model is the standard model employed by option traders to value exchange-traded options. However, despite its widespread use and popularity, the Black–Scholes model has been criticized for making some simplifying assumptions that do not hold in practice. With respect to equities, these assumptions include constant volatility, the random walk nature of returns, no dividends or transaction costs, a constant risk-free rate, and perfect liquidity.

Because of these and other shortcomings, much work has gone into attempting to develop improved, more realistic option valuation models. Of interest, starting with Malliaris and Salchenberger (1993) and Hutchison et al. (1994), a significant and continuously growing body of research[1] has been generated that investigates the use of artificial neural networks (ANNs) as a non-parametric alternative to the Black–Scholes model for the valuation of options. Moreover, a third model for option valuation, that pre-dates the Black–Scholes model, resurfaced recently as a potential alternative. Unlike the Black-Scholde model, the Bachelier (1900) model[2] accommodates negative price values for the asset underlying an option, which led to its brief comeback during early 2020 for option valuation after the contract futures price for West Texas Intermediate (WTI) briefly dipped into negative territory.[3] For a recent exploration of potential adaptations and uses of the Bachelier model in practice, see Nyarko, et al. (2023).

The use of the Bachelier model to price options, although temporary and resulting from unusual circumstances, once more raises questions surrounding the predictive abilities of parametric models (such as Bachelier and Black–Scholes) against those of non-parametric models (such as the ANN). Given that the global trade in exchange-traded options alone reached 56.17 billion option contracts in 2022 (World Federation of Exchanges 2022), the efficient pricing of options is of great economic importance and, hence, worth investigating. Thus, the purpose of this study is to comprehensively compare the predictive accuracy of the Bachelier model to (i) the Black–Scholes model, and (ii) an ANN designed to predict the price of options. Further, given the dominance of the U.S. in the global financial market and following many other studies,[4] this study focuses specifically on the pricing of European call options on the Standard & Poor's (S&P) 500 index, which are among the most widely traded options in the world.

[1] For a review of research on neural networks in option pricing, see Ruf and Wang (2020).

[2] Originally published in French. For a translation into English, see Davis and Etheridge (2006).

[3] For example, the CME Group, Inc., which operates the Chicago Board of Trade, Chicago Mercantile Exchange, the Commodity Exchange, and the New York Mercantile Exchange, moved to the Bachelier model on 21 April 2020 and moved back to its prior model on 13 August 2020 (Yuan 2020).

[4] See, for example, Bakshi, Cao, and Chen (1997), Garcia and Gençay (2000), and Vach (2019).

Whereas most existing studies compare two of the option pricing methodologies (especially the Black–Scholes model and ANNs[5]), there is a lack of studies comparing all three, and a particular gap in studies that include the Bachelier model in this comparison. Thus, the main contribution of this study is to rigorously compare all three models in the context of S&P 500 index European call options using three metrics of predictive ability and the Diebold–Mariano statistic (Diebold & Mariano 1995). As it is not the aim of this study to find the best ANN structure for pricing options, an ANN previously identified as optimal by Vach (2019) is used to represent this approach.

Considering the size of the U.S. options market and the important role it plays in the global financial market, this research is of great practical interest to industry practitioners and option traders. It is further relevant to the large academic body of knowledge on option pricing, as it attempts to address the gap in the literature regarding a comprehensive three-way comparison of the main option pricing approaches.

The remainder of this article is structured as follows. Sect. "Literature review" examines the prior research covering option pricing, with specific reference to the three models discussed in this study. Sect. "Data and methodology" details the data and methodology adopted in this study, followed by the results and their corresponding analyses in Sect. "Results and analysis". Sect. "Conclusion" concludes the article.

Literature review

The Bachelier and the Black–Scholes models are both parameterized stochastic option pricing models, each of which has a closed-form solution. Both models make assumptions regarding 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), whereas the Bachelier model was developed with this notion assumed implicitly decades before the notion of risk appetite was developed.

The two models differ in the fundamental principles that underly them. Thus, the Black–Scholes model follows from 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). The price of a call option on an underlying with a price of S_t for the period t^*, with exercise price X and risk-free rate r, using Black–Scholes model, together with its parameters, is depicted in Equation (1), while the same is depicted in Equation (2) using the Bachelier model, below.

[5] For some studies comparing the Black–Scholes model and specific ANNs, see Malliaris and Salchenberger (1993), Geigle (2000) and Bennell and Sutcliffe (2004).

$$
\begin{aligned}
C_{S_t,t^*} &= S_t\Phi\left(d_1\right) - Xe^{-rt^*}\Phi\left(d_2\right) \\
d_1 &= \frac{\ln\frac{S_t}{X} + \left(r + \frac{1}{2}\sigma^2\right)(t^*)}{\sigma\sqrt{t^*}} \\
d_2 &= \frac{\ln\frac{S_t}{X} + \left(r - \frac{1}{2}\sigma^2\right)(t^*)}{\sigma\sqrt{t^*}}
\end{aligned} \tag{1}
$$

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

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

where $\phi(d)$ is the probability density function of the normal distribution at point d.

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 parameterize 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. (1994) admit that the ANN is not a substitute for 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 suggest 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). They 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.

Garcia and Gençay (2000) 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, concluding that an ANN model is superior to the Black–Scholes model when implementing hedging strategies using options.

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), who conclude that such modeling 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.

The importance of these studies is that they 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. For example, Malliaris and Salchenberger (1993) find 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. Thus, the Black–Scholes model seems to over-price deep-out-of-the-money options, while both models underprice in-the-money options, with the Black–Scholes model being slightly more accurate. 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.

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 (2013) finds that non-parametric models such as the ANN offers superior results when using stochastic methods to estimate volatility, specifically relative to more traditional models such as the Black–Scholes model.

Vach (2019) examines several structures as candidates for the optimal structure of an option pricing ANN. 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 are incorporated.[6] Vach (2019) 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 to 100 days). Finally, the yield on monthly U.S. Treasury bills is used as a proxy for 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 is used 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 utilized by the ANN depend on

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

the layer in question, with the sigmoid function employed in the hidden layers and the Softplus function employed in the output layer. In addition, a non-zero dropout rate and modularity are found not to add to the predictive ability of the ANN when pricing options.

In contrast to the ANN literature, the literature regarding comparisons of the Black–Scholes and Bachelier models focuses mainly on the similarities and differences of these models (see, for example, Choi et al. 2022; specifically, the detail regarding their derivations and assumptions). There is, therefore, a clear gap in the literature for an actual comparison of performance and analyses regarding the models' predictive abilities over a meaningful study period, as we undertake in this article.

Data and methodology

This study follows several others, especially that of Baruníkova and Baruník (2011), by using the S&P 500 index as a proxy to the U.S. equities market and evaluating European call options based on this 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. The period in question covers almost 15 years (3671 trading days[7]) and 230,571 options.[8] The longer period allows for better calibration of the ANN. Moreover, the period is split into training, validation, and testing sets, as indicated below.

Other data collected from the Bloomberg database include daily rates for the 6-month London Inter-bank Offer Rate (LIBOR) as proxy for the risk-free rate, and maturities, prices, and strikes for European call options based on the S&P 500 index for various values of moneyness[9] (ranging from 80 to 120%). Maturities for the options range from one month to two years.

Following Baruníkova and Baruník (2011), several exclusions to the data were considered. We exclude any option whose price quote is lower than \$0.375 to remove any potential adverse effects of the discreteness of an option's price on the valuation. This condition is applied in this study for option prices based on both the Black–Scholes model and for those based on the Bachelier model. This filter excludes 4401 options, leaving 226,170 options. Last, we 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. This leads to no exclusions of any options in the data

[7] 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.

[8] The two parameterized models used approximately 20% of the data available, a total of 45,836 options, to provide comparable analyses to the testing dataset of the ANN used in this study.

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

used in this study based on both models. Hence, the total number of options left in the dataset is 226,170.

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 parameterized model, together with the volatility inputs outlined below[10]). The ANN is constructed of two hidden layers with 30 neurons each. Vach (2019) chooses the activation function of hidden layers arbitrarily to be the sigmoid function. However, this study employs the exponential linear unit (ELU) activation function, as the use of the sigmoid function resulted in a dead gradient[11] while training the data. The hidden layers do not make use of a dropout.[12] The output layer is constructed of a single output neuron, utilizing the Softplus activation function, as it ensures a positive output (as is the case for option prices in general), as used by Vach (2019). 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 (2019).

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 we do adopt the convention of scaling strikes by exercise prices, we do not sub-divide the network into nine different ANNs but, rather, we use a single ANN to price across all moneyness and maturities. However, we do sub-divide the database based on moneyness when applying the parameterized models 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). Although Baruníkova and Baruník (2011) use 1.03 as their boundary value to distinguish between ATM and ITM call options, this study adopts the more recent value proposed by Vach (2019).

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 uses 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 (2019) finds that

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

[11] 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.

[12] 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).

Table 1 Descriptive statistics

Measure	Number of observations	Mean	Standard deviation	Minimum	Maximum
r_t	3671	0.03%	1.18%	−9.47%	10.96%
r_f	3671	1.91%	1.69%	0.32%	5.64%
C	226170	\$161.65	\$119.40	\$0.54	\$759.10

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 for 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

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 nonetheless follows the approach of Vach (2019) by utilizing both moving average and GARCH measures to calculate volatility. This allows this study to offer more robust results regarding the predictive accuracy of the models in question.

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.

$$r_t = \ln\left(\frac{S_t}{S_{t-1}}\right) \tag{3}$$

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, Eq. (3) 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.

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.

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

$$\sigma_{HV(\tau)} = \sqrt{\frac{1}{\tau-1}\sum_{t=1}^{\tau}\left(r_\tau - \bar{r}_{(t.\tau)}\right)^2} \tag{4}$$

where $\sigma_{HV(\tau)}$ is the historical volatility for a historical period of length τ; r_t is the log return (as described in Eq. (3)); 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

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%

This table depicts some key statistics of the various moving averages and GARCH volatilities calculated, namely the mean, standard deviation, minimum, and maximum. The annualized 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

moving average historical volatility experienced over the preceding 100 days. The measure calculated in Eq. (4) measures a τ-day daily volatility, which must be annualized. This is achieved by calculating the product of the square root of 252 and the τ-day daily volatility.[13]

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 on the underlying asset, stated as follows.

$$r_t = \bar{r}_t + \sqrt{\sigma_t}\varepsilon_t \tag{5}$$

where r_t is as described in Eq. (3); $\bar{r}_t$ 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.

$$\sigma_{t+1} = \sqrt{\omega + \alpha\sigma_t^2\varepsilon_t^2 + \beta\sigma_t^2} \tag{6}$$

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.[14]

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

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

14 The GARCH model parameters must adhere to the following: (i) $\alpha, \beta, \omega > 0$; (ii) $\alpha + \beta < 1$; and (iii) the long-term average of the process is defined to be $\sqrt{\omega/[1-(\alpha+\beta)]}$. For more detail, see Engle (2001).

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 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.

When it comes to determining which model is superior in its predictive ability, we use 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 Eqs. (7), (8), and (9), 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 (measure-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.

$$\mathrm{MAE} = \frac{1}{N}\sum_{n=1}^{N}\left|y_n - \hat{y}_n\right| \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(y_n - \hat{y}_n\right)^2} \tag{8}$$

$$\mathrm{MAPE} = \frac{1}{N}\sum_{n=1}^{N}\left|\frac{y_n - \hat{y}_n}{y_n}\right| \tag{9}$$

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 using the relevant model.

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 (Diebold & Mariano 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 Diebold 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,[15] the resulting difference is then termed the loss differential, denoted by Diebold and Mariano (1995) as d. The average loss differential for the two models' time series is then denoted by $\overline{d}$. The authors then define the autocovariance of the loss differential at lag k, denoted γ_k, as follows.

$$\gamma_k = \frac{1}{N} \sum_{i=k+1}^{n} \left(d_i - \overline{d}\right)\left(d_{i-k} - \overline{d}\right) \tag{10}$$

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.

$$\mathrm{DM} = \frac{\bar{d}}{\sqrt{\widehat{Var}\left[\bar{d}\right]}} \tag{11}$$

where DM is the Diebold–Mariano statistic; and $\widehat{Var}\left[\overline{d}\right]$ is a consistent estimator of the sampling distribution of the average loss differential measure, derived using the autocovariance function in Equation (10) (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.

Results and analysis

The Black–Scholes and Bachelier models were tested using five volatility measures over 20% of the data and over all moneyness levels to correspond to the volatility conditions, testing period, and moneyness levels used to train, calibrate, and test the

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

Table 3 Forecasting error measures for the Black–Scholes model, the bachelier model, and the artificial neural network

Volatility Measure		MAE	RMSE	MAPE (%)
Model	Measure/Period			
Black–Scholes	$\boldsymbol{\sigma}_{\mathbf{HV}(5)}$	83.994	103.708	51.259
	$\boldsymbol{\sigma}_{\mathbf{HV}(20)}$	77.854	95.687	48.629
	$\boldsymbol{\sigma}_{\mathbf{HV}(60)}$	72.727	88.261	46.467
	$\boldsymbol{\sigma}_{\mathbf{HV}(100)}$	70.322	84.521	45.509
	$\boldsymbol{\sigma}_{\mathbf{GARCH}}$	74.968	89.640	47.252
Bachelier	$\boldsymbol{\sigma}_{\mathbf{HV}(5)}$	84.048	103.634	51.930
	$\boldsymbol{\sigma}_{\mathbf{HV}(20)}$	77.892	95.581	49.232
	$\boldsymbol{\sigma}_{\mathbf{HV}(60)}$	72.769	88.190	47.064
	$\boldsymbol{\sigma}_{\mathbf{HV}(100)}$	70.369	84.478	46.116
	$\boldsymbol{\sigma}_{\mathbf{GARCH}}$	75.022	89.586	47.890
ANN	Training and Validation	0.0001839	0.0002253	0.0185
	Testing	0.0001854	0.0002268	0.0186

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, the Bachelier model, and the artificial neural network (ANN), as used to calculate the prices of European call options. For the parametric models, the models used the daily log returns of the S&P 500 index over the period 1 March 2005 to 31 December 2019, using approximately 20% of options to test the model and 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, of which 45,836 options were used for testing, representing approximately 20% of the available dataset. For the ANN, the model used 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 226170 split between the training and validation periods (180936) and the testing period (45234). Note that the forecasting error measures' values were calculating using the raw data, without the use of a logarithmic scale

ANN. The results of the three models are presented in Table 3, below, depicting the results over each of the three error measures, followed by the relevant analyses.

The results for the Black–Scholes and Bachelier models show that the three error measures agree, concluding that the 100-day volatility measure produces the most accurate forecasts under both models. The results for the Black–Scholes and Bachelier models broken down by moneyness level are not significantly different to those presented in Table 3, above.

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 (2019). This may play a role in the fact that the 100-day moving average historical volatility measure can be concluded as superior to the GARCH-based volatility measure.

Table 4 Diebold–Mariano statistics comparing the forecasting accuracy of the models

Models tested	Diebold–Mariano Statistic	*p*-value
Black–Scholes versus Bachelier	1.51	0.06597
Bachelier versus Black–Scholes	−1.51	0.1319
Black Scholes versus ANN	185.45	2.2e–16
Bachelier versus ANN	187.00	2.2e–16

This table depicts the results of the Diebold–Mariano test, testing the forecast accuracies of the models compared. The null hypothesis of this test is that the two models compared have the same forecasting accuracy, while the alternative hypothesis is that the second model named has greater forecasting accuracy relative to the first model named. The models were used to calculate the prices of European call options based on the S&P 500 index for the period 1 March 2005 to 31 December 2019, each of the models being tested on approximately 20% of the total options over the period. The Black–Scholes model and Bachelier model were calibrated using a 100-day moving average historical volatility measure, while the artificial neural network (ANN) was calibrated using a 5-, 20-, 60-, 100-day moving average historical volatility measures together with a GARCH-based volatility measure

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.

While the Black–Scholes and Bachelier models perform best using the 100-day moving average historical volatility measure, even using this measure, they fall far short of the performance exhibited by the ANN. 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 studied. The results achieved with the ANN are noteworthy considering that a single ANN was used for all maturity and moneyness level 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 4 depicts the Diebold–Mariano tests, as captured by the test statistics for each pair of models compared and the associated *p*-values attached to such tests. The tests allow us to conclude that the ANN produced statistically superior forecasts relative to each of the parameterized models at the 1% significance level, while there was no significant difference in predictive accuracy when comparing the Bachelier and the Black–Scholes model at the 5% significance level.

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

The conclusions on the lack of superiority of the Black–Scholes model relative to the Bachelier model and the superiority of the ANN model relative to the Bachelier model are the novel contributions of this study. These show that, while the Bachelier model allows for the use of negative strike prices on the underlying, the user of such a model is not better off than if using the Black–Scholes model outside this niche case. However, the choice between the parameterized models becomes largely irrelevant, as we find that the ANN constructed in this study to value European call options on the S&P 500 index to be far superior to both other models. 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.

Conclusion

In conclusion, motivated by renewed interest in the Bachelier model following its brief introduction by the CME Group, Inc., (2020), as a result of the COVID-19 pandemic, this study compared the predictive option pricing abilities of the Bachelier and Black–Scholes models, as well as an optimized ANN model, on a data set of 226,170 European call options on the S&P 500 index between 1 March 2005 and 31 December 2019.

The study examined all call options available during the period for various maturities (ranging from one month to two years) and moneyness levels (ranging from 80 to 120%), using approximately 20% of the dataset to test each of the three models considered. Using five different volatility measures combined with three forecasting error estimation measures, the 100-day moving average historical volatility measure was found to be superior for both the Black–Scholes and the Bachelier models. However, it was found that the ANN used in this study, with data split between the training set, validation set, and testing set according to a 60:20:20 ratio, when evaluated across three forecast error measures, performed remarkably well compared to both the Black–Scholes and Bachelier models. Further, for the pricing of European call options on the S&P 500 index over the period examined, the Diebold–Mariano test confirmed that there is no statistical difference in the predictive accuracy between the Black–Scholes and Bachelier models at the 5% significance level, but that the ANN model is statistically superior to both at the 1% significance level.

Therefore, this study shows that, while the brief re-introduction of the Bachelier model to the markets was novel, at least with regards to pricing European call options on the S&P 500 index, this model would have yielded statistically comparable pricing to the standard Black–Scholes model and, hence, would not have provided any obviously better outcomes. However, this study also shows that the optimized ANN developed by Vach (2019) is markedly superior to both other models in the context in which it was tested within this study, and, therefore, is worth considering by option traders in the pricing of call options, potentially also beyond only those on the S&P 500 index.

The major limitation of this study is related data availability for options on the S&P 500 index, specifically in terms of moneyness levels and times to maturity, which limited the scope of the study. Further research into the effects of different autoregressive models, such as EGARCH and FIGARCH, when predicting volatility values, as well as the use of the Harvey et al. (1997) test, may 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.

Funding Open access funding provided by University of the Witwatersrand. This research did not receive any specific funding from any funding agency in the public, commercial or not-for-profit sectors.

Declarations

Conflict of interests The authors declare that they are not aware of any competing interests that exist with regards to this research.

References

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

Bakshi, G., C. Cao, and Z. Chen. 1997. Empirical performance of alternative option pricing models. *Journal of Finance* 52 (5): 2003–2049.

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

Bennell, J., and C. Sutcliffe. 2004. Black–Scholes versus artificial neural networks in pricing FTSE 100 options. *Intelligent Systems in Accounting, Finance and Management* 12: 243–260.

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

Choi, K., M. Kwak, C. Tee, and Y. Wang. 2022. A Black–Scholes user's guide to the Bachelier model. *The Journal of Futures Markets* 42 (5): 959–980.

CME Group Inc. 2020. *Switch to bachelier options pricing model - effective April 22, 2020*. Chicago: CME Group Inc.

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

Diebold, F.X., and R.S. Mariano. 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13 (3): 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): 157–168.

Figlewski, S. 1997. Forecasting volatility. *Financial Markets, Institutions & Instruments* 6 (1): 1–88.
Garcia, R., and R. Gençay. 2000. Pricing and hedging derivative securities with neural networks and a homogeneity hint. *Journal of Econometrics* 94: 93–115.
Geigle, D.S. 2000. *An artificial neural network approach to the valuation of options and forecasting volatility*. Fort Lauderdale, Florida: Nova Southeastern University.
Hahn, J.T. 2013. *Option pricing using artificial neural networks: an Australian perspective*. Robina, Australia: Bond University.
Harvey, D., S. Leybourne, and P. Newbold. 1997. Testing the equality of prediction mean squared errors. *International Journal of Forecasting* 13: 281–291.
Heston, S.L., and S. Nandi. 2000. A closed-form GARCH option valuation model. *The Review of Financial Studies* 13 (3): 585–625.
Hutchinson, J.M., A.W. Lo, and T. Poggio. 1994. A Nonparametric approach to pricing and hedging derivative securities via learning networks. *Journal of Finance* 49 (3): 851–889.
Long, J.B. 1974. Session Topic: risk, information and capital budgeting: discussion. *Journal of Finance* 29 (2): 485–488.
Malliaris, M., and L. Salchenberger. 1993. A Neural network model for estimating option prices. *Journal of Applied Intelligence* 3: 193–206.
Nyarko, N.A., B. Divelgama, J. Gnawali, B. Omotade, S.T. Rachev, and P. Yegon. 2023. Exploring dynamic asset pricing within Bachelier's market model. *Journal of Risk and Financial Management* 16: 1–18.
Ruf, J. & Wang, W., 2020. Neural networks for option pricing and hedging: A literature review. *Journal of Computational Finance,* 24 (1): pp. 1–46.
Schachermayer, W., and J. Teichmann. 2008. How close are the option pricing formulas of Bachelier and Black–Merton–Scholes? *Mathematical Finance* 18 (1): 155–170.
Smith, C.W. 1976. Option pricing: a review. *Journal of Financial Economics* 3: 3–51.
Vach, D. 2019. *Artificial neural networks in option pricing*. Prague: Charles University.
World Federation of Exchanges, 2022. FY 2022 Market Highlights. [Online] Available at: https://www.worldexchanges.org/storage/app/media/FY%202022%20Market%20Highlights%20Report%20website.pdf
Yuan, G.X. 2020. *The CME vulnerability: the impact of negative oil futures trading*. Guangzhou: World Scientific Publishing.