A Comparative Study between Hierarchical Kernel Learning and Hybrid Models for Option Pricing

Ramasahayam Riddhima Reddy
Department of electronics and communications engineering
MGIT college
Hyderabad,Telangana,India

Abstract—This paper delves into the realm of option pricing, presenting a comparative exploration of two distinct methodologies. Specifically, it focuses on the application of Hierarchical Kernel Learning (HKL) with two kernel functions: the Polynomial Kernel and the ANOVA Kernel. Furthermore, the paper introduces a hybrid models that combine Expectation-Maximization (EM) and K-Means clustering techniques with Support Vector Regression (SVR) . A comprehensive comparative analysis of the hybrid model sheds light on their capacity to improve option pricing accuracy across diverse market conditions. The insights derived from this research provide valuable contributions to understanding the performance of different kernel functions within the HKL framework. Moreover, it underscores the potential advantages of employing hybrid models in capturing nuanced market dynamics, thus enhancing the accuracy of option pricing. Ultimately, this study advances the field of option pricing, serving the needs of financial professionals and researchers seeking advanced tools for the valuation of derivatives.

Keywords—options pricing; machine learning;hierarchical kernel learning;hybrid models

Introduction--

In tandem with the growing significance of derivatives in the realm of financial markets, a plethora of pricing techniques have been developed to address the essential need for accurately estimating the true value of financial instruments. Futures and options are extensively utilized by investors and traders to amplify their investments and seek substantial returns while also managing risk exposure. In the Indian financial markets, the National Stock Exchange (NSE) and Bombay Stock Exchange (BSE) stand as two major exchanges, with the NSE alone witnessing an average monthly turnover of approximately Rs. 7.5 trillion in futures and options. Consequently, the precise pricing of derivative products becomes of paramount importance due to the sheer volume of transactions and the significant monetary involvement.

The seminal work in option pricing was pioneered by Black and Scholes, who introduced the renowned Black-Scholes formula for option pricing. This model and its variants have served as stalwarts in producing reasonably fair values for options for over three decades. However, as time passed, researchers endeavoured to refine the model by challenging its underlying assumptions. Although various modifications and alternative models were proposed, none could wholly replicate the behaviour of actual option prices. These models assert that option prices depend on five key variables: the value of the underlying asset (S), the standard deviation of its expected returns (σ), the exercise price of the option (K), the time until option maturity (T), and the interest rate on the default-free bond (r). The relationship between option prices and these five variables is a complex and nonlinear one.

Financial markets exhibit intricate and stochastic behaviour, resulting in multivariate and highly nonlinear option pricing functions. Parametric models such as Black-Scholes model describe stationary nonlinear relationships between theoretical option prices and various variables. There is also evidence suggesting that market participants alter their option pricing attitudes over time. Parametric Option Pricing Models (OPMs) may fall short in adapting to these swiftly changing market dynamics. Efforts are ongoing to develop nonparametric techniques capable of surmounting the limitations of parametric OPMs. Furthermore, there is a persistent need among market participants for more accurate OPMs applicable in real-world scenarios.

To address these challenges, machine learning techniques such as Support Vector Regression (SVR), Hierarchical Kernel Learning (HKL), emerge as potent nonparametric data-driven approaches in empirical options pricing research. Support Vector Regression stands out as a powerful methodology for approximating complex functions without the need for a priori determination of model complexity, as seen in other nonparametric regression techniques. HKL is particularly valuable for non-linear variable selection.

.

Background--

Options are financial derivatives that provide investors and traders with the right, but not the obligation, to buy (call option) or sell (put option) an underlying asset, such as stocks, commodities, or currencies, at a predetermined price (strike price) on or before a specific date (expiration date). They play a crucial role in modern financial markets, enabling participants to hedge risk, speculate on price movements, and enhance portfolio strategies.

Options pricing relies on factors such as the current price of the underlying asset, the option's strike price, time until expiration, volatility, and interest rates. The Black-Scholes model, developed in the early 1970s, was a groundbreaking development in options pricing. It established the theoretical framework for calculating option prices and introduced the concept of implied volatility.

Options are actively traded in various financial markets, including options exchanges, and have a wide range of applications. Traders use them to speculate on price movements, while investors use them for risk management and income generation. The versatility and flexibility of options have made them an essential tool in modern finance, providing a means to navigate the complexities of financial markets and optimize investment strategies.

Literature Review

The conventional option-pricing models find their roots in the pioneering work of Black and Scholes in 1973. Their model marked a significant milestone as the first comprehensive option pricing model with all parameters measurable. However, this model and its variations have been subject to systematic biases, as reported by numerous researchers. For instance, it has been observed that Implied Volatility derived from the Black-Scholes model, as a function of the moneyness ratio (S/X) and time to expiration (T), often exhibits a U-shaped pattern, commonly referred to as the volatility smile. Additionally, empirical studies have shown that implicit stock return distributions exhibit negative skewness and greater excess kurtosis than accounted for in the Black-Scholes lognormal distribution. The Black-Scholes model assumes continuous diffusion of the underlying asset, a normal distribution of returns, constant standard deviation (volatility), and no impact on option prices from supply and demand, all of which are assumptions frequently challenged in real-world scenarios.

The model is based on the assumption that the underlying asset follows a geometric Brownian motion described by the stochastic differential equation:

$$dS = \mu dt + \sigma S dW \quad (1)$$

Where W represents Brownian motion and dW is the uncertainty in the stock price.

Applying Ito's Lemma and the no-arbitrage condition yields the second-order Black-Scholes partial differential equation (PDE):

$$\partial V/\partial t + (1/2)\sigma^2 S^2 \partial^2 V/\partial S^2 + rS\partial V/\partial S - rV = 0 \quad (2)$$

The Black-Scholes formula is derived by solving this PDE. It expresses the value of a call option on a non-dividend-paying underlying stock as:

$$C(S, t) = SN(d1) - Xe^{(-r(T-t))}N(d2) \quad (3)$$

Where:

- C(S, t): Premium paid for the European call option
- S: Spot price of the underlying asset
- X: Exercise price of the call option

- r: Continuously compounded risk-free interest rate
- T-t: Time remaining until the option's expiration date
- σ^2: Yearly variance rate of return for the underlying asset
- N(.): Standard normal cumulative distribution

The original model is founded on assumptions that do not always hold in the real world. Various extensions and modifications have been proposed to relax these constraints.

Model description:

Hierarchical models: This approach is designed to enhance the accuracy and precision of option pricing models by incorporating the hierarchical decomposition of the kernels.

Hierarchical Kernel Learning (HKL)

Hierarchical Kernel Learning (HKL) is a machine learning framework that extends the concept of Multiple Kernel Learning (MKL) by introducing a hierarchical structure to the selection and combination of multiple kernels. It aims to find a suitable function that approximates complex, nonlinear relationships in data by leveraging a set of kernels and their convex combinations. The hierarchical structure allows for more flexibility and adaptability in modelling complex data patterns.MKL typically involves combining multiple basic kernels into a single composite kernel using a single optimization function. However, the exponential growth in the number of basic kernels with the dimension of the input space can make this approach intractable.

HKL addresses this challenge by organizing basic kernels into a hierarchical structure represented as a directed acyclic graph (DAG). In this hierarchy, certain kernels are considered ancestors, while others are descendants. Kernels are selected based on specific rules that ensure a kernel can only be chosen after all of its ancestor kernels have been selected. Additionally, subsets of kernels are chosen only after all their subsets have been selected. These rules guide the selection process and make it computationally feasible.

HKL is particularly valuable in scenarios where a large number of potential kernels and non-linear relationships need to be efficiently handled. It is often applied to tasks involving non-linear variable selection in machine learning. The hierarchical structure and selection rules help capture complex relationships between variables and data points while maintaining polynomial time complexity.

HKL with polynomial kernel: A polynomial kernel is a kernel function that calculates the similarity or inner product between data points by raising the dot product of the data points to a certain power. The general formula for a polynomial kernel between two data points, denoted as x and x', is given as:

k(x, x') = (γ * (x^T * x' + r))^d (4)

- γ (gamma) is a scaling factor that controls the influence of the polynomial term.
- r is a coefficient that can be added to the dot product.
- d is the degree of the polynomial, determining the maximal degree of polynomial terms considered.

In the context of HKL, the polynomial kernel used considers a specific form where xi and xi' are the data points being compared, and qj represents certain parameters for each dimension or feature. The full kernel, denoted as k(x, x'), is defined as a product (the ∏ symbol indicates multiplication) over all these parameters qj for each dimension.

HKL with ANOVA Kernel:

The ANOVA kernel is designed to capture interactions among subsets of features in the data. It is particularly useful when dealing with data where interactions between multiple features are essential for understanding the underlying relationships. ANOVA kernels are expressed as a sum of terms, with each term representing interactions between a specific subset of features. The kernel equation is typically defined as:

k(x, x') = ∑ [I ⊆ {1, 2, ..., p}] (γ_I * ∏ [i∈I] xi * xi') (5)

- I represents a subset of feature indices from 1 to p, where p is the total number of features.
- γ_I is a parameter that scales the contribution of each subset I to the kernel value.
- The product is taken over all dimensions i within the selected subset I

2.Clustering hybrid model:

a) EM & SVR model: This is a hybrid model uses EM algorithm described in to cluster the series into two cluster based on moneyness ratio(S/K) & time to maturity(T). SV regression is then applied to each of these clusters separately to determine best model parameters.

EM:The Expectation-Maximization (EM) algorithm is a statistical iterative method used to estimate model parameters when data is incomplete or contains hidden variables. Its primary goal is to find the Maximum Likelihood (ML) estimates that make the observed data most probable. The EM algorithm operates through cycles of two steps: the

Expectation (E-step) and the Maximization (M-step).

In the E-step, the algorithm estimates missing or hidden data given the observed data and the current model parameter estimates. This step derives conditional expectations to assess the most likely values for the unobserved data.

In the M-step, the likelihood function is maximized, assuming the missing data are known using values estimated in the E-step. This optimization process seeks the best model parameters that fit the observed data.

Support Vector Regression (SVR)

Support Vector Regression (SVR) is a robust non-parametric approach used in empirical options pricing and other regression problems. It operates based on the idea of optimizing a model that strikes a balance between fitting the data as closely as possible and allowing for deviations within a certain margin, denoted as ε (epsilon). This margin represents the level of tolerance for errors, and SVR aims to keep these errors within this margin.

Training Data: SVR works with training data consisting of pairs of input patterns (xi) and corresponding output values (yi). In the context of options pricing, this could represent data such as index call options and related econometric indicators.

ε-Insensitive Loss Function: SVR employs an ε-insensitive loss function. This loss function penalizes errors but only considers errors larger than ε. This is crucial in situations where you want to limit your risk and are willing to accept deviations within a specific margin (ε).

Dimensionality Independence: One of the advantages of SVR is that its capacity is not dependent on the dimensionality of the feature space. It is controlled by parameters that do not rely on the dimensionality of the feature space, which makes it suitable for high-dimensional data.

Generalization and Model Complexity: SVR seeks to optimize generalization bounds for regression problems. It handles function estimation effectively without requiring prior determination of model complexity, unlike some other non-parametric regression techniques.

Empirical Options Pricing: SVR has gained importance in options pricing research. It is particularly valuable because it can model both linear and non-linear relationships in the data. Additionally, SVR's statistical properties allow it to generalize well to unseen data.

Comparison with Least Squares SVR: In empirical research for pricing call options, SVR models, particularly ε-insensitive and Least Squares SVR, have been compared. It was found that Least Squares SVR outperformed ε-insensitive techniques for out-of-sample pricing performance.

In summary, Support Vector Regression (SVR) is a flexible and powerful approach for various regression tasks, including options pricing. It allows you to control the tolerance for errors, making it suitable for scenarios where risk management is crucial. Its capacity is independent of feature space dimensionality, and it optimizes generalization bounds, making it a valuable tool in empirical research and modeling complex relationships in financial data.

b) K-means clustering & SVR:

This is a hybrid model uses K-means algorithm described to cluster the series into two cluster based on moneyness ratio(S/K) & time to maturity(T). SV regression is then applied to each of these clusters separately to determine best model parameters.

K-means clustering: The K-means clustering method is a popular unsupervised machine learning technique used to partition a dataset into a specified number of clusters, denoted as "k." The goal is to create clusters that have the greatest possible distinction from each other. The K-means method starts with random clusters and proceeds by iteratively optimizing the cluster assignments and centroid positions. The key idea is to create clusters that minimize the internal variability by grouping similar data points together. Simultaneously, it maximizes the variability between clusters, ensuring that the clusters are as distinct as possible from each other.

Data: The analysis used data on S&P CNX Nifty index call options traded on National Stock Exchange (NSE) over the period from January 2010 to July 2020 .The options with 1-month to expiry are considered for the purpose of this study. For the purpose of this study the options data is classified into five series depending on the difference between the index price and strike price. The dataset was then scaled in between [0, 1]. The data points with 0 days to maturity were removed from the sample data.

To manage and analyse the data effectively, the dataset was categorized into five distinct series based on the difference between the index price and the strike price. These series include:

1. **Deep In-the-Money (DITM)**: Options where the index price is significantly higher than the strike price.

In-the-Money (ITM): Options where the index price is higher than the strike price.

At-the-Money (ATM): Options where the index price is approximately equal to the strike price.

Out-of-Money (OTM): Options where the index price is lower than the strike price.

Deep Out-of-Money (DOTM): Options where the index price is significantly lower than the strike price.

Methodology:

In this study, the methodology is designed to enhance the accuracy of option pricing through the application of Hierarchical Kernel Learning (HKL) and innovative hybrid models.

The kernel selection process involves the choice of two kernel functions, namely the Polynomial and ANOVA Kernels, offering increased flexibility in modelling option pricing. This diverse kernel selection enables the capture of non-linear relationships and the decomposition of kernels to effectively approximate complex option pricing processes.

HKL employs a hierarchical structure that integrates multiple kernels, providing a comprehensive approach to modelling. This hierarchical structure enhances the precision of option pricing.

The hybrid models introduced in this research incorporate Expectation-Maximization (EM), K-Means clustering, and Support Vector Regression (SVR) techniques to improve option pricing accuracy. Clustering is performed using EM and K-Means methods, segmenting financial series data into two clusters based on moneyness ratio (S/K) and time to maturity (T). Subsequently, SVR is applied independently to each cluster to determine the optimal model parameters for precise option pricing.

A crucial component of the methodology involves a comprehensive comparative analysis of these hybrid models. This analysis evaluates their performance and their potential to enhance option pricing accuracy across various market conditions.

Performance characteristics:

Performance measurement is a critical aspect of evaluating the accuracy of different models used in option pricing. In this study, several error measurement parameters are employed to quantify the pricing performance of these models, with a focus on the following key metrics:

Mean Squared Error (MSE): Mean Squared Error is a commonly used performance measurement criteria in various studies. It quantifies the dissimilarity between an estimator's predictions and the true values of the quantity being estimated. The MSE is calculated as the average of the squared differences between the observed values and the predicted values.

Explained Variance: Explained Variance, a statistical metric, assesses the extent to which a mathematical model accounts for the variation or apparent randomness within a given dataset. It measures how well the model explains the variance in the observed data relative to the variance in the true values.

Results:

model	DITM	ITM	ATM	OTM	DOTM
standard BS	24.6%	8.6%	25.5%	3%	6.4%
HKL-poly	41%	22.3%	34.7%	28%	26.1%
HKL-ANOVA	47.2%	38.5%	41.8%	34%	40%
EM SVR clustering model	58.4%	49.6%	63.5%	51%	58.8%
KM SVR clustering model	56.9%	46.4%	73.4%	47.8%	53.5%

HKL Models: The HKL models (HKL-poly and HKL-ANOVA) show significant improvements over the standard Black-Scholes model (BS) in various categories, especially in ATM and DOTM.

Clustering Models: The clustering models (EM SVR clustering model and KM SVR clustering model) perform even better, particularly in ATM and DOTM, where they show substantial improvements over both HKL models and the standard Black-Scholes model.

ATM and DOTM: Both the EM and KM clustering models provide the highest improvements in the ATM and DOTM categories, demonstrating the effectiveness of clustering techniques in enhancing option pricing accuracy.

Conclusion:

The overall conclusion of the analysis is that the hybrid models, particularly those combining clustering techniques (EM and KM) with Support Vector Regression (SVR), provide substantial improvements in the accuracy of option pricing. These models outperform both the standard Black-Scholes model and the Hierarchical Kernel Learning (HKL) models in various categories of options, with the most significant enhancements observed for At-The-Money (ATM) and Deep Out-Of-The-Money (DOTM) options.

The results indicate that traditional models, such as the Black-Scholes model, may not adequately capture the complex dynamics of financial markets, especially in the case of options that are ATM or DOTM. This limitation highlights the need for more advanced and adaptive modelling techniques.

By incorporating clustering methods, these hybrid models demonstrate their effectiveness in segmenting financial series data into distinct clusters based on moneyness ratio and time to maturity. These clusters provide a more nuanced understanding of market dynamics, allowing the SVR component of the hybrid model to adapt more accurately to different market conditions.

In summary, the study underscores the importance of considering advanced machine learning techniques, specifically hybrid models that leverage clustering and SVR, to enhance option pricing accuracy, especially for options that are ATM or DOTM. This research contributes valuable insights for financial professionals and researchers seeking improved tools for derivative valuation in dynamic and evolving market conditions.

REFERENCES

1)Park, H., Kim, N., & Lee, J. (2014). Parametric models and non-parametric machine

2)learning models for predicting option prices: Empirical comparison study over KOSPI

3) Zeynep lt¨uzer Samur and G¨ul Tekin Temur. The use of artificial neural network in option pricing: The case of s&p 100 index options. World Academy of Science,Engineering and Technology, 2009.

4) Pawel Radzikowski. Non-parametric methods of option pricing. INFORMS & KO-RMS, 2000.

5) Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector ma-chines. ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27, 2011.

6)] Panayiotis Ch. Andreou, Chris Charalambous, and Spiros H. Martzoukos. Criticalassessment of option pricing methods using artificial neural networks. Springer-Verlag Berlin Heidelberg, 2002.

7) Panayiotis C. Andreou, Chris Charalambous, and Spiros H. Martzoukos. Pricing and trading european options by combining artificial neural networks and parametricmodels with implied parameters. European Journal of Operational Research, 2008