EMERGING MARKETS FINANCE AND TRADE
https://doi.org/10.1080/1540496X.2019.1695598

Routledge
Taylor & Francis Group

An Intelligent Learning and Ensembling Framework for Predicting Option Prices

Xiangyu Wei, Zhilong Xie, Rui Cheng, Di Zhang, and Qing Li

Financial Intelligence and Financial Engineering Research Key Laboratory of Sichuan Province, School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu, Sichuan, China

ABSTRACT
Estimating option prices and implied volatilities are critical for option risk management and trading. Common strategies in previous studies have relied on parametric models, including the stochastic volatility model (SV), jump-diffusion model (JD), and Black-Scholes model (BS). However, these models are built on several strict and idealistic assumptions, including lognormality and sample-path continuity. In addition, previous studies on option pricing mainly relied on its own market-level indicators without considering the effect of other concurrent options. To address these challenges, we propose an intelligent learning and ensembling framework based on convolutional neural network (CNN). Specifically, the customized nonparametric learning approach is first utilized to estimate option prices. Second, several traditional parametric models are also applied to estimate these prices. The estimated prices are combined by a CNN to obtain the final estimations. Our experiments based on Chinese SSE 50 ETF options demonstrate that the proposed intelligent framework outperforms the traditional SV model, JD model, and BS model with at least 41.52% performance enhancement in terms of RMSE.

KEYWORDS
CNN; ensembling; implied volatility; option pricing

1. Introduction

The option is one of the most important financial derivatives with versatile functions. Specifically, it is frequently applied to speculate or to hedge current holdings. It is also utilized to generate income through option writing. The earliest options exchange can be traced back to the establishment of the Chicago Board Options Exchange in 1973. In 2017, the notional principal of the annual turnover of global exchange-traded options reached up to $619,040 billion. It is of great necessity to estimate the implied volatility (IV) and the price of an option for risk management and option transaction.

The earliest option pricing model is the Black-Scholes model (BS) (Black and Scholes 1973). It assumes that the underlying returns follow a geometric Brownian

CONTACT Qing Li liq_t@swufe.edu.cn Financial Intelligence and Financial Engineering Sichuan Key Laboratory and the School of Information, Southwestern University of Finance and Economics, Chengdu, Sichuan 610072, China

motion (GBM), and the volatility is constant over time. Unfortunately, these assumptions are too idealistic to be achieved in practice. Later, many polished parametric models, including jump-diffusion (JD) (Merton 1976) and stochastic volatility (SV) (Heston 1993), were proposed. However, these models were still built on some strict and unrealistic assumptions, including path continuity and no-arbitrage conditions. Such simplification is too naive and unrealistic to capture the complicated and volatile options markets in the real world (Hutchinson, Lo, and Poggio 1994; Park, Kim, and Lee 2014). In addition, previous studies estimate an option's price mainly relied on the market-level indicators of this specific option without considering the performance of other concurrent options.

To address these challenges, we proposed an intelligent learning and ensembling framework based on convolutional neural network (CNN). In particular, the customized nonparametric learning approach is first utilized to estimate option prices. Second, several traditional parametric models are also applied to estimate these prices. The estimated prices are combined by a CNN to obtain the final estimations. This framework has several unique contributions, as follows:

- The customized nonparametric learning approach is free of the unrealistic simplifications assumed in the traditional parametric models.
- Our approach can consider the influence of concurrent options to estimate the implied volatility and the price of a target option.
- Our approach utilizes the power of both parametric and nonparametric models for option pricing via ensembling techniques.

Our experiments based on Chinese SSE 50 ETF options demonstrate that the proposed intelligent framework outperforms the traditional BS model, jump-diffusion model, and SV model with at least 41.52% performance enhancement in terms of root mean square error (RMSE).

2. Related Work

In this study, we focus on estimating the implied volatilities and prices of options. Previous studies can be roughly divided into two types: traditional parametric models and nonparametric models, which rely on modern machine learning approaches (Park, Kim, and Lee 2014; Rubinstein 1994).

2.1. *Traditional Parametric Models*

2.1.1. *Parametric Models for Implied Volatility*

Volatility is an important risk indicator for financial assets. Implied volatility (IV) is the estimated volatility given a security's price and is most commonly used when pricing options. High volatility means a large price swing, but the price could swing very high, very low, or both. Low volatility means that the price likely will not make broad, unpredictable changes. When the price is known, the implied volatility can be derived from the option pricing model because it is the only factor in the model that cannot be directly observed in the market. However, in most cases, the price is unknown, and the implied volatility can only be determined by other factors that can be observed from the market, including the strike price and the time to maturity.

The determined volatility function (DVF) is the most popular method for estimating IV based on option strikes and time to maturity (Derman, Kani, and Chriss 1996; Dupire 1994; Rubinstein 1994). However, the DVF focuses on linear patterns, which are too simple to capture complicated implied volatility patterns. Later, several more complicated parametric models were proposed to estimate IV (Christoffersen and Jacobs 2004). It is still hard for these models to adequately capture the dynamics of implied volatilities (Homescu 2011).

2.1.2. *Parametric Models for Option Pricing*

The earliest work on option pricing can be traced to the Black-Sholes (BS) model (Black and Scholes 1973). BS assumes volatility remains constant over the option's life, which is not the case because volatility fluctuates with the level of supply and demand (Jackwerth and Rubinstein 1996; Madan, Carr, and Chang 1998).

One alternative solution is to estimate the volatility based on the real options market situations to replace the constant volatility. A good example is the ad hoc BS (AHBS) model, which estimates implied volatilities using the DVF (Andreou, Charalambous, and Martzoukos 2014; Kim and Kim 2004; Kim and Lee 2014). Specifically, the traditional BS model estimates option prices in terms of the underlying price S, the strike price K, the risk-free rate r, the time to maturity τ, and the constant volatility σ. The AHBS model uses the DVF function to replace the constant volatility σ with the estimated volatility σ_{IV}. Several studies have further demonstrated the effectiveness of the AHBS model (Kim 2010; Singh and Dixit 2016).

Both the BS and AHBS models rely on the GBM assumption. That is, the price movement is a continuous-time stochastic process in which the logarithm of the randomly varying quantity follows a Brownian motion (also called a Wiener process) with drift. However, the GBM assumption does not capture extreme movements, such as stock market crashes (Heston 1993; Merton 1976). Some researchers have released such constraints. The jump-diffusion (JD) model and the stochastic volatility (SV) model are two good examples of this approach. In particular, the JD model noted that the underestimation of extreme moves yields tail risk. Therefore, the JD model assumes that the movement of underlying assets follows a stochastic process with jumps to Brownian motion (Merton 1976). Similarly, the SV model realized that the assumption of a stationary process yields volatility risk, which can be hedged with volatility hedging. SV assumes that volatility follows a random diffusion process (Heston 1993). Later, many models were proposed to release the strict assumptions of the traditional BS model, including the variance gamma model (Madan, Carr, and Chang 1998; Madan and Milne 1991) and the generalized autoregressive conditional heteroskedasticity (GARCH) pricing models (Heston and Nandi 2000).

Although many efforts have been devoted to releasing market constraints on these parametric models, they are still restricted by the assumptions of frictionless market and risk-neutrality theories that are difficult to achieve in practice (Jackwerth and Rubinstein 1996). Although the AHBS model has proven efficient in many cases (Christoffersen and Jacobs 2004; Jackwerth and Rubinstein 1996; Kim 2010; Li and Pearson 2007), some researchers found that the AHBS model fails when using different markets and experimental datasets (Kim and Kim 2004; Singh and Dixit 2016). The strict assumptions can partially explain the inconsistent performance of parametric models in practice.

2.2. *Machine Learning Based Nonparametric Models*

With the rapid development of information technology (IT), researchers have introduced machine learning techniques for estimating implied volatilities and option prices. Such nonparametric models are essentially data-driven approaches that are free of the aforementioned parametric model constraints.

2.2.1. *Nonparametric Models for Implied Volatility*

Realizing the limitations of parametric models, some researchers have utilized modern machine learning techniques to estimate the implied volatilities. One of the pilot studies utilizing machine learning techniques to estimate option volatilities was the work of Malliaris and Salchenberger (1996). It applied neural networks (NN) to forecast S&P100 implied volatilities. Ahn et al. (2012) applied NN to measure KOSPI 200 (Korean index) options under Greek inputs and achieved promising predictive performance. Mostafa, Dillon, and Chang (2015) demonstrated NN capabilities in predicting option volatilities and utilized the estimated volatilities to further predict option prices with the BS model.

In addition to NN, many other traditional machine learning techniques have been utilized. Audrino and Colangelo (2010) implemented regression trees to forecast implied volatilities and conduct an empirical study on S&P500 index options. Zeng and Klabjan (2019) designed an online adaptive primal support vector regression (SVR) to predict volatility, and experiments on E-mini S&P 500 options show its efficiency.

Table 1 shows the details of the relevant studies applying machine learning techniques to predict implied volatilities in terms of features, markets, time intervals, and machine learning methods. Note that, the target of all of these machine learning methods is the implied volatility σ_{IV}.

Table 1. Research on the performance of IV forecasting by machine learning methods.

Literature	Features	Markets	Time intervals	Methods
Malliaris and Salchenberger (1996)	13 variables[(a)]	S&P100	1992.1–1992.12	BP-NN
Ahn et al. (2012)	Option Greeks	KOSPI 200	2003–2004	ANN(logistic)
Audrino and Colangelo (2010)	$S/K, \tau$	S&P500	1996.1–2003.8	Regression trees
Mostafa, Dillon, and Chang (2015)	$S/K, \tau, \sigma_{HV}$	London FTSE 100	2000.2–2001.12	MLP
Zeng and Klabjan (2019)	Tick data	E-mini S&P 500 [(b)]	2014.1.27–1.31	SVR (KPSVR, BKPSVR, EKPSVR),

NOTE: (a). Variables including (history, middle, distant) volatilities (put, call, market), price and the sum of some of the variables. (b). High frequency data were used.

2.2.2. *Nonparametric Models for Option Pricing*

Rather than predicting implied volatilities, some researchers have predicted option prices using nonparametric models. The earliest work can be traced back to Hutchinson, Lo, and Poggio (1994). In this study, several NN were applied, including multilayer perceptron (MLP), projection pursuit regression (PPR) and radial basis function (RBF) networks, to estimate option prices, and an empirical test on S&P 500 futures options demonstrated the superiority of NN, especially when traditional parametric models fail. Similar results were also observed by Yao, Li, and Tan (2000), who applied NN to Nikkei 225 futures options in Japan. Gençay and Qi (2001) utilized NN with Bayesian regulation, early stopping, and bagging mechanisms to study the S&P 500 index options and found that NN achieved better performance than traditional parametric models. Culkin and Das (2017) connected a number of feed-forward neural networks to estimate option prices. In addition, Yang, Zheng, and Hospedales (2017) proposed the gated neural network under economic rationality guarantees. Their empirical study on EU call options showed that the proposed model significantly outperformed some existing models.

Some researchers have integrated traditional parametric models with NN for better performance. For example, instead of directly taking the market observable variables as NN inputs, Liang et al. (2009) utilized the estimated option prices, which were calculated via traditional parametric models (binomial tree, finite difference method and BS models). The experiments on the Hong Kong market demonstrated the efficiency of this combined method. Wang (2009) estimated the volatilities calculated using different approaches (GARCH, GJR-GARCH and gray GJR-GARCH) as inputs of NN along with the observable fundamental market factors. Wang et al. (2012) used different volatility models, including historical volatility, DVF, GARCH, and GM-GARCH models.

Park, Kim, and Lee (2014) compared nonparametric models, including SVR and NN, to three classic parametric models, including BS, SV and JD models, and the experimental study revealed that nonparametric methods significantly outperform traditional parametric methods. Recently, an experimental study of Liu et al. (2019) suggested that other than traditional pricing models, NN also performs better than wavelet-based pricing models.

Table 2 shows the details of the relevant studies applying machine learning techniques to predict option prices in terms of features, markets, time intervals, and machine learning methods.

Traditional parametric methods are typically based on overly idealized assumptions, including no-arbitrage conditions, lognormality, or sample-path continuity. In addition, previous studies on estimating an option price mainly relied on the market-level indicators of this specific option without considering the influence of concurrent options (Cassese and Guidolin 2006). In fact, many options are traded at the same moment and affect each other simultaneously.

In this study, we propose an intelligent learning and ensembling framework based on CNN to estimate future option trends. Our experiments based on Chinese SSE 50 ETF options demonstrate the superiority of the proposed framework compared to traditional parametric methods.

The remainder of this article is organized as follows: Section 3 describes the structure of the proposed pricing framework, including the feature settings, the CNN structure and the ensemble learning methodologies. Section 4 presents the empirical results based on SSE 50 ETF options. Section 5 concludes with the main findings and speculation on future work.

Table 2. Research on the performance of option pricing by machine learning methods.

Literature	Features	Markets	Time intervals	Methods
Hutchinson, Lo, and Poggio (1994)	$S/K, \tau$	S&P500 future	1987.1–1991.12	MLP, PPR, and RBF
Yao, Li, and Tan (2000)	$S/K, \tau$	Nikkei225 Future	1995.1–1995.12	NN(tanh)
Gencay and Qi (2001)	$S/K, \tau$	S&P500	1992.1–1994.12	NN with Bayesian regulation, early stopping, and bagging
Liang et al. (2009)	C_{BT}, C_{FD}, C_{MC}	Hong Kong	2006.1–2007.12	LNN, MLP, and SVR
Wang (2009)	$S/K, \tau, r, \sigma_G$ [a]	TXO [b]	2005.1–2006-12	BP-NN(sigmoid)
Wang (2011)	$S/K, r_f, \sigma, r, \tau$ [c]	GBP/USD currency	2009.1–2009.7	SVR(RBF) and BP-NN
Wang et al. (2012)	F, K, τ, r, σ [d]	TXO	2008.1–2009.12	BP-NN (logistic, tanh)
Park, Kim, and Lee (2014)	F, K, τ, r, σ	KOSPI	2001.1–2010.12	GPR, SVR (EXP, RBF), ANN (sigmoid, tanh)
Yang, Zheng, and Hospedales (2017)	$K/S, \tau$	S&P500	1996.4.1–2016.5.31	Gated neural network
Liu et al. (2019)	$Se^{-\delta T}/K, T, r, \sigma$ [e]	DAX-30 index	2009.1–2012.12	Hybrid NN

NOTE: (a). σ_G is the volatility estimated from GARCH, GJR-GARCH, and Gray-GJR-GARCH models. (b). High frequency data were used. (c). r_f represents the foreign exchange interest rate. (d). Different volatilities, including historical volatility, implied volatility, volatility derived from the deterministic volatility function, GARCH and GM-GARCH models. (e). δ represents the dividend rate.

3. System Design of the CNN-based Framework

Figure 1 presents an overview of the proposed intelligent learning and ensembling framework. The customized nonparametric learning approach is first utilized to estimate implied volatilities and prices. Second, several traditional parametric models, including the JD model, SV model, and AHBS model, are also applied to estimate these prices. To utilize the power of both parametric and nonparametric models, the pre-estimated prices are mingled by a CNN to obtain the future option prices.

3.1. *Features for Predicting Implied Volatilities and Option Prices*

In practice, several options are traded at the same time, which may interact with the pricing of relevant options (Cassese and Guidolin 2006). Considering such real-time interference among different options is very important. In this study, we model options with matrices, as illustrated in Figure 2. In these matrices, row O_i represents the option with i^{th} strike, and column F_j represents the j^{th} input feature of an option, and $F_{i,j}$ is the j^{th} feature value for the i^{th} option strike. Table 3 lists all of these input features.

3.1.1. *Features of Implied Volatility Estimation*

Typically, implied volatility can be calculated by using the BS pricing model once we know the option prices (Kim and Kim 2004; Liu et al. 2019; Wang et al. 2012). However, option prices cannot be obtained in advance and applied to forecast future implied volatilities. Many predictive models forecast implied volatilities based on four key features, including the strike price K, the underlying price S, the risk-free rate r, and time to maturity τ (Yao, Li, and Tan 2000).

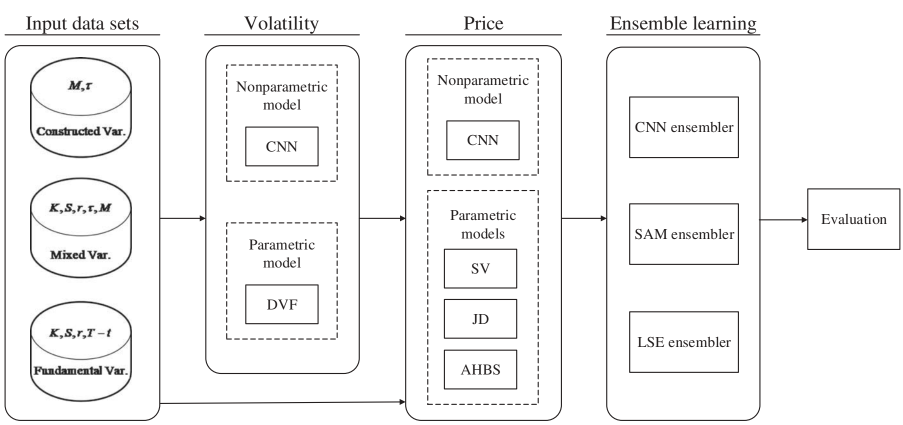

Figure 1. The proposed intelligent learning and ensembling framework.

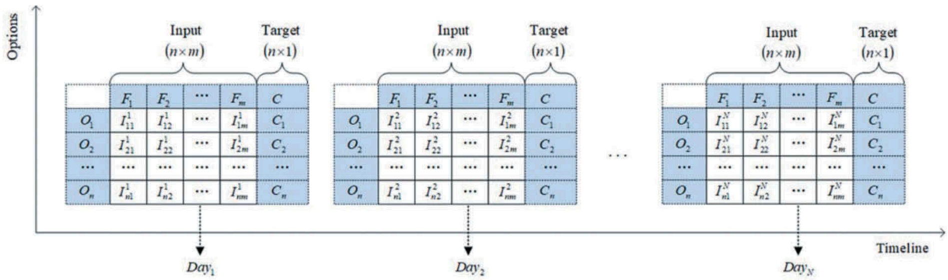

Figure 2. The option matrices.

Table 3. Related option indicators.

Features	Description	Output/Input
Price C	Prices for call options, unobservable.	Output
Implied volatility σ_{IV}	The volatility implied by the market price of the option based on an option pricing model.	Output/input
Strike price K	The price at which an option contract can be exercised.	Input
Underlying price S	The spot price of the underlying asset.	Input
Risk-free rate r	The rate earned on a riskless asset.	Input
Maturity T	The final payment date of options.	Input
Trading moment t	The transaction moment.	Input
Time to maturity τ	The time remaining until an option contract expires, which equals $T-t$.	Input
Moneyness M	A description relating strike price to the underlying price, which equals S/K.	Input

Note that options on the same matrix should be arranged in ascending or descending rules, as those with neighboring strikes may share more similarities. In addition, we estimate two indicators to track the option trends, implied volatility σ_{IV} and option price C. Each indicator is estimated by different feature combinations. The following two subsections explain the features applied to estimate the implied volatilities and option prices.

The most popular method for predicting implied volatility σ_{IV} is using a DVF (Dupire 1994; Rubinstein 1994). The DVF assumes that implied volatility is highly dependent on time to maturity τ and moneyness M, which is calculated as S/K. Specifically,

$$\sigma_{IV} = f_{DVF}(M, \tau) \tag{1}$$

where f_{DVF} is a linear function with binomial or trinomial forms. σ_{IV} is the dependent variable. Later, variants of DVF with different combinations of these four key features were proposed to further improve the predictive accuracy, including Dupire (1994); Rubinstein (1994); Kim and Kim (2004); Wang et al. (2012); Andreou, Charalambous, and Martzoukos (2014); Liu et al. (2019). In this study, we adopt three feature combinations in the proposed CNN predictive model:

- The fundamental set: it includes the classic four features $\{K, S, \tau, r\}$, which can be directly observed from options markets.
- The constructed set: it includes two features $\{M, \tau\}$, which are typically utilized in DVF.
- The combined set: it combines both the fundamental variables $\{K, S, \tau, r\}$ that can be directly observed from the market and the constructed indicator M indicating the option moneyness.

3.1.2. *Features of Option Pricing*

The BS model, which is widely used in both industry and academia, is the most classic approach to estimate option prices (Black and Scholes 1973). Specifically,

$$C_{BS} = f_{BS}(K, S, r, \tau, \sigma) \tag{2}$$

where C_{BS} is the call option price under the BS model, f_{BS} is the BS pricing formula, and σ is the constant representing the volatility of the underlying assets.

Realizing that the constant volatility assumption contradicts reality, some researchers have extended BS to AHBS (ad hoc Black-Scholes), which replaces the constant volatility σ with the estimated implied volatility σ_{IV} obtained by some predictive models including the DVF (Andreou, Charalambous, and Martzoukos 2014; Kim and Kim 2004; Kim and Lee 2014). Specifically,

$$C_{AHBS} = f_{BS}(K, S, r, \tau, \sigma_{IV}) \tag{3}$$

where C_{AHBS} is the call option price under the AHBS model and σ_{IV} is the estimated IV. AHBS releases the unrealistic assumption of constant volatility and achieves better performance (Kim 2010; Singh and Dixit 2016).

Similar to BS and AHBS, we applied two feature sets in the proposed framework to estimate option prices:

- BS-based set: it adopts the variables $\{K, S, \tau, r\}$ in the classic BS model.
- AHBS-based set: it adopts the variables $\{K, S, \tau, r, \sigma_{IV}\}$ in the AHBS model. Here, σ_{IV} can be estimated by DVF or CNNs.

3.2. *CNN-based Predictive Model to Estimate IV and Prices*

The fluctuation of an option is affected by the relevant options. A good example is SSE 50 ETF options, which has many options with different strike prices reflecting the different investors' expectations on the future trend of the SSE 50 Exchange-Traded Funds. The expectation on one option can interfere with the relevant options with close strikes, which is essentially an option comovement problem for option estimation. In addition, security comovement is translation invariance within different horizontal and vertical ranges (Zhang, Zohren, and Robert 2019). Here, the horizontal range means time, and the vertical range is the strike price, as shown in Figure 2. For the vertical range, the hidden relations between the options with strike prices of 2.5 and 2.55 can exist between the options with strike prices of 2.55 and 2.6 (An et al. 2014). Such relationships among neighboring options are also time series patterns. In other words, it can coexist during a certain period and disconnect with external factors. For instance, the patterns among deep out-of-money options may exist until the options are expired.

In our preliminary experiments, we predicted the implied volatilities and prices using RNN, LSTM and CNN. We found that the CNN-based framework has supreme performance. A good explanation is that the unique structure of CNN is able to consider the comovements of options traded at the same moment when estimating the implied volatilities and prices. In this study, we proposed a CNN-based predictive model to address the option comovement and translation invariance issues for option estimations. Figure 3 shows the CNN structure comprised of convolutional layers, pooling layers and dense layers.

The convolutional layer: The convolutional layer is the core building block of a CNN that allows the network to concentrate on low-level features in the first hidden layer and then assemble them into higher-level features in the next hidden layer. This hierarchical structure allows us to integrate the movements of neighboring options of one option to estimate its future prices. As shown in Figure 2, the option-feature matrix is the input feature map. The convolutional layer utilizes kernels (or filters), a matrix with K_w width and K_h height, to assemble features into a receptive field with a size of $K_h \times K_w$ to form the extracted feature maps, which is also known as the convolutional operation. As shown in Figure 3, kernels are applied to assemble adjacent options to form feature maps. Specifically, the neuron $x^l_{i,j}$ in row i, column j

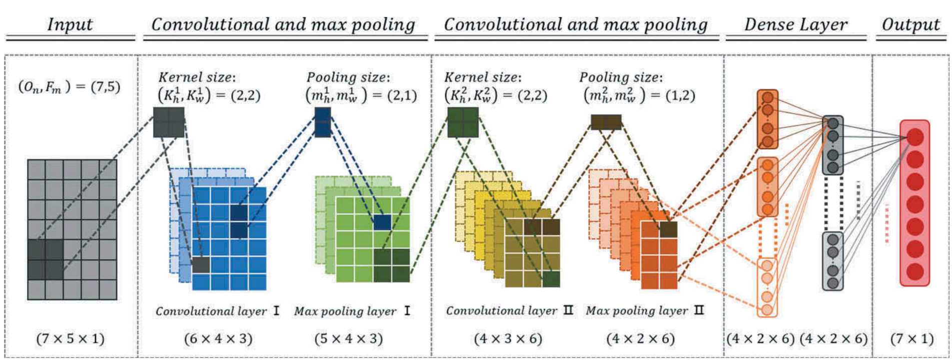

Figure 3. The proposed predictive model based on CNN.

of a given l^{th} layer is connected to the outputs of the neurons in the previous layer located in rows i to $i{+}K_h - 1$, columns j to $j + K_w - 1$.

$$x_{i,j}^{l} = b + \sum_{u=1}^{k_h}\sum_{v=1}^{k_w} x_{i',j'}^{l-1}\cdot w_{u,v} \tag{4}$$

where $x_{i,j}^{l-1}$ is the output of the neuron located in the previous $(l-1)^{th}$ layer, b is the bias, and $w_{u,v}$ is the connection weight in row u and column v of the receptive field. For any layer L_l with a size of $\left(L_h^l, L_w^l, 1\right)$, after being convoluted by a kernel with a size of (K_h, K_w), the layer L_{l+1} has a size of $\left(L_h^l - K_h + 1, L_w^l - K_w + 1, F^l\right)$ when the number of filters is equal to F^l under the condition that the stride is set to 1 (Ciresan et al. 2011). Here, the stride is the distance between two consecutive receptive fields. In this study, we set the stride to 1 to make full use of the features.

The max-pooling layer: A pooling layer follows a convolutional layer to reduce the feature dimension by downsampling the features extracted by the convolutional layers. The pooling operation not only reduces the complexity of the convolutional layers but also restrains the phenomenon of overfitting (Liu et al. 2017). Options traded at the same moment may share some common features; for instance, the underlying price and the risk-free rate are the same for those options traded at the same moment. Therefore, we use the max-pooling to select superior invariant features and improve generalization (Scherer, Müller, and Behnke 2010). Specifically, the max-pooling function calculates the superior invariant feature x^{l+1} by selecting the maximum from the feature map $x_{u,v}^{l}$ in the l^{th} layer with a size of $m_h \times m_w$, which is also the size of the max-pooling layer. The mechanisms can be correspondingly denoted by

$$x_{i,j}^{l+1} = \max\left(x_{u,v}^{l}\right), with\begin{cases} u \in [i, i + m_h - 1] \\ v \in [ssj, j + m_w - 1] \end{cases} \tag{5}$$

The dense layer: The previous operation obtains multiple feature matrices; we need to flatten these features into a vector to map the final output. The dense layer, also known as the fully connected layer, helps to achieve this goal and pairwisely connects the neurons between two adjacent layers (Cui and Fearn 2018). Two dense layers are used in the proposed CNN structure. The feature maps extracted from previous convolutional layers and max-pooling layers need to be flattened by a fully connected layer. The activation functions, including a rectified linear unit (ReLU), tanh, and sigmoid, can be used to deal with the nonlinearity reflection between previous features and the dense layer. The mechanisms can be denoted by,

$$X^{l+1} = f\left(X^l W\right) \tag{6}$$

where W is the weight matrix, X^l represents the feature maps in the lth layer, and f is the activation function. The commonly used activation functions including sigmoid, tanh and ReLU. These activations are also used in previous convolutional layers. Tanh activation is simply a linear mapping to the sigmoid activation, and ReLU is a non-linear activation.

3.3. *Ensemble Mechanism*

As aforementioned, traditional parametric models, including SV, JD and AHBS, are restricted by certain unrealistic hypotheses, such as GBM, path continuity and no-arbitrage conditions. Such models estimate an option's price mainly relied on the market-level indicators of this specific option without considering the effect of other concurrent options.

In contrast, the proposed nonparametric CNN model is free of strict and unrealistic assumptions. However, the performance of deep learning models is determined by the available training data. Due to the short history of Chinese options markets, we adopted the ensembling mechanism to take advantages of both parametric models and nonparametric models.

Intuitively, assume a complex question is asked to thousands of random people. In many cases, you will find that the aggregated answer is better than an expert's answer. Similar to this wisdom of the crowd, the ensemble mechanism aims at obtaining better prediction performances by strategically combining multiple learning models. It reduces the risk of selecting incorrect models by aggregating all candidate models (Chatterjee et al. 2015; Dietterich 2000; Qureshi et al. 2017). There are usually two steps to construct an ensemble model. The first step is to train several base learners, which can use the same or different algorithms, and the second step is to aggregate the predictions of these base learners with a meta-algorithm to obtain the final results (Qureshi et al. 2017; Ren, Zhang, and Suganthan 2016). The most popular meta-algorithms are the simple arithmetic mean (SAM) and the least square error weighting (LSE) means. Instead of these linear aggregations, some researchers have taken the further step by aggregating the base learners predictions via machine learning approaches, including support vector machines (SVMs) and neural networks (NNs) (Bo 2006; Wang and Wu 2012).

In this study, we propose a CNN-based ensembling mechanism to aggregate the estimations of both the nonparametric model (CNN) and parametric models (SV, JD, and AHBS). Due to the considerable difference between the nonparametric model and parametric models, simple linear meta-algorithms (SAM, LSE) cannot properly interpret the data and achieve satisfactory results (Section 4.6). Here, we propose a novel ensemble learning by using CNN as a meta learner to detect patterns hidden among the existing estimations. Figure 4 shows the ensemble learning mechanisms comprised of base learning progress and meta learning progress.

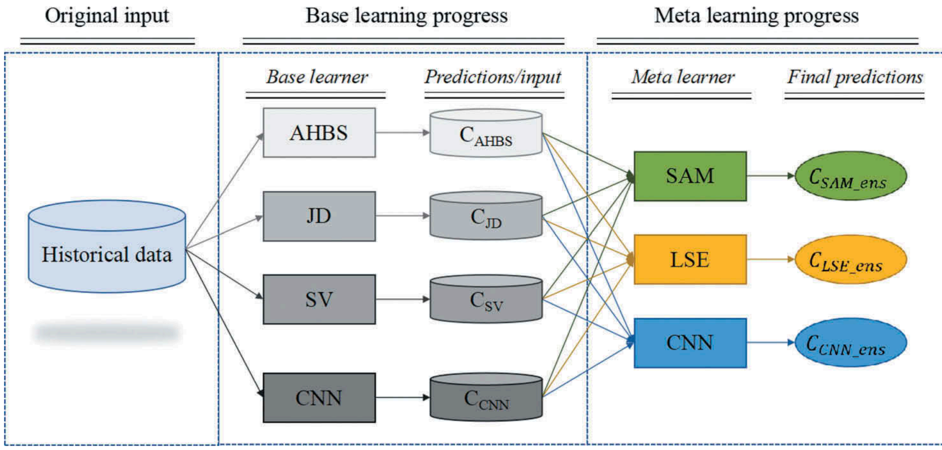

Figure 4. The ensemble learning mechanisms.

4. Empirical Study

4.1. *Experimental Data*

To gauge the performance, we examine the proposed framework with Shanghai Stock Exchange (SSE) 50 ETF options, which is the first stock option contract that was established on February 9th, 2015 on Chinese mainland. Call options with a maturity of Sept. 2016 were extracted from the Bloomberg database[1]. These samples include 1,770 observations with a maximum of 163 trading days. Table 4 summarizes the statistics of the implied volatilities and the prices of these samples.

In Table 4, ST, MT and LT represent the different time categories for the maturity of options; ITM, ATM, and OTM are the categories of option moneynesses.

Table 5 shows the detailed definitions based on the work of (Bakshi, Cao, and Chen 1997; Park, Kim, and Lee 2014; Rubinstein 1994). For a call option, out the money (OTM) means that the market price of the underlying assets is lower than the strike price, which may suffer from profit losses. At the money (ATM) is the condition in which the strike price of an option is equal to (or

Table 4. Statistics of the experimental data.

		Implied volatility		Option price			
Index		Mean	Std.	Mean	Std.	Num.	PCT.
Moneyness	ITM	0.1191	0.1803	0.2451	0.0935	990	55.93%
	ATM	0.1700	0.0533	0.0905	0.0484	417	23.56%
	OTM	0.1981	0.0723	0.047	0.0413	363	20.51%
Maturities	ST	0.1867	0.2282	0.1746	0.1540	526	29.72%
	MT	0.1143	0.0769	0.1569	0.1039	901	50.90%
	LT	0.1737	0.0843	0.1872	0.0657	343	19.38%
Total		0.1473	0.1450	0.1680	0.1163	1770	100.00%

Table 5. Time to maturity and moneyness.

Maturity	$\tau < 60$	$60 \leq \tau \leq 180$	$\tau > 180$
	ST (Short term)	MT (Medium term)	LT (Long term)
Moneyness	$M < 0.97$	$0.97 \leq M \leq 1.03$	$M > 1.03$
	OTM (Out the money)	ATM (At the money)	ITM (In the money)

nearly equal to) the market price of the underlying assets. If an investor is said to be ATM, they stand to suffer profit losses. For a call option, in the money (ITM) is the situation in which an option's strike price is below the market price of the underlying assets. ITM option has intrinsic value, and there is potential to achieve gains by exercising the option.

As shown in Table 4, compared with the ATM and OTM options, the ITM options have the largest portion, the lowest implied volatilities, and the highest prices. In addition, the medium-term options have the largest number, and the lowest implied volatilities and prices.

4.2. *Performance Metrics*

To gauge the performance of the proposed framework, the RMSE was adopted in this study. Specifically,

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(p_i - \hat{p}_i\right)^2} \tag{7}$$

where p_i is the real value of the option price C (or the implied volatility σ_{IV}) and $\hat{p}_i$ is the estimated call option price $\hat{C}$ (or the estimated implied volatility $\hat{\sigma}_{IV}$). N is the number of samples when it is used as a loss function in the training steps.

Equation (7) also acts as the measure of evaluation in prediction. Specifically, if N is the number of options on a certain day, the RMSE measures the daily predictions' RMSE, hence referred to as daily RMSE in later sections. If N is the number of options in a certain category (moneyness or maturity), the total RMSE can be used to measure the performance of the model in this specific category. These two kinds of estimators are used to evaluate the models' performances in the following subsections.

4.3. *Sliding Window*

Options prices dynamically change as the time to maturity varies. For example, options with a long time to maturity may have more stable characteristics compared to those with a short time to maturity. Therefore, we adopted a sliding window mechanism to train the predicted models. Figure 5 demonstrates the dynamic training mechanism.

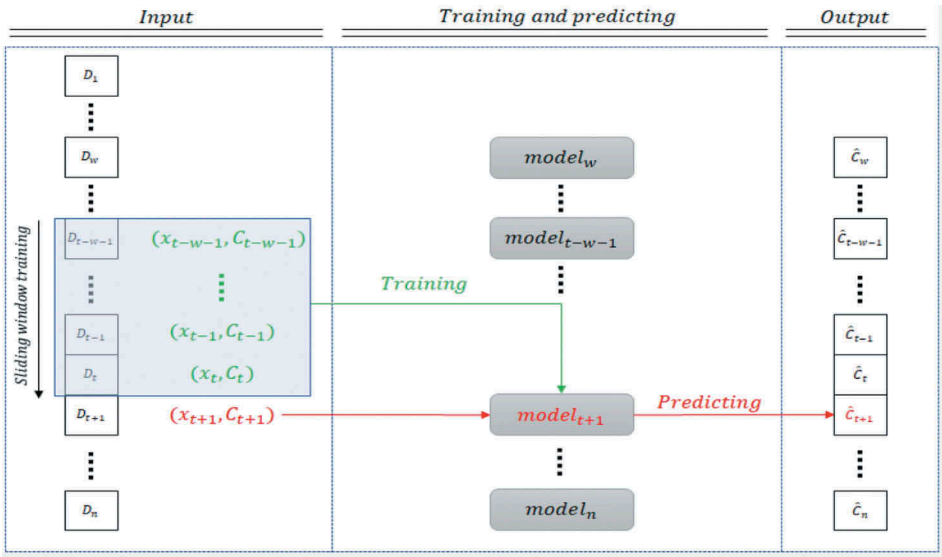

Figure 5. The dynamic training scheme.

In Figure 5, w is the sliding training window size, and the best-performing model m_{t+1} is obtained by training with data from D_{t-w-1} to D_t. In this study, we tested option pricing performances under 10 different sizes of training windows, from 5 to 50 days, with an interval equal to 5 days. The input features were based on the classic BS model, which is $\{K, S, \tau, r\}$, where σ is omitted because it is constant during the training scheme. Figure 6 shows the estimated results in terms of the average daily RMSE. In Figure 6, the RMSE under the 10-day-ahead training scheme obtained the best RMSE (0.018).

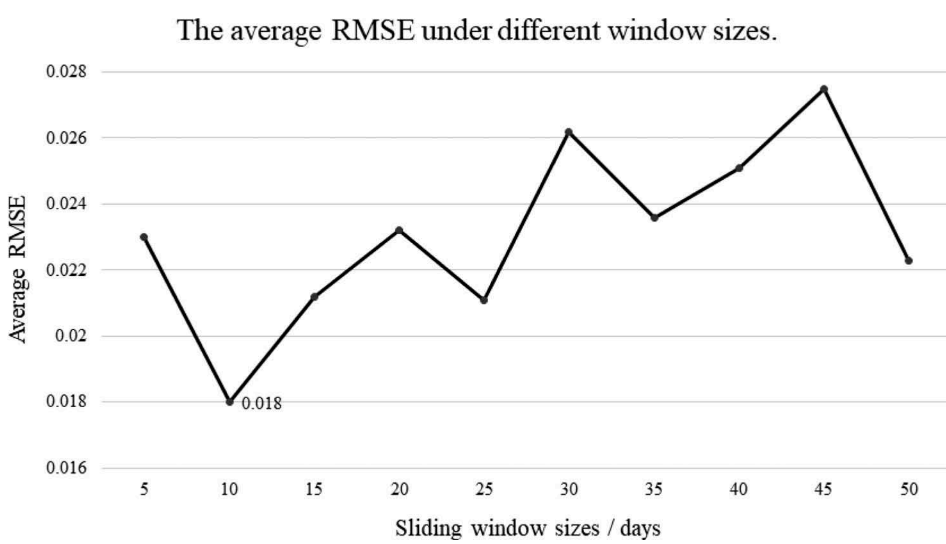

Figure 6. The average RMSE of option pricing under different window sizes.

One possible explanation for the results could be that the pattern may not last a long time in financial markets, especially for newly established instruments. That is, if the training window is set to a large window size, the learning may not capture the correct market trend because of noise information (Liang et al. 2009; Yang, Zheng, and Hospedales 2017). In this study, we carried out our empirical test based on a 10-day-ahead training scheme.

4.4. *Volatility Estimation*

In this section, we estimate the implied volatilities using the proposed CNN-based framework based on three different feature sets (Section 3.1.1). Table 6 shows the performance in terms of RMSE, where Cstd. represents the constructed feature input $\{M, \tau\}$, Fdmt. represents the fundamental features $\{K, S, \tau, r\}$, and Comb. is the combined set, which combines both the fundamental and the constructed variables, that is, $\{K, S, \tau, r, M\}$.

Table 6 consists of two panels. Table 6a shows the daily RMSE statistics of these 153 days' predictions. Table 6b analyzes the RMSE under different moneyness and maturity categories, along with the overall statistics of these 1703 option transactions. The following can be observed:

- The CNN models performed better than the traditional DVF model in terms of both daily and total RMSE. In particular, the three CNN models with different feature sets performed better than the DVF model in terms of the average daily RMSE over the 153 days (Table 6a). The CNN model with a combined feature set (Comb.) decreased from 0.1334 to 01287, with an enhancement of 3.55% in terms of the total RMSE of 1703 option transactions.
- From Table 6a, the CNN model with combined input feature set (Comb.) obtained the best performance, which shows that utilizing

Table 6. Volatility estimation.

		CNN for IV			Baseline	
Index		Cstd.	Fdmt.	Comb.	DVF	Days
(a) Statistics of daily RMSE for 153 days.						
Daily RMSE	Mean	0.0782	0.0763	0.0750	0.0813	153
	Min.	0.0148	0.0059	0.0081	0.0153	153
	Max.	0.9139	0.9957	0.9546	0.9881	153
	Std.	0.1042	0.1071	0.1058	0.1053	153
(b) RMSE under different moneyness and maturity categories.						
Moneyness	ITM	0.1657	0.1670	0.1644	0.1684	964
	ATM	0.0592	0.0552	**0.0511**	**0.0624**	393
	OTM	0.0549	0.0552	0.0566	0.0646	346
Maturity	ST	0.2173	0.2194	0.2167	0.2231	526
	MT	0.0574	0.0542	**0.0519**	**0.0588**	901
	LT	0.0630	0.0653	0.0625	0.0608	276
Total	RMSE	0.1302	0.1308	**0.1287**	**0.1334**	1703

more features or combining the feature (M) derived from the fundamental features $\{K, S, \tau, r\}$ can further improve the predictive performance. A possible explanation is that the convolutional mechanism of CNN can dynamically merge the fundamental features to obtain superior fused features along with the constructed feature M, which is derived from the fundamental features in a fixed economic principle.

- There are 1703 option transactions in the evaluation corpus, which can be classified into different categories: ST, MT, LT, ITM, ATM, and OTM. Here, the first three subcategories (ST, MT, LT) represent the maturity of the options, and the last three (ITM, ATM, OTM) represent the option moneyness, which is the ratio of strike price to the underlying price. For ATM and MT options, the CNNs achieved superior performances in forecasting the next day's implied volatilities compared with the traditional DVF model. In particular, for ATM options, CNN (with the combined feature input) reduced RMSE of DVF (0.0624) by 18.1% to 0.0511. For MT options, CNN (also with the combined feature input) reduced the RMSE of the DVF (0.0588) by 11.7% to 0.0519.

4.5. *Option Pricing with CNN*

Following implied volatility forecasting, in this section, we estimate option prices by using three traditional parametric models, SV, JD and AHBS, along with the CNN models based on two feature sets (BS-based set and AHBS-based set). In section 3.1.2, the BS-based feature set consists of $\{K, S, \tau, r\}$, and the AHBS-based set contains $\{K, S, \tau, r, \sigma_{IV}\}$. Here, feature σ_{IV} is calculated from the CNN model with the combined features in the previous section (4.4). Table 7 presents the pricing performances in terms of RMSE.

Table 7. Option pricing.

		CNN for prices		Baseline models			
Index		BS set	AHBS set	SV	JD	AHBS	Days
(a) Statistics of daily RMSE for 153 days.							
Daily RMSE	Mean	0.0180	0.0169	0.0411	0.0297	0.0354	153
	Min.	0.0030	0.0036	0.0030	0.0018	0.0020	153
	Max.	0.0820	0.0819	0.1075	0.1194	0.1542	153
	Std.	0.0134	0.0128	0.0228	0.0216	0.0355	153
(b) RMSE under different moneyness and maturity categories.							
Moneyness	ITM	0.0268	0.0248	0.0368	0.0314	0.0640	964
	ATM	0.0167	0.0161	0.0538	0.0425	0.0182	393
	OTM	0.0117	0.0119	0.0527	0.0346	0.0217	346
Maturity	ST	0.0229	0.0202	0.0182	0.0153	0.0363	526
	MT	0.0206	0.0195	0.0437	0.0349	0.0590	901
	LT	0.0262	0.0262	0.0740	0.0556	0.0388	276
Total	RMSE	0.0223	**0.0209**	0.0447	**0.0349**	0.0499	1703

Table 7 consists of two panels. Table 7a shows the daily RMSE statistics of these 153 days' predictions. Table 7b analyzes the RMSE under the moneyness and maturity categories, along with the statistics of these 1703 options.

The following can be observed:

- The CNN models outperformed the traditional parametric models (SV, JD, and AHBS) in terms of both daily and total RMSE. The JD model outperformed the other two parametric models, including the SV and AHBS models. However, both the CNN models with different feature sets were better than the best-performing JD model in terms of the average daily RMSE over the 153 days (Table 7). The CNN model with the AHBS-based input set decreased from 0.0349 to 0.0209, with an enhancement of 40.11% in terms of the total RMSE of 1703 option transactions.
- As shown in Table 7a, the CNN model with AHBS-based input feature set obtained better performance than the BS-based model, which suggests that utilizing the additional feature σ_{IV} calculated from the previous section (4.4) can further improve the predictive performance. Similar to the results in Table 6a, a possible explanation is that the derived feature σ_{IV} calculated from another CNN model represents the option risks, which can contribute to the evaluation. The convolutional mechanism of CNN is able to dynamically merge the input factors to obtain superior fused characteristics that lead to more precise results.
- Similar to the results in Table 6b, Table 7b shows the pricing performances in terms of categories, including maturity (ST, MT, LT) and moneyness (ITM, ATM, OTM) categories. In Table 7b, CNN shows significant superiority in terms of total RMSE and most moneyness and maturity categories. However, traditional parametric models work better than CNN for some option types. Specifically, the minimum daily RMSE of JD (0.0018) and AHBS (0.002) were smaller than those of the CNN model with AHBS-based input (0.0036). Furthermore, Table 7b implies that for short-term options, SV and JD outperform the CNN models. These phenomena suggest that it is very important to combine the advantages of both parametric and nonparametric models to further improve the pricing accuracy.

4.6. *Ensemble Learning with CNN*

By strategically combining multiple models, the ensembling methods can reduce the risk of selecting some incorrect models and obtain better prediction performances. The ensemble model typically consists of two steps: the first step is to train several base learners, and the second step is to aggregate the predictions of these base learners with a meta algorithm to obtain final results (Qureshi et al. 2017; Ren, Zhang, and Suganthan 2016). Here, the base learners include SV, JD, AHBS and CNN. Three meta algorithms are used: SAM, LSE and CNN ensembler.

Table 8 shows the pricing performances in terms of daily and total RMSE, where $\mathrm{CNN}_{\mathrm{ENS}}$ represents the ensembling model with CNN as the meta learner and $\mathrm{CNN}_{\mathrm{AHBS}}$ represents the base learning model with the AHBS-based set as input, corresponding to that in the previous section (4.5).

Similar to the previous sections (4.5 and 4.4), Table 8 also consists of two panels. Table 8a shows the statistics of daily RMSE of these 143 days' predictions after we use a 10-day-ahead training scheme. Table 8b analyzes the RMSE under moneyness and maturity categories, along with the total statistics of these 1626 samples.

Table 8 implies that due to the large difference between the nonparametric model and parametric models, simple linear meta algorithms (SAM, LSE) fail to further improve the performance of $\mathrm{CNN}_{\mathrm{AHBS}}$. However, $\mathrm{CNN}_{\mathrm{ENS}}$ can detect patterns hidden among the existing estimations and further improve the predictive capability. In particular, the RMSE of the $\mathrm{CNN}_{\mathrm{ENS}}$ ensembling model decreased from 0.0208 to 0.0193, with an enhancement of 7.21% in terms of the total RMSE of 1626 option transactions compared with the base learner $\mathrm{CNN}_{\mathrm{AHBS}}$. $\mathrm{CNN}_{\mathrm{ENS}}$ also reduced the RMSE of the best-performing parametric model JD (0.033) by 41.52% to 0.0193.

We also compare real option prices with two estimated prices with a strike price of 2.05 during the period between March 4, 2016 and September 28, 2016 as shown in Figure 7. The first estimated price is obtained by the JD model, which is the best-performing parametric model among BS, SV and JD. The second is based on the proposed CNN-based ensembling framework. Figure 8 demonstrates the differences between the estimated prices and the real prices. We apply the relative biases to illustrate such difference when comparing the proposed approach with the baseline (JD).

Table 8. Ensemble learning.

	Performances of ensemblers			Performances of base learners				
Index	$\mathrm{CNN}_{\mathrm{ENS}}$	SAM	LSE	$\mathrm{CNN}_{\mathrm{AHBS}}$	SV	JD	AHBS	Days
Mean	0.0158	0.0211	0.0167	0.0168	0.039	0.0279	0.0359	143
Min.	0.0028	0.0022	0.0022	0.0036	0.0030	0.0018	0.002	143
Max.	0.0723	0.0776	0.1126	0.0819	0.1075	0.1085	0.1542	143
Std.	0.0119	0.0128	0.0168	0.0127	0.0218	0.0198	0.0366	143
ITM	0.0229	0.0252	0.0279	0.0249	0.0351	0.0299	0.0648	932
ATM	0.0152	0.0244	0.0157	0.0155	0.0532	0.0406	0.0170	368
OTM	0.0102	0.0195	0.0163	0.0107	0.0501	0.0317	0.0215	326
ST	0.0182	0.0169	0.0243	0.0202	0.0182	0.0153	0.0363	526
MT	0.0180	0.0243	0.0220	0.0195	0.0437	0.0349	0.0590	901
LT	0.0266	0.0356	0.0275	0.0270	0.0748	0.0524	0.0414	199
Total	**0.0193**	0.0240	0.0235	**0.0208**	0.0430	**0.0330**	0.0506	1626

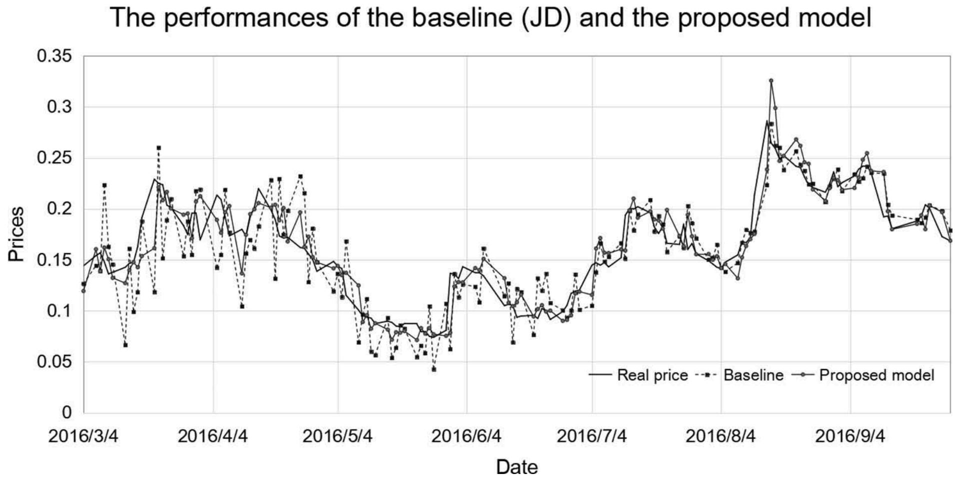

Figure 7. Pricing performances of the baseline model and the proposed model when strike price is equal to 2.05.

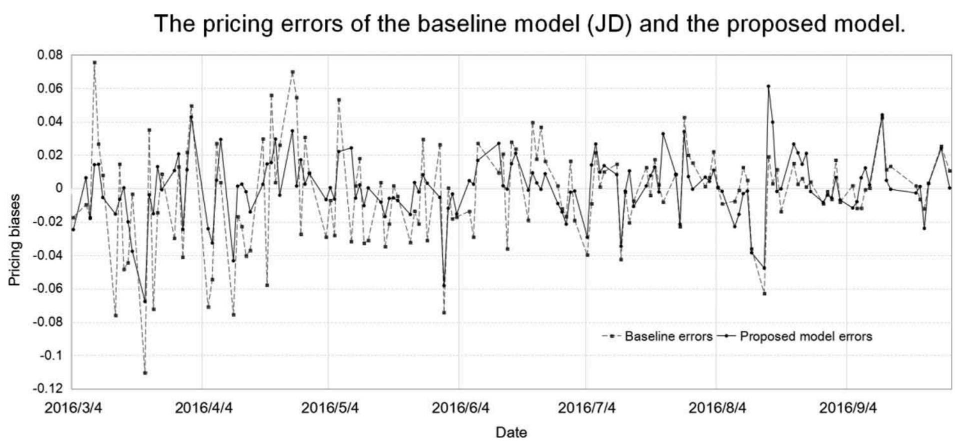

Figure 8. Pricing errors for the options when the strike price is equal to 2.05.

5. Conclusions

Previous studies on option pricing are mostly based on some strict and unrealistic assumptions, which are too limiting to capture the complicated and volatile options markets in the real world. In addition, these studies estimate an option's price mainly relied on the market-level indicators of this specific option without considering the comovements of other concurrent options. In this study, we proposed an intelligent learning and ensembling framework to address these challenges. This research was based on a nonparametric CNN model. We considered the comovements of options traded at the same moment when estimating the implied volatilities and prices. We also utilized the advantages of both parametric models and nonparametric CNN to further enhance the pricing accuracy.

Specifically, the customized nonparametric learning approach was first utilized to estimate option prices. Second, several traditional parametric models, including JD, SV and AHBS, were also applied to estimate these prices. To utilize the power of both parametric and nonparametric models, the pre-estimated prices were combined by a CNN to obtain the final estimations.

The customized nonparametric learning structure was tested with SSE 50 ETF options. The empirical result showed that, first, the CNN outperforms traditional parametric models in estimating implied volatilities and option prices. Specifically, the CNN model decreased from 0.0349 to 0.0209, with an enhancement of 40.11% in terms of the total RMSE of 1,703 option transactions compared with the best-performing JD model. Second, derived features, including moneyness M and implied volatility σ_{IV}, can further improve estimation accuracy because the convolutional mechanism of the CNN can dynamically merge the input features to obtain superior fused characteristics that lead to more precise results. Third, classic SAM and LSE become invalid when dealing with ensembling problems within nonparametric and parametric models, while the CNN acts as a rather qualified ensembler and can efficiently reduce prediction RMSE. In particular, the CNN_{ENS} ensembler reduced the RMSE of the base learner CNN_{AHBS} by 7.21%, and the final pricing results of the proposed framework reduced the RMSE by 41.52% compared with the traditional best-performing JD model.

Note

1. Data are obtained from Bloomberg terminals, the website is www.bloomberg.com.

Funding

This work has been supported by National Natural Science Foundation of China (NSFC) [71671141, 71873108], Fundamental Research Funds for the Central Universities [JBK 171113, JBK 170505, JBK 1806003, JBK 1902031, JBK 1907201876], Sichuan Province Science and Technology Department [2019YJ0250], and the Financial Innovation Center of the Southwestern University of Finance and Economics.

References

Ahn, J. J., D. H. Kim, K. J. Oh, and T. Y. Kim. 2012. Applying option Greeks to directional forecasting of implied volatility in the options market: An intelligent approach. *Expert Systems with Applications* 39 (10):9315–22. doi:10.1016/j.eswa.2012.02.070.

An, B. J., A. Ang, T. G. Bali, and N. Cakici. 2014. The joint cross section of stocks and options. *The Journal of Finance* 69 (5):2279–337. doi:10.1111/jofi.12181.

Andreou, P. C., C. Charalambous, and S. H. Martzoukos. 2014. Assessing the performance of symmetric and asymmetric implied volatility functions. *Review of Quantitative Finance and Accounting* 42 (3):373–97. doi:10.1007/s11156-013-0346-z.

Audrino, F., and D. Colangelo. 2010. Semi-parametric forecasts of the implied volatility surface using regression trees. *Statistics and Computing* 20 (4):421–34. doi:10.1007/s11222-009-9134-y.

Bakshi, G., C. Cao, and Z. Chen. 1997. Empirical performance of alternative option pricing models. *The Journal of Finance* 52 (5):2003–49. doi:10.1111/j.1540-6261.1997.tb02749.x.

Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81:637–59. doi:10.1086/260062.

Bo, L. 2006. Novel algorithm for constructing support vector machine regression ensemble. *Journal of Systems Engineering and Electronics* 17 (3):541–45. doi:10.1016/S1004-4132(06)60093-5.

Cassese, G., and M. Guidolin. 2006. Modelling the implied volatility surface: Does market efficiency matter?: An application to MIB30 index options. *International Review of Financial Analysis* 15 (2):145–78. doi:10.2139/ssrn.535222.

Chatterjee, S., A. Dash, and S. Bandopadhyay. 2015. Ensemble support vector machine algorithm for reliability estimation of a mining machine. *Quality and Reliability Engineering International* 31 (8):1503–16. doi:10.1002/qre.1686.

Christoffersen, P., and K. Jacobs. 2004. The importance of the loss function in option valuation. *Journal of Financial Economics* 72 (2):291–318. doi:10.2139/ssrn.424461.

Ciresan, D. C., U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber. (2011). Flexible, high performance convolutional neural networks for image classification. *International joint conference on artificial intelligence* (pp. 1237–1242). Barcelona, Spain.

Cui, C., and T. Fearn. 2018. Modern practical convolutional neural networks for multivariate regression: Applications to NIR calibration. *Chemometrics and Intelligent Laboratory Systems* 182:9–20. doi:10.1016/j.chemolab.2018.07.008.

Culkin, R., and S. R. Das. 2017. Machine learning in finance: The case of deep learning for option pricing. *Journal of Investment Management* 15 (4):92–100.

Derman, E., I. Kani, and N. Chriss. 1996. Implied trinomial trees of the volatility smile. *Journal of Derivatives* 3 (4):7–22. doi:10.3905/jod.1996.407952.

Dietterich, T. G. (2000). *Ensemble methods in machine learning*. International workshop on multiple classifier systems, 1–15. Berlin, Heidelberg: Springer.

Dupire, B. 1994. Pricing with a smile. *Risk* 7 (1):18–20.

Gençay, R., and M. Qi. 2001. Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging. *IEEE Transactions on Neural Networks* 12 (4):726–34. doi:10.1109/72.935086.

Heston, S. L. 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6 (2):327–43. doi:10.1093/rfs/6.2.327.

Heston, S. L., and S. Nandi. 2000. A closed-form GARCH option valuation model. *Review of Financial Studies* 13 (3):585–625. doi:10.2139/ssrn.210009.

Homescu, C. 2011. *Implied volatility surface: Construction methodologies and characteristics*. Available at SSRN 1882567. doi:10.2139/ssrn.1882567

Hutchinson, J. M., A. W. Lo, and T. Poggio. 1994. A nonparametric approach to pricing and hedging derivative securities via learning networks. *The Journal of Finance* 49 (3):851–89. doi:10.1111/j.1540-6261.1994.tb00081.x.

Jackwerth, J. C., and M. Rubinstein. 1996. Recovering probability distributions from option prices. *The Journal of Finance* 51 (5):1611–31. doi:10.1111/j.1540-6261.1996.tb05219.x.

Kim, I. J., and S. Kim. 2004. Empirical comparison of alternative stochastic volatility option pricing models: Evidence from Korean KOSPI 200 index options market. *Pacific-Basin Finance Journal* 12 (2):117–42. doi:10.1016/S0927-538X(03)00042-8.

Kim, S. 2010. The performance of traders' rules in options market. *Journal of Futures Markets* 29 (11):999–1020. doi:10.1002/fut.20403.

Kim, S., and C. Lee. 2014. On the importance of the traders' rules for pricing options: Evidence from intraday data. *Asia-Pacific Journal of Financial Studies* 43 (6):873–94. doi:10.1111/ajfs.12075.

Li, M., and N. D. Pearson. 2007. A "horse race" among competing option pricing models using S&P 500 index options. *SSRN Electronic Journal*. doi:10.2139/ssrn.952770.

Liang, X., H. Zhang, J. Xiao, and Y. Chen. 2009. Improving option price forecasts with neural networks and support vector regressions. *Neurocomputing* 72 (13–15):3055–65. doi:10.1016/j.neucom.2009.03.015.

Liu, Q., N. Zhang, W. Yang, S. Wang, Z. Cui, X. Chen, and L. Chen. (2017). A review of image recognition with deep convolutional neural network. *International conference on intelligent computing* (pp. 69–80). Cham: Springer. doi:10.1007/978-3-319-63309-1_7

Liu, X., Y. Cao, C. Ma, and L. Shen. 2019. Wavelet-based option pricing: An empirical study. *European Journal of Operational Research* 272 (3):1132–42. doi:10.1016/j.ejor.2018.07.025.

Madan, D. B., P. P. Carr, and E. C. Chang. 1998. The variance gamma process and option pricing. *Review of Finance* 2 (1):79–105. doi:10.1023/A:1009703431535.

Madan, D. B., and F. Milne. 1991. Option pricing with VG martingale components. *Mathematical Finance* 1 (4):39–55. doi:10.1111/mafi.1991.1.issue-4.

Malliaris, M., and L. Salchenberger. 1996. Using neural networks to forecast the S&P 100 implied volatility. *Neurocomputing* 10 (2):183–95. doi:10.1016/0925-2312(95)00019-4.

Merton, R. C. 1976. The impact on option pricing of specification error in the underlying stock price returns. *The Journal of Finance* 31 (2):333–50. doi:10.1111/j.1540-6261.1976.tb01890.x.

Mostafa, F., T. Dillon, and E. Chang. (2015). Computational intelligence approach to capturing the implied volatility. *IFIP international conference on artificial intelligence in theory and practice* (pp. 85–97). Cham: Springer. doi:10.4070/kcj.2015.45.1.85

Park, H., N. Kim, and J. Lee. 2014. Parametric models and non-parametric machine learning models for predicting option prices: Empirical comparison study over KOSPI 200 index options. *Expert Systems with Applications* 41 (11):5227–37. doi:10.1016/j.eswa.2014.01.032.

Qureshi, A. S., A. Khan, A. Zameer, and A. Usman. 2017. Wind power prediction using deep neural network based meta regression and transfer learning. *Applied Soft Computing* 58:742–55. doi:10.1016/j.asoc.2017.05.031.

Ren, Y., L. Zhang, and P. N. Suganthan. 2016. Ensemble classification and regression-recent developments, applications and future directions. *IEEE Computational Intelligence Magazine* 11 (1):41–53. doi:10.1109/MCI.2015.2471235.

Rubinstein, M. 1994. Implied binomial trees. *The Journal of Finance* 49 (3):771–818. doi:10.1111/j.1540-6261.1994.tb00079.x.

Scherer, D., A. Müller, and S. Behnke. (2010). Evaluation of pooling operations in convolutional architectures for object recognition. *International conference on artificial neural networks* (pp. 92–101). Berlin, Heidelberg: Springer.

Singh, S., and A. Dixit. 2016. Performance of the Heston's stochastic volatility model: A study in Indian index options market. *Theoretical Economics Letters* 6 (02):151–65. doi:10.4236/tel.2016.62018.

Wang, C. P., S. H. Lin, H. H. Huang, and P. C. Wu. 2012. Using neural network for forecasting TXO price under different volatility models. *Expert Systems with Applications* 39 (5):5025–32. doi:10.1016/j.eswa.2011.11.038.

Wang, L., and J. Wu. (2012). Application of hybrid RBF neural network ensemble model based on wavelet support vector machine regression in rainfall time series forecasting. *Fifth international joint conference on computational sciences and optimization IEEE* (pp. 867–71). doi:10.1109/CSO.2012.195

Wang, P. 2011. Pricing currency options with support vector regression and stochastic volatility model with jumps. *Expert Systems with Applications* 38 (1):1–7. doi:10.1016/j.eswa.2010.05.037.

Wang, Y. H. 2009. Nonlinear neural network forecasting model for stock index option price: Hybrid GJR–GARCH approach. *Expert Systems with Applications* 36 (1):564–70. doi:10.1016/j.eswa.2007.09.056.

Yang, Y., Y. Zheng, and T. M. Hospedales. (2017). Gated neural networks for option pricing: Rationality by design. *Thirty-first AAAI conference on artificial intelligence*. San Francisco, California, USA.

Yao, J., Y. Li, and C. L. Tan. 2000. Option price forecasting using neural networks. *Omega* 28 (4):455–66. doi:10.1016/S0305-0483(99)00066-3.

Zeng, Y., and D. Klabjan. 2019. Online adaptive machine learning based algorithm for implied volatility surface modeling. *Knowledge-Based Systems* 163:376–91. doi:10.1016/j.knosys.2018.08.039.

Zhang, Z., S. Zohren, and S. Robert. 2019. DeepLOB: Deep convolutional neural networks for limit order books. *IEEE Transactions on Signal Processing* 67 (11):3001–12. doi:10.1109/TSP.2019.2907260.