REM WORKING PAPER SERIES

Conformal prediction of option prices

João A. Bastos

REM Working Paper 0304-2023

December 2023

REM – Research in Economics and Mathematics
Rua Miguel Lúpi 20,
1249-078 Lisboa,
Portugal

ISSN 2184-108X

REM – Research in Economics and Mathematics

Rua Miguel Lupi, 20
1249-078 LISBOA
Portugal

Telephone: +351 - 213 925 912
E-mail: rem@iseg.ulisboa.pt

https://rem.rc.iseg.ulisboa.pt/

https://twitter.com/ResearchRem

https://www.linkedin.com/company/researchrem/

https://www.facebook.com/researchrem/

Conformal prediction of option prices

João A. Bastos*

Lisbon School of Economics & Management (ISEG) and REM
Universidade de Lisboa

Abstract

The uncertainty associated with option price predictions has largely been overlooked in the literature. This paper aims to fill this gap by quantifying such uncertainty using conformal prediction. Conformal prediction is a model-agnostic procedure that constructs prediction intervals, ensuring valid coverage in finite samples without relying on distributional assumptions. Through the simulation of synthetic option prices, we find that conformal prediction generates prediction intervals for gradient boosting machines with an empirical coverage close to the nominal level. Conversely, non-conformal prediction intervals exhibit empirical coverage levels that fall short of the nominal target. In other words, they fail to contain the actual option price more frequently than expected for a given coverage level. As anticipated, we also observe a decrease in the width of prediction intervals as the size of the training data increases. However, we uncover significant variations in the width of these intervals across different options. Specifically, out-of-the-money options and those with a short time-to-maturity exhibit relatively wider prediction intervals. Then, we perform an empirical study using American call and put options on individual stocks. We find that the empirical results replicate those obtained in the simulation experiment.

Keywords: Conformal prediction; Machine learning; Option price; Quantile regression; American options.

1 Introduction

The use of machine learning methods for pricing and hedging options has become well-established in the finance literature, with roots in the influential works of Malliaris and Salchenberger (1993) and Hutchinson et al. (1994). These methods excel at estimating complex and nonlinear relationships, making them valuable tools for pricing options, even when closed-form pricing formulas are available. Among the various machine learning algorithms, the feed-forward neural network trained with back-propagation (Rumelhart et al., 1986) stands out as the most commonly used model for predicting option prices (e.g., Anders et al., 1998; Garcia and Gençay, 2000; Andreou et al., 2010). Nonetheless, alternative machine learning models, such as support vector machines (Vapnik, 1995), have demonstrated their competitiveness relative to neural networks in pricing options

*ISEG, Rua do Quelhas 6, 1200-781 Lisboa, Portugal. E-mail address: jbastos@iseg.ulisboa.pt, Phone: +351 213 925 800.

(Wang, 2011; Park et al, 2014). Moreover, a recent study by Ivaşcu (2021) compared various machine learning algorithms for predicting option prices, revealing that tree-based ensembles, such as random forests (Breiman, 2001) and gradient boosting machines (Friedman, 2001), outperformed neural networks.[1]

From a risk management perspective, it is surprising that the literature on option pricing with machine learning has primarily focused on point predictions, neglecting the quantification of the associated uncertainty. This oversight is particularly significant, as the findings of this paper reveal substantial variations in this uncertainty across different option contracts. Moreover, when the option price is known, prediction intervals can serve as a valuable speculative tool. If the option price falls below the prediction interval, it suggests potential undervaluation, indicating a possible long position. Conversely, if the option price exceeds the prediction interval, it may indicate overvaluation, suggesting a potential short position.

In this paper, we address the problem of quantifying uncertainty in option price predictions using conformal prediction (Papadopoulos et al., 2002; Vovk et al., 2005, 2009; Lei et al., 2013; Lei and Wasserman, 2014). To achieve this objective, we employ conformal quantile regression (Romano et al., 2019), a technique for constructing statistically rigorous prediction intervals for machine learning models. In contrast to other techniques, conformal quantile regression enables the creation of flexible and adaptive intervals capable of capturing the varying levels of uncertainty associated with different option contracts. To the best of our knowledge, this is the first application of conformal prediction in the field of asset pricing. The predictive model is a gradient boosting machine based on decision trees. Preliminary results showed that this model predicts more accurately option prices in our dataset than feed-forward neural networks. This is not surprising since tree-based model still outperform neural networks on many problems with tabular data (Grinsztajn et al., 2022).

We conduct a comparison between the properties of intervals derived from conformal quantile regressions and those obtained from normal (or *non-conformal*) quantile regressions. Through simulations and a real dataset of American options on individual stocks, we show that conformalized quantile models consistently provide prediction intervals with coverage levels near the nominal target. In contrast, non-conformalized quantile models give prediction intervals that fall short in covering the test data. In other words, these intervals fail to contain the actual option price more frequently than expected for a given coverage level. We further note significant variations in the uncertainty associated with the predicted values. Specifically, we find wider prediction intervals for out-of-the-money options.[2] Indeed, pricing out-of-the-money options can be more challenging than pricing in-the-money options since the former have little intrinsic value and tend to lose value rapidly as they approach their expiration date. Similarly, we find large intervals for options with a short time-to-maturity, which are susceptible to pin risk (Golez and Jackwerth, 2012). These findings shed light on the neglected variation in the uncertainty

[1]Naturally, machine learning methods find various other applications in finance, with one of the most common being stock return prediction. For recent studies on this topic, see Ribeiro et al (2021), Kanwal et al (2022), Chaudhari and Thakkar (2023), Gülmez (2023), Nakayama et al. (2023), Deng et al. (2024).

[2]Out-of-the-money options have minimal intrinsic value, specifically, call options with an underlying asset price lower than the strike price or put options with an underlying asset price higher than the strike price. Conversely, in-the-money options hold intrinsic value and can be immediately exercised – call options when the underlying asset price exceeds the strike price or put options when the underlying asset price is lower than the strike price. At-the-money options possess a strike price that closely aligns with the price of the underlying asset.

associated with option price predictions given by machine learning models.

The remainder of this paper is structured as follows. The next section introduces some potential approaches for measuring the uncertainty of option price predictions and their limitations. In Section 3, the conformal prediction framework for option prices is detailed. Section 4 presents the results derived from simulated data, while Section 5 describes an empirical application focused on American options associated with individual stocks. Finally, Section 6 summarizes the key findings from the study.

2 Preliminaries

The problem at hand involves a set of n option prices $\{Y_i\}_{i=1}^n$ and corresponding explanatory variables $\{\boldsymbol{X}_i\}_{i=1}^n$, such as spot price, strike price, and time-to-maturity. Our objective is to construct a prediction interval $\mathcal{C}(\boldsymbol{X}_{n+1}) \subseteq \mathbb{R}$ for a new option contract with known characteristics $\boldsymbol{X}_{n+1}$ but an unknown price Y_{n+1}. Given a significance level α, the goal is to ensure that the prediction interval has high probability of containing the unknown price

$$\Pr(Y_{n+1} \in \mathcal{C}(\boldsymbol{X}_{n+1})) \geq 1 - \alpha, \tag{1}$$

where $1-\alpha$ represents the desired nominal coverage. Ideally, this interval should possess favorable finite-sample properties and be independent of the joint distribution $P_{\boldsymbol{X},Y}$ over which the probability in Equation (1) is calculated.

For European options, a straightforward approach to address this problem is by assuming a data generating process for the underlying asset price. By generating n asset paths and calculating the option payoffs at maturity for each path, we can compute the mean payoff at maturity $\bar{Y}_T$ and its standard deviation $s(Y_T)$. Subsequently, a prediction interval for the option value at maturity can be constructed as:

$$\left[\bar{Y}_T - z_{\frac{\alpha}{2}} \frac{s(Y_T)}{\sqrt{n}}, \bar{Y}_T + z_{1-\frac{\alpha}{2}} \frac{s(Y_T)}{\sqrt{n}}\right] \tag{2}$$

where $z_{\frac{\alpha}{2}}$ and $z_{1-\frac{\alpha}{2}}$ represent the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ quantiles of the payoff distribution at maturity. Finally, by discounting the boundaries of this interval, a prediction interval for the option before expiration can be obtained. It is important to note that this approach heavily relies on the assumption made regarding the data generating process. If this assumption is incorrect, the resulting intervals will not provide valid coverage. Given the typically unknown dynamics of the underlying asset price, an alternative approach could involve bootstrapping the training set and reestimating the option price on each sample (see, e.g., Tibshirani, 1996). However, this approach would yield prediction intervals with a constant width, which is not realistic and fails to capture changing market conditions.

A different approach was proposed by Healy et al. (2003). They introduced a two-step strategy that involved training a neural network to predict option prices and calculate the squared residuals. Subsequently, they trained a second network with two outputs: the first output aimed to predict option prices, while the second output targeted the squared residuals derived from the first network predictions. This approach aimed to provide both a point prediction and a measure of uncertainty for the price of a new option, which could then be used to construct prediction intervals. However, this approach poses two main issues. First, employing two output units in the network results in a doubling of the connections between the last hidden layer and the output layer. Consequently, this increased complexity requires a larger amount of data to effectively train the model.

Second, there are no theoretical guarantees that the resulting prediction intervals have valid coverage, meaning their reliability is not assured.

A strategy that avoids increasing the size of the model is to predict the conditional quantiles of option prices. Let $F(Y|\boldsymbol{X})$ denote the conditional distribution function of option prices Y given a set of explanatory variables $\boldsymbol{X}$. The conditional quantile function is defined as

$$q_\alpha(\boldsymbol{X}) = \inf\{Y \in \mathbb{R} : F(Y|\boldsymbol{X}) \geq \alpha\}. \tag{3}$$

By training two models to learn the quantile functions $q_{\frac{\alpha}{2}}(\boldsymbol{X})$ and $q_{1-\frac{\alpha}{2}}(\boldsymbol{X})$, we can obtain a conditional prediction interval for Y_{n+1} with a nominal coverage level of $1-\alpha$, given by

$$\left[q_{\frac{\alpha}{2}}(\boldsymbol{X}_{n+1}), q_{1-\frac{\alpha}{2}}(\boldsymbol{X}_{n+1})\right]. \tag{4}$$

Regarding the functional form for $q_\alpha(\boldsymbol{X})$ we could consider the conventional quantile regression of Koenker and Bassett (1978). But this would impose a linear association between the conditional quantiles of the target variable and the predictors, and option prices are non-linear on their explanatory variables. Fortunately, many non-linear machine learning algorithms can be trained to learn quantiles instead of the mean response. The key is to minimize a "pinball loss" function rather than a quadratic loss function.

Unfortunately, intervals estimated using "plain" quantile machine learning models do not guarantee valid coverage in most circumstances (Takeuchi et al., 2006; Meinshausen, 2006; Steinwart and Christmann, 2011). Simulation experiments and empirical results below show that this approach leads to intervals that seriously undercover the observed option prices, resulting in a higher frequency of missing the actual option prices than expected for a given nominal coverage level. To address this miscoverage problem, an effective solution is the use of *conformal prediction*. Conformal prediction is a model-agnostic technique that constructs prediction intervals without relying on distributional assumptions, ensuring valid coverage in finite samples (Papadopoulos et al., 2002; Vovk et al., 2005, 2009; Lei et al., 2013; Lei and Wasserman, 2014). The finite-sample validity of conformal prediction is particularly beneficial for thinly traded options and newly created contracts where limited data points are available.

Early conformal regression approaches suffer from the limitation of producing intervals with constant or weakly varying lengths across the regressor space. This restrictiveness is problematic, as the findings in the paper highlight significant variations in model uncertainty among different option contracts. Indeed, the empirical analysis reveals relative prediction intervals spanning from a few percent to a staggering thirty percent. To overcome this limitation, Romano et al. (2019) introduced the concept of *conformalized quantile regression*. This procedure enables the derivation of confidence intervals with accurate coverage in finite samples from any quantile regression model. Notably, the only assumption required for this procedure is that the observations $\{(\boldsymbol{X}_i, Y_i)\}_{i=1}^{n+1}$ are exchangeable. By adopting conformalized quantile regression, the limitations of early conformal regression approaches can be mitigated, allowing for more flexible and adaptive intervals that capture the varying levels of uncertainty across different option contracts.

3 Methodology

Let Y represent an option price, and $\boldsymbol{X}$ denote a vector of explanatory variables. We consider a sample of n options, denoted as $\{(\boldsymbol{X}_i, Y_i)\}_{i=1}^{n}$, which are used to train a machine learning model. Additionally, we have a new option contract with known characteristics

$\boldsymbol{X}_{n+1}$ but an unknown price Y_{n+1}. We assume that all observations in the dataset, including the new option, $\{(\boldsymbol{X}_i, Y_i)\}_{i=1}^{n+1}$, are exchangeable and drawn from a joint distribution $P_{\boldsymbol{X},Y}$. This assumption holds automatically if the $(\boldsymbol{X}_i, Y_i)$ are independent and identically distributed. Our objective is to construct a prediction interval $\mathcal{C}(\boldsymbol{X}_{n+1}) \subseteq \mathbb{R}$ that satisfies Equation (1) for any desired coverage level $1-\alpha$. Moreover, this relationship should hold for any joint distribution $P_{\boldsymbol{X},Y}$. The following discussion is based on the work of Romano et al. (2019).

3.1 Split conformal prediction

A simple approach for constructing conformal intervals is the "split conformal prediction" method (Papadopoulos et al., 2002). This method involves splitting the training data into two subsets: $S_1 = \{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_1\}$ for estimation, and $S_2 = \{(\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_2\}$ as the calibration set for obtaining "conformity scores", where $\mathcal{I}$ represents a set of observation indices. A regression model $Y = f(\boldsymbol{X})$ is then trained on S_1, where any regression function can be used. Next, conformity scores are computed for each observation in the calibration set using the absolute residuals of the trained model $\hat{f}(\boldsymbol{X})$:

$$\hat{\varepsilon}_i = |Y_i - \hat{f}(\boldsymbol{X}_i)| \;\; \forall\, i \in \mathcal{I}_2. \tag{5}$$

For a given significance level α, the quantile $q_{1-\alpha}(\mathcal{I}_2)$ is calculated from the empirical distribution of the conformity scores,

$$q_{1-\alpha}(\mathcal{I}_2) = \frac{(n_2+1)(1-\alpha)}{n_2}\text{th-quantile of } \hat{\varepsilon}_i : i \in \mathcal{I}_2, \tag{6}$$

where n_2 is the number of observations in the calibration set. Finally, the prediction interval for the unknown price Y_{n+1} of option $\boldsymbol{X}_{n+1}$ is constructed as:

$$\mathcal{C}(\boldsymbol{X}_{n+1}) = \left[\hat{f}(\boldsymbol{X}_{n+1}) - q_{1-\alpha}(\mathcal{I}_2),\; \hat{f}(\boldsymbol{X}_{n+1}) + q_{1-\alpha}(\mathcal{I}_2)\right]. \tag{7}$$

One limitation of the split conformal prediction approach is that it produces constant prediction intervals, which fails to capture the varying width of intervals observed empirically for option prices. However, the split conformal prediction method can be extended to address this limitation by introducing a locally adaptive approach known as "locally adaptive conformal prediction" (Papadopoulos et al., 2008). This method scales the absolute residuals $\hat{\varepsilon}_i$ by their dispersion at $\boldsymbol{X}_i$, allowing for non-constant prediction intervals. But this procedure has several limitations, as explained in Romano et al. (2019, pp. 8).

3.2 Conformal quantile prediction

In conformal quantile prediction, the training dataset is also divided into two subsets: one for estimation and the other for calibration to obtain conformity scores. This procedure uses any regression model for quantiles. Initially, two regression models for quantiles, namely $\hat{q}_{\frac{\alpha}{2}}(\boldsymbol{X})$ and $\hat{q}_{1-\frac{\alpha}{2}}(\boldsymbol{X})$, are trained using subset S_1 for a given coverage level of $1-\alpha$. Subsequently, the conformity scores are calculated using:

$$\hat{\varepsilon}_i = \max\left[\hat{q}_{\frac{\alpha}{2}}(\boldsymbol{X_i}) - Y_i, Y_i - \hat{q}_{1-\frac{\alpha}{2}}(\boldsymbol{X_i})\right],\; \forall\, i \in \mathcal{I}_2. \tag{8}$$

Next, Equation (6) is employed to compute the $q_{1-\alpha}(\mathcal{I}_2)$ quantile of the empirical distribution of these conformity scores. Finally, the conformalized prediction interval for Y_{n+1} is given by:

$$\mathcal{C}(\boldsymbol{X}_{n+1}) = \left[\hat{q}_{\frac{\alpha}{2}}(\boldsymbol{X}_{n+1}) - q_{1-\alpha}(\mathcal{I}_2),\ \hat{q}_{1-\frac{\alpha}{2}}(\boldsymbol{X}_{n+1}) + q_{1-\alpha}(\mathcal{I}_2)\right]. \tag{9}$$

Theorem (Romano et al., 2019): If $(\boldsymbol{X}_i, Y_i)_{i=1}^{n+1}$ are drawn interchangeably, the prediction interval $\mathcal{C}(\boldsymbol{X}_{n+1})$ given in Equation (9) satisfies:

$$\Pr(Y_{n+1} \in \mathcal{C}(\boldsymbol{X}_{n+1})) \geq 1 - \alpha,$$

If the conformity scores $\{\hat{\varepsilon}_i : i \in \mathcal{I}_2\}$ are almost surely distinct, then $\mathcal{C}(\boldsymbol{X}_{n+1})$ is nearly perfectly calibrated, with the following relationship:

$$\Pr(Y_{n+1} \in \mathcal{C}(\boldsymbol{X}_{n+1})) \leq 1 - \alpha + \frac{1}{1+n_2}.$$

3.3 Machine learning model for quantiles

To obtain the conformalized prediction intervals described in Equation (9), a regression model for quantiles is required. Here, we employ a modified version of a "gradient boosting machine" (Friedman, 2001) as the quantile regression model. The gradient boosting machine combines multiple base models to form a robust "committee" of models. Typically, decision trees (Breiman et al., 1983; Quinlan, 1986) are used as the base models. Decision trees consist of a series of if-then-else conditions based on the features of the observations, which ultimately lead to a prediction. A decision tree can contain numerous branches, each with multiple sequential tests on the features. The prediction $\hat{Y}$ of the gradient boosting machine is obtained by summing the predictions of a set of K decision trees $\{f_k(\boldsymbol{X})\}_{k=1}^{K}$:

$$\hat{Y} = \sum_{k=1}^{K} f_k(\boldsymbol{X}). \tag{10}$$

The initial tree, $f_1(\boldsymbol{X})$, is a standard decision tree trained on the original data. The subsequent decision trees, $\{f_k(\boldsymbol{X})\}_{k=2}^{K}$, are incrementally added to the committee. However, each new tree is trained on the errors produced by the trees already present in the committee. This process aims to rectify the errors made by the existing committee of trees. During each iteration, the tree to be added is the one that minimizes the regularized loss function.

$$\sum_{i=1}^{n} L\left(Y_i, \hat{Y}_i^{(k-1)} + f_k(\boldsymbol{X}_i)\right) + \gamma T + \frac{1}{2}\lambda||\boldsymbol{w}_k||^2. \tag{11}$$

When aiming to predict the mean response, the loss $L(\cdot)$ is typically defined as the squared-error loss:

$$L\left(Y_i, \hat{Y}_i^{(k-1)} + f_k(\boldsymbol{X}_i)\right) = \left(Y_i - \hat{Y}_i^{(k-1)} - f_k(\boldsymbol{X}_i)\right)^2. \tag{12}$$

However, in the context of predicting quantiles, we use a different loss known as the "pinball loss", which is defined as follows:

$$L\left(Y_i, \hat{Y}_i^{(k-1)} + f_k(\boldsymbol{X}_i)\right) = \begin{cases} \alpha\left(Y_i - \hat{Y}_i^{(k-1)} - f_k(\boldsymbol{X}_i)\right), & \text{if } Y_i \geq \hat{Y}_i^{(k-1)} + f_k(\boldsymbol{X}_i) \\ (1-\alpha)\left(\hat{Y}_i^{(k-1)} + f_k(\boldsymbol{X}_i) - Y_i\right), & \text{otherwise} \end{cases} \tag{13}$$

The last two terms in Equation (11) are regularization terms that aim to penalize complex trees, thereby preventing the committee from overfitting the training data. The parameter γ serves as a penalty on the number of terminal nodes in a tree, denoted by T, while λ serves as a penalty on the magnitude of the tree weights $\boldsymbol{w}_k$. To add new trees, a gradient descent algorithm is used to minimize the loss function. There are several efficient implementation of gradient boosting. In this paper, we use the "Light Gradient Boosting Machine", or LightGBM (Ke et al., 2017). LightGBM is known for its speed and efficiency, surpassing the popular "Extreme Gradient Boosting" (XGBoost) (Chen and Guestrin, 2016) in terms of training time. Moreover, both implementations exhibit comparable accuracy.[3] LightGBM has a range of "hyperparameters" – parameters that are not learned during the training process and need to be optimized.

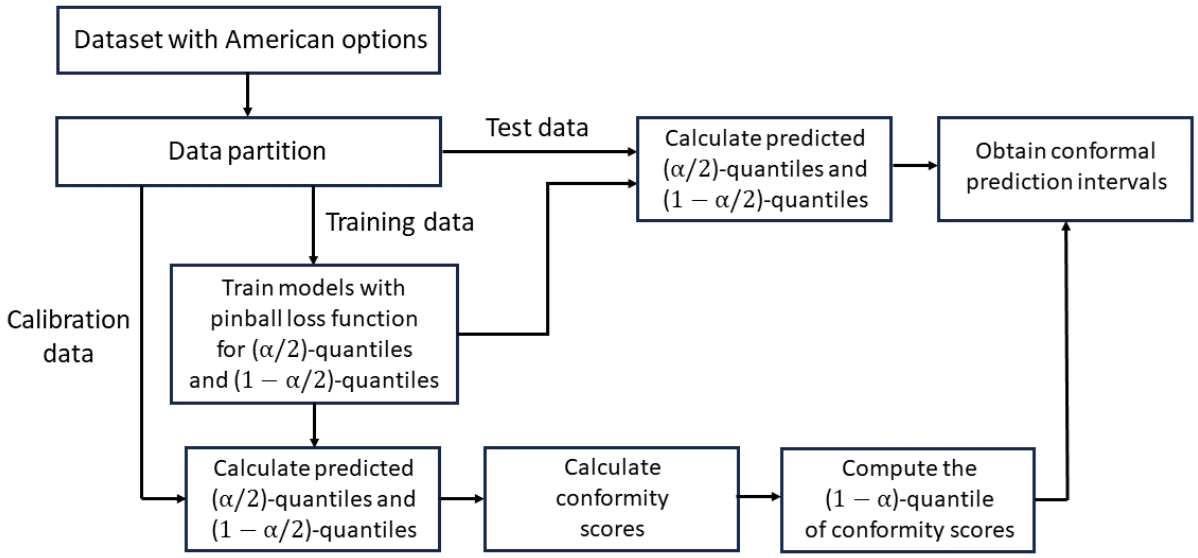

Figure 1: Flowchart for conformal quantile prediction. The significance level of the confidence intervals is $1 - \alpha$.

Figure 1 shows a flowchart summarizing the procedure for calculating conformal prediction intervals. Initially, the data are partitioned into three distinct datasets. The training data serves as the basis for training the models with pinball loss functions. Subsequently, the calibration data is used to predict quantiles and derive conformity scores. The properties of the conformal prediction intervals are evaluated using an independent test dataset.

4 Simulation experiments

With Monte Carlo simulations we created a controlled environment to compare the properties of conformalized quantile regression with those of the non-conformalized counterpart. The Monte Carlo experiments focused on simulating call options on stocks and assuming that their prices are determined by the Black-Scholes formula for pricing European options. It was assumed that the underlying asset does not pay dividends. In this particular scenario, where there are no ex-dividend events during the life of the option, it is always optimal to exercise the option at expiry. As a result, the Black-Scholes formula

[3]source: https://lightgbm.readthedocs.io/en/latest/Experiments.html

can be used to value American call options, and the simulation results are applicable to both European and American call options.

Variable	**Simulated process**
Spot price	$S \sim U(500, 1500)$
Volatility	$\sigma \sim U(0.1, 1)$
Time-to-maturity	$\tau \sim U(14/252, 2)$
Interest rate	$r \sim U(0.001, 0.05)$
Strike price	$K \sim S/z,\ z \sim N(1, 0.1)$

Table 1: Statistical distributions for generating Black-Scholes prices of non-dividend paying call options. $U(a, b)$ is the uniform distribution bounded by a and b, and $N(\mu, s^2)$ is the normal distribution with mean μ and variance s^2.

The simulation data for the option parameters are generated according to the processes outlined in Table 1. The data generation process is the following. A specified number of sets $\{S, \sigma, \tau, r\}$ is generated based on the processes described in Table 1. For each set, four strike prices K are generated using the process specified in the last row of Table 1 – that is, the strike prices are constrained to be in the vicinity of the spot price S. For each set $\{S, \sigma, \tau, r, K\}$, the call prices C_{BS} are calculated using the Black-Scholes formula. To investigate the finite-sample properties of the procedures under various sample sizes, datasets of four different sizes were generated: 20,000, 50,000, 100,000, and 200,000 observations.

In all simulation experiments and the empirical application, a random set with 20% of the observations is used as validation data to derive prediction intervals. In the case of conformal models, an additional 20% of the observations are randomly held out as calibration data. As a result, non-conformal models are trained using 80% of the observations, while conformal models are trained using only 60%. One might argue that non-conformal models, being trained on more data, would yield more accurate point predictions. This is indeed true, especially for small datasets. However, our primary focus is to obtain prediction intervals with valid coverage, even if it requires sacrificing some training data to achieve this goal. It is worth noting that when a machine learning model is deployed in a production environment, it is usually trained using all available data before deployment.

We make the assumption that the homogeneity of degree one in the spot and strike prices of the Black-Scholes formula partially holds, regardless of the true pricing model. Therefore, both spot and option prices are scaled by the strike price – the machine learning models are trained to learn the relationship between C_{BS}/K and the inputs $\{S/K, \sigma, \tau, r\}$. After training, the predicted prices and intervals can be recovered in monetary units by multiplying them by the strike price. To determine the optimal hyperparameters for the models, a grid-search approach was performed. The objective was to minimize the mean absolute error in the validation data. The following hyperparameters were included in the grid-search:

1. Number of decision trees in the ensemble: The initial trees added to the committee typically provide significant improvements in out-of-sample accuracy, but the marginal gains decrease as more trees are added.

2. Maximum depth of the decision trees: Using very large trees may lead to overfitting of the training data and may not necessarily result in the best out-of-sample

accuracy.

3. Maximum number of leaves in one tree: A large number of leaves in a tree may also lead to overfitting and may not provide the best out-of-sample accuracy.

4. Learning rate: This parameter controls the step size in the gradient descent algorithm and can affect the convergence and overall performance of the model.

Hyperparameter	**Search space**
Number of trees in the ensemble	$\{100, 500, 1000, 2500, 5000\}$
Maximum number of leaves in one tree	$\{32, 64, 128, 256\}$
Maximum tree depth	$\{8, 16, 32, 64\}$
Learning rate	$\{0.005, 0.01, 0.05, 0.1, 0.5\}$

Table 2: Hyperparameters' search space for the Light Gradient Boosting Machine.

By performing a grid-search over these hyperparameters, the optimal values were determined, resulting in the models with the lowest mean absolute error on the validation data. The hyperparameters' search space is shown in Table 2.

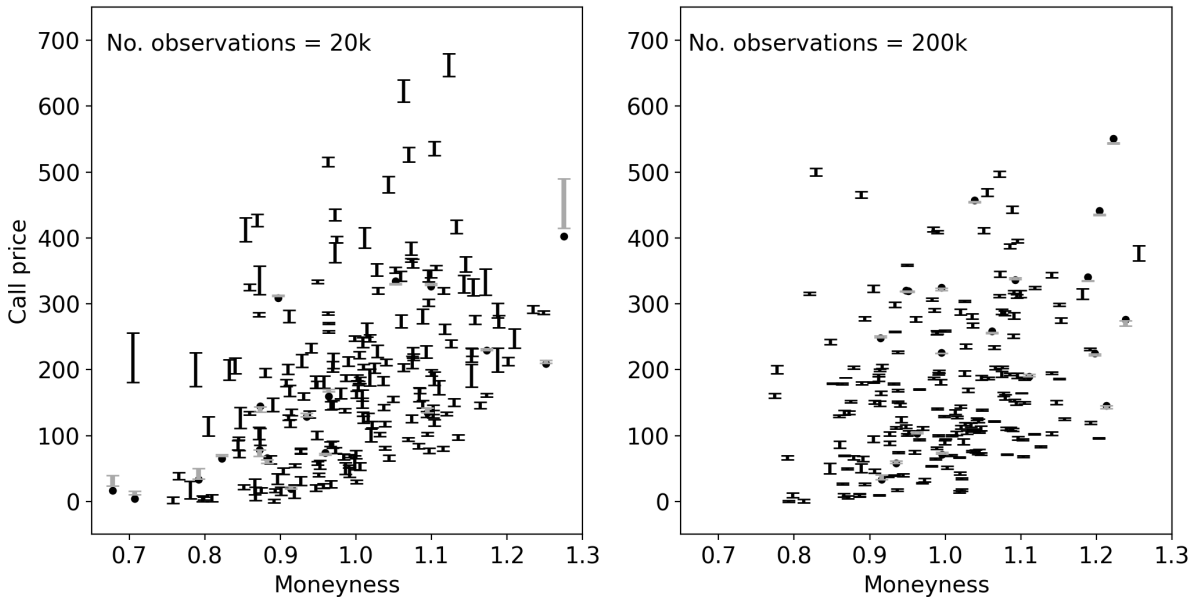

Figure 2: Conformal prediction intervals against option moneyness (S/K) for the simulated datasets with 20,000 and 200,000 observations. The bars in gray represent the cases where the prediction interval failed to cover the actual value. The dots represent actual option prices missed by the intervals. The nominal coverage is 0.9.

Once the models are trained, prediction intervals can be calculated for the test data. Figure 2 provides an illustration of the conformal prediction intervals for a randomly selected sample of 200 observations plotted against the option moneyness (S/K).[4] The left plot corresponds to the dataset with 20,000 observations, while the right plot corresponds to the larger dataset with 200,000 observations. The nominal coverage of these

[4]Conformal prediction intervals were obtained with the MAPIE (Model Agnostic Prediction Interval Estimator) package: https://mapie.readthedocs.io/en/latest/index.html.

intervals is set at 0.90, meaning that if the conformal intervals are valid, they should encompass approximately 90% of the actual option prices. In the plots, the bars in gray represent the cases where the prediction interval failed to cover the actual value. The dots represent the option prices missed by the intervals. The adaptability of interval widths to model uncertainty is clearly demonstrated by the observed large variations in the size of the intervals. This variability reflects the level of uncertainty associated with the model's predictions – the intervals widen or narrow depending on the degree of prediction uncertainty. The intervals for the larger dataset with 200,000 observations are typically smaller. This indicates that with larger training datasets, the machine learning models gain more "confidence" in their predictions, leading to narrower intervals.

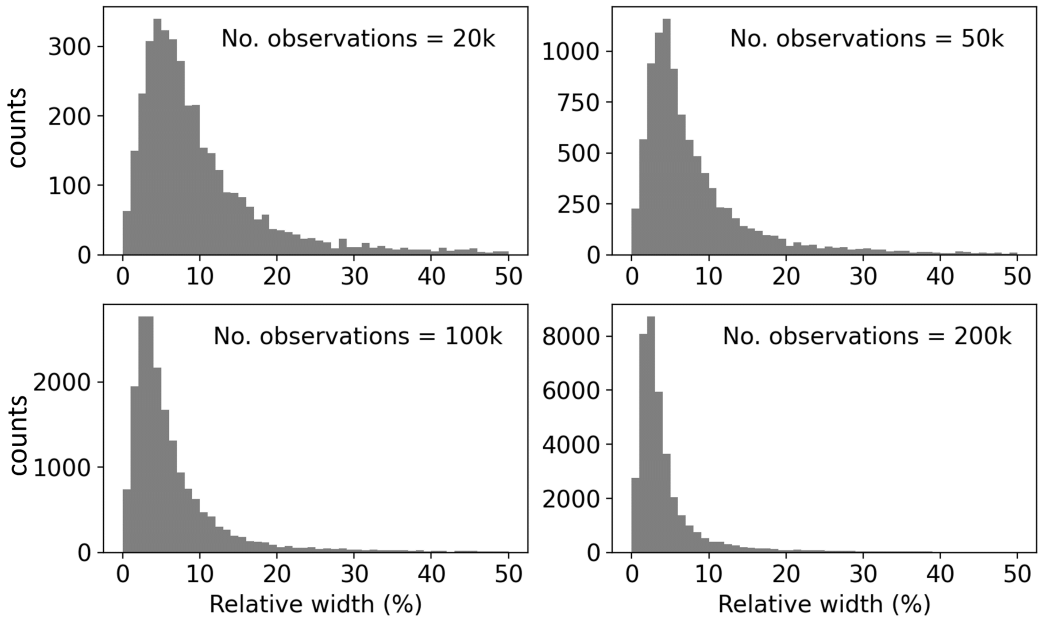

Figure 3: Distributions of relative interval widths (in %) for the simulated data and different sample sizes. The nominal coverage is 0.9.

Figure 3 displays the distributions of relative interval widths – expressed as a percentage – for the four simulated datasets with varying sample sizes. These distributions exhibit an asymmetric shape with positive skewness. As anticipated, the relative interval widths tend to concentrate at smaller values as the sample size increases, indicating that the models improve their ability to accurately map the input variables to option prices.

4.1 Empirical coverages

To evaluate the validity of the coverage guarantees provided by the models, we can calculate the empirical coverage, which is defined as the proportion of actual option prices that fall within the prediction interval:

$$\frac{1}{n_{\text{test}}} \sum_{i \in \mathcal{I}_{\text{test}}} \mathbf{1}(Y_i \in \mathcal{C}(\mathbf{X}_i)). \tag{14}$$

Here, n_{test} represents the number of observations in the test set, and $\mathcal{I}_{\text{test}}$ is the set of indices corresponding to this set. The indicator function $\mathbf{1}(Y_i \in \mathcal{C}(\mathbf{X}_i))$ is equal to 1 if the actual option price Y_i falls within the prediction interval $\mathcal{C}(\mathbf{X}_i)$, and 0 otherwise. Additionally, it is valuable to examine the overall width of the intervals. Since the distributions of relative interval widths in Figure 3 exhibit significant skewness, we report the median relative interval width over the test set. This provides an overall measure of interval widths that is less affected by extreme values and captures the typical range of interval widths observed in the data. All the reported results in the following analysis correspond to a nominal coverage level of 0.9.

	Empirical coverage		**Relative width**	
Dataset size	**CQR**	**NQR**	**CQR**	**NQR**
20,000	0.91	0.61	7.9%	2.4%
50,000	0.90	0.66	5.9%	2.1%
100,000	0.89	0.70	4.3%	1.8%
200,000	0.90	0.74	2.9%	1.4%

Table 3: Empirical coverages and median relative widths of the prediction intervals given by the conformal quantile regressions (CQR) and the non-conformal quantile regressions (NQR) for the simulated data. Nominal coverage level is 0.9.

Table 3 reports the empirical coverage of the prediction intervals for the simulated test data, as well as the median relative interval widths. Considering Equation (9), these are given by:

$$\text{median}\left(\frac{\left[\hat{q}_{1-\frac{\alpha}{2}}(\boldsymbol{X}) + q_{1-\alpha}(\mathcal{I}_2)\right] - \left[\hat{q}_{\frac{\alpha}{2}}(\boldsymbol{X}) - q_{1-\alpha}(\mathcal{I}_2)\right]}{Y}\right). \tag{15}$$

CQR refers to the intervals derived from the conformal machine learning models for quantiles, while NQR refers to those obtained using plain (or non-conformal) machine learning models for quantiles. For the conformalized quantile regressions (CQR), regardless of the sample size, the empirical coverage in the test set closely aligns with the nominal level of 0.9. This indicates that the conformal models provide reliable finite-sample coverage guarantees. In contrast, the non-conformalized quantile regressions (NQR) severely undercover the actual values, especially for smaller sample sizes. However, as the sample size increases, the empirical coverage improves, although it still deviates from the nominal level. This suggests that the non-conformal models struggle to provide accurate coverage guarantees, even when trained on large datasets.

Regarding the median relative interval width, it is expected that the non-conformal models yield narrower intervals compared to the conformal models. This is because the non-conformal models tend to underestimate the uncertainty. On the other hand, the conformal models exhibit wider intervals, reflecting their ability to capture the inherent uncertainty in estimating option prices. Furthermore, the widths of both conformal and non-conformal models decrease as the sample size increases – with larger training datasets, the machine learning models gain more confidence in their predictions, resulting in narrower prediction intervals.

4.2 Conditional coverage

The marginal coverage guarantee expressed in Equation (1) should be distinguished from the conditional coverage guarantee:

$$\Pr\left(Y_{n+1} \in \mathcal{C}(\boldsymbol{X}_{n+1}) | \boldsymbol{X}_{n+1} = \boldsymbol{x}\right)) \geq 1 - \alpha, \tag{16}$$

that should be satisfied for any $\boldsymbol{x}$. However, achieving this conditional coverage guarantee is impossible without making strong assumptions about the joint distribution $P_{\boldsymbol{X},Y}$ (Barber et al., 2021). To obtain valid conditional coverages across different regions of the regressor space, it would be necessary to train individual models on data specific to each of these regions and conformalize the intervals accordingly. Nonetheless, it remains an interesting empirical question to explore the behavior of conditional coverages in specific regions of the regressor space when the model is trained on the entire dataset. Indeed, the data may be too limited to effectively train individual models on subsets that cover small regions of a regressor domain. This is the case of thinly traded options or newly created contracts for which few trades were recorded.

	Moneyness					
	Empirical coverage			**Relative width**		
Dataset size	OTM	ATM	ITM	OTM	ATM	ITM
20,000	0.93	0.93	0.88	12%	7%	6%
50,000	0.91	0.92	0.87	8%	5%	5%
100,000	0.90	0.93	0.86	6%	4%	3%
200,000	0.91	0.93	0.87	4%	3%	2%
	Time-to-maturity					
	Empirical coverage			**Relative width**		
Dataset size	short	medium	long	short	medium	long
20,000	0.93	0.89	0.91	29%	16%	7%
50,000	0.88	0.92	0.90	21%	14%	5%
100,000	0.86	0.90	0.89	15%	10%	4%
200,000	0.88	0.90	0.90	9%	5%	3%

Table 4: Conditional coverages and median relative widths of the conformal prediction intervals for different levels of moneyness (upper panel) and time-to-maturity (lower panel), obtained for the simulated call options. The levels of moneyness (M) are defined as: out-of-the-money (OTM), $M < 0.98$; at-the-money (ATM), $0.98 \leq M < 1.02$; and in-the-money (ITM), $M \geq 1.02$. The different levels of time-to-maturity (τ) are defined as: short-term, $\tau < 30$ days; medium-term, $30 \leq \tau < 60$ days; and long-term, $\tau \geq 60$ days. The nominal coverage level is 0.9.

The upper panel of Table 4 presents the conditional coverages and median relative widths for different levels of moneyness obtained from the conformal models. Moneyness is defined as the ratio of the underlying asset price and the option's strike price (S/K) and measures the intrinsic value of an option. In-the-money call options have a moneyness exceeding 1, while in-the-money put options have a moneyness below 1. Conversely, out-of-the-money call options possess a moneyness less than 1, and out-of-the-money put options hold a moneyness greater than 1. At-the-money options exhibit a moneyness close to 1. To calculate the conditional coverages, the test data was divided into three bins based on moneyness (M): out-of-the-money ($M < 0.98$), at-the-money ($0.98 \leq M <$

1.02), and in-the-money ($M \geq 1.02$). The conditional coverages measure the proportion of prices falling within the prediction interval for each moneyness bin. The conditional widths represent the median relative widths within each moneyness bin. Notably, the conditional coverages align rather closely with the expected nominal level. Regarding the conditional interval widths, we observe again that the relative intervals decrease as the number of observations increases. The lower panel of Table 4 reports the conditional coverages and median relative widths of the conformal prediction intervals for different levels of time-to-maturity (τ): short-term ($\tau < 30$ days), medium-term ($30 \leq \tau < 60$ days), and long-term ($\tau \geq 60$ days). The conditional coverages represent the proportion of prices falling within the prediction interval for each maturity bin, while the conditional widths denote the median relative widths within each maturity bin. Once again, the conditional coverages closely align with the expected nominal level and, as the number of observations increases, the relative width of the interval decreases.

5 Empirical application

5.1 Data

The dataset used in this study comprises information on American option contracts obtained through the Yahoo Finance API[5]. The dataset includes data for both call and put contracts. The underlying assets are stocks of eight large companies: Amazon, AMD, Boeing, Disney, Meta, Netflix, PayPal, and Salesforce. The data collection period for the option prices spans from November 11, 2020, to February 12, 2021. Each option price was matched with the closest trading price of the corresponding underlying stock. None of these companies had dividend events during the options' lifespan at the time of data collection. Consequently, for call contracts, the empirical results can be directly compared to the findings from the simulation exercise. We follow Hutchinson et al. (1994) and use the 3-month Treasury Bill rate from the Federal Reserve Bank of St. Louis as a measure of the short-term interest rate. Throughout the period under examination, this rate remained at historically low levels, fluctuating between 0.04% and 0.16%.

The volatility of the underlying assets was determined using the implied volatility obtained from the binomial options pricing model (Cox et al., 1979). To exclude highly illiquid contracts, the following criteria were applied. Observations with a maturity of fewer than 10 trading days until expiry were removed. Additionally, contracts that were deep in-the-money or deep out-of-the-money ($S/K < 0.5$ or $S/K > 1.5$) were also excluded from the analysis.

Table 5 presents summary statistics for the calls and puts datasets, including the number of observations, mean spot price, mean strike price, mean option price, mean annual implied volatility, and mean time-to-maturity in years. All prices are denominated in US dollars. The number of observations varies across the companies, with Amazon having the highest number and PayPal having the lowest. The mean annual implied volatility ranges from 0.38 to 0.54 across the companies. The mean time-to-maturity ranges from 0.15 (approximately 38 trading days) to 0.27 (approximately 68 trading days). For the call options, the mean spot prices are lower than the mean strike prices, indicating that these options are typically out-of-the-money. Conversely, for the put options, the mean spot prices are higher than the mean strike prices, indicating that

[5]https://pypi.org/project/yfinance/

Call options						
Company	Count	Spot price	Strike price	Option price	Maturity	Volatility
Amazon	14676	3215	3498	96.7	0.18	0.38
AMD	10477	90	100	6.0	0.26	0.54
Boeing	9278	212	246	10.9	0.29	0.52
Disney	6583	168	182	8.5	0.29	0.40
Meta	9941	271	304	11.3	0.27	0.41
Netflix	7061	521	568	19.9	0.19	0.43
PayPal	5399	234	251	13.3	0.26	0.45
Salesforce	6575	230	261	8.9	0.29	0.41
Put options						
Company	Count	Spot price	Strike price	Option price	Maturity	Volatility
Amazon	11537	3225	2994	74.5	0.15	0.38
AMD	6700	90	84	5.1	0.22	0.53
Boeing	6050	211	192	10.4	0.24	0.53
Disney	4380	166	151	4.9	0.24	0.40
Meta	7692	271	245	9.5	0.25	0.43
Netflix	5820	524	476	14.4	0.18	0.43
PayPal	3644	234	211	7.3	0.23	0.45
Salesforce	4367	230	211	8.4	0.25	0.43

Table 5: Summary statistics of the dataset: number of observations; mean spot price; mean strike price; mean option price; mean time-to-maturity (years); and mean annual implied volatility. Prices are in US dollars.

these options are also typically out-of-the-money.

Figure 4 displays the conformal prediction intervals plotted against moneyness for a random sample of 200 options on Amazon stocks. The left plot corresponds to call options, while the right plot corresponds to put options. In the plots, the bars in gray represent the cases where the prediction interval failed to cover the actual value. The dots represent the option prices missed by the intervals. These instances should account for approximately 10% of the plotted data points. The adaptability of the prediction intervals to model uncertainty is evident, as considerable variations in the interval size can be observed across the moneyness range.

Figure 5 presents the conformal prediction intervals plotted against moneyness for 200 options on PayPal stocks. The models used in this case were trained using a smaller dataset, consisting of only 3239 calls and 2126 puts. This corresponds to approximately 37% and 30% of the respective numbers of observations for Amazon calls. The confidence intervals in this case tend to be larger, reflecting the higher uncertainty associated with a model trained on a smaller number of samples. Nevertheless, it is expected that the conformal prediction intervals adjust to cover approximately 90% of the option prices, maintaining the desired nominal coverage level.

5.2 Empirical coverage

To compute the empirical coverage and the median relative widths for the real data, we need to consider the sampling variability caused by the small number of observations in certain datasets, such as those for PayPal and Salesforce. While conformal regression

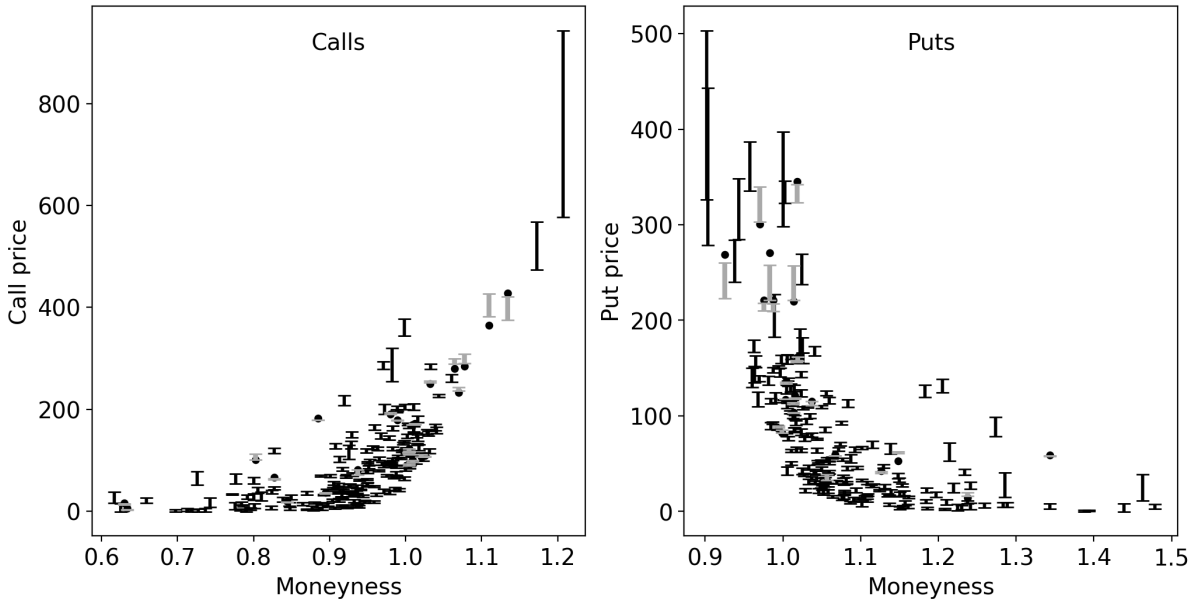

Figure 4: Conformal prediction intervals against option moneyness (S/K) for a random sample of 200 options on Amazon stocks. The bars in gray represent the cases where the prediction interval failed to cover the actual value. The dots represent actual option prices missed by the intervals. The nominal coverage is 0.9. Left plot: calls; Right plot: puts.

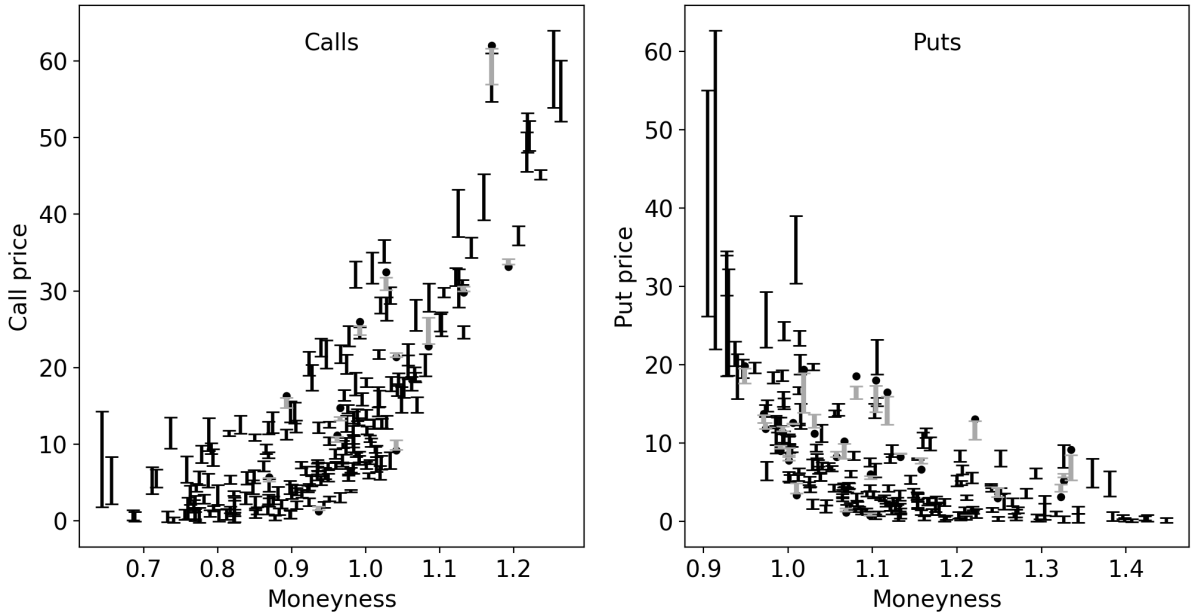

Figure 5: Conformal prediction intervals against option moneyness (S/K) for a random sample of 200 options on PayPal stocks. The bars in gray represent the cases where the prediction interval failed to cover the actual value. The dots represent actual option prices missed by the intervals. The nominal coverage is 0.9. Left plot: calls; Right plot: puts.

provides non-asymptotic coverage guarantees, there are two factors contributing to the finite-sample variability. The first factor arises from the random partitioning of the com-

plete dataset into training and test sets. This affects both conformal and non-conformal models. The second factor of finite-sample variability is due to the random selection of calibration samples from the training data to derive conformity scores. Consequently, the prediction intervals conditioned on the calibration data become random variables. In Vovk (2012), it is shown that, given a training dataset $(X_i, Y_i)_{i=1}^n$ and random calibration samples

$$\Pr(Y_{n+1} \in \mathcal{C}(X_{n+1}|\{(X_i, Y_i)\}_{i=1}^n) \sim \text{Beta}(n+1-k, k), \tag{17}$$

where $k = (n+1)\alpha$. Further details and an illustration can be found in Angelopoulos and Bates (2023). To address this issue, we performed the relevant calculations 100 times using different random splits of the data into training, calibration, and test sets. By averaging the results, we mitigate the impact of finite-sample variability and obtained more reliable estimates for the empirical coverage and median relative widths.

	Call options			
	Empirical coverage		**Relative width**	
	CQR	**NQR**	**CQR**	**NQR**
All companies	0.90	0.69	6%	3%
Amazon	0.90	0.62	9%	4%
AMD	0.90	0.62	10%	5%
Boeing	0.90	0.62	16%	7%
Disney	0.90	0.59	16%	7%
Meta	0.90	0.63	15%	6%
Netflix	0.90	0.60	14%	6%
PayPal	0.90	0.59	15%	7%
Salesforce	0.90	0.60	19%	8%
	Put options			
	Empirical coverage		**Relative width**	
	CQR	**NQR**	**CQR**	**NQR**
All companies	0.90	0.67	7%	3%
Amazon	0.90	0.61	10%	4%
AMD	0.90	0.60	12%	5%
Boeing	0.90	0.58	15%	6%
Disney	0.90	0.56	22%	9%
Meta	0.90	0.57	15%	6%
Netflix	0.90	0.58	16%	6%
PayPal	0.90	0.54	25%	8%
Salesforce	0.90	0.57	19%	7%

Table 6: Empirical coverage and median relative width of the prediction intervals given by the conformal quantile regressions (CQR) and the non-conformal quantile regressions (NQR). Nominal coverage level is 0.9. All figures refer to average values over 100 random splits of the data into train, calibration and test sets.

Table 6 reports the empirical coverage in the test data for each company and option type. Additionally, it shows the results obtained when the models are trained and tested using the full dataset (top row), which comprises 69,990 calls and 50,190 puts. This table reveals a scenario similar to the one observed in the simulation exercise. Specifically, it shows that the non-conformal intervals significantly undercover the actual prices, whereas

the conformalized intervals achieve coverage levels near the nominal level. The median relative width for the full dataset is the lowest since the models have access to more information, enabling them to better map the relationship between option prices and their corresponding characteristics.

5.3 Conditional coverage

The simulation experiment has confirmed that conformal prediction does not provide conditional coverage guarantees. However, despite the absence of exact empirical coverages, we still aim to examine how the interval widths vary across different regions of the input space, without training separate models for each region.

	Call options					
	Empirical coverage			**Relative width**		
	OTM	ATM	ITM	OTM	ATM	ITM
All companies	0.92	0.92	0.80	9%	3%	4%
Amazon	0.92	0.90	0.75	17%	5%	6%
AMD	0.93	0.92	0.81	18%	7%	6%
Boeing	0.93	0.87	0.78	24%	8%	9%
Disney	0.95	0.88	0.79	32%	10%	7%
Meta	0.93	0.88	0.79	26%	7%	8%
Netflix	0.94	0.89	0.76	24%	7%	8%
PayPal	0.93	0.91	0.82	29%	10%	10%
Salesforce	0.93	0.89	0.78	29%	9%	9%
	Put options					
	Empirical coverage			**Relative width**		
	ITM	ATM	OTM	ITM	ATM	OTM
All companies	0.79	0.88	0.92	5%	4%	8%
Amazon	0.74	0.87	0.92	10%	5%	14%
AMD	0.80	0.89	0.92	8%	7%	16%
Boeing	0.79	0.89	0.93	9%	9%	22%
Disney	0.76	0.86	0.92	12%	11%	30%
Meta	0.77	0.88	0.93	8%	8%	21%
Netflix	0.78	0.85	0.92	11%	8%	20%
PayPal	0.81	0.84	0.92	18%	13%	29%
Salesforce	0.79	0.87	0.93	11%	10%	28%

Table 7: Conditional coverages and median relative widths of the conformal prediction intervals for different levels of moneyness (M): out-of-the-money (OTM); at-the-money (ATM); and in-the money (ITM). For call options, the levels of moneyness are OTM: $M < 0.98$, ATM: $0.98 \leq M < 1.02$, and ITM: $M \geq 1.02$. For put options, the levels of moneyness are OTM: $M \geq 1.02$, ATM: $0.98 \leq M < 1.02$, and ITM: $M < 0.98$. The nominal coverage level is 0.9. All figures refer to average values over 100 random splits of the data into train, calibration and test sets.

Table 7 provides the conditional coverages and median relative widths for different levels of moneyness (M). Once again, the test data was divided into three bins based on moneyness (M) to obtain conditional coverages. For call options, the moneyness levels are categorized as out-of-the-money ($M < 0.98$), at-the-money ($0.98 \leq M < 1.02$), and

in-the-money ($M \geq 1.02$). For put options, the moneyness levels are categorized as out-of-the-money ($M \geq 1.02$), at-the-money ($0.98 \leq M < 1.02$), and in-the-money ($M < 0.98$). In-the-money options exhibit greater coverage distortions, tending to undercover actual prices. Nevertheless, they are closer to the nominal level than those provided by the non-conformalized models. On the other hand, the interval widths are generally larger for out-of-the-money options, with some figures approaching 30%.

	Call options					
	Empirical coverage			**Relative width**		
	short	medium	long	short	medium	long
All companies	0.95	0.92	0.83	7%	5%	5%
Amazon	0.94	0.91	0.80	10%	7%	9%
AMD	0.95	0.92	0.83	13%	9%	8%
Boeing	0.94	0.92	0.85	26%	18%	12%
Disney	0.94	0.92	0.86	27%	16%	11%
Meta	0.93	0.93	0.85	24%	15%	10%
Netflix	0.93	0.91	0.81	19%	12%	9%
PayPal	0.94	0.92	0.84	20%	15%	13%
Salesforce	0.94	0.92	0.86	26%	19%	14%
	Put options					
	Empirical coverage			**Relative width**		
	short	medium	long	short	medium	long
All companies	0.95	0.92	0.79	8%	6%	6%
Amazon	0.94	0.90	0.71	12%	8%	10%
AMD	0.95	0.91	0.81	15%	12%	9%
Boeing	0.95	0.93	0.82	25%	14%	10%
Disney	0.95	0.93	0.82	34%	24%	14%
Meta	0.94	0.94	0.83	24%	16%	10%
Netflix	0.93	0.91	0.79	22%	13%	12%
PayPal	0.94	0.92	0.82	32%	27%	17%
Salesforce	0.95	0.92	0.83	28%	22%	12%

Table 8: Conditional coverages and median relative widths of the conformal prediction intervals for different levels of time-to-maturity (τ): short-term ($\tau < 30$ days); medium-term ($30 \leq \tau < 60$ days); and long-term ($\tau \geq 60$ days). The nominal coverage level is 0.9. All figures refer to average values over 100 random splits of the data into train, calibration and test sets.

Table 8 shows the conditional coverages and median relative widths of the conformal prediction intervals for different levels of time-to-maturity (τ): short-term ($\tau < 30$ days), medium-term ($30 \leq \tau < 60$ days), and long-term ($\tau \geq 60$ days). Again, the values reported in the table are averages derived from 100 random splits of the data into training, calibration, and test sets. For short-term and medium-term options, the conditional coverages closely align with the nominal level. However, for long-term options, the conformalized models tend to slightly undercover the data, although not to the same extent as the non-conformalized models. The intervals are generally larger for short-term options. This outcome is somewhat expected since option prices vary sharply near the strike price as the time-to-maturity approaches zero.

6 Conclusions

This paper presents several findings regarding the uncertainty in option price predictions generated by machine learning models. The instrument to achieve this goal was conformalized quantile regression, a model-agnostic approach that constructs prediction intervals, ensuring valid coverage in finite samples without relying on distributional assumptions. Through simulation experiments and an empirical study involving American call and put options on individual stocks, it was shown that the conformalized models indeed provide empirical coverages close to the nominal level. The study uncovered significant variations in the uncertainty associated with the predictions made by the machine learning models. First, larger relative intervals were observed when the number of training observations for the models was smaller. This indicates that the procedure adapts the width of the intervals to ensure empirical coverage close to the nominal level when the model is less certain about its predictions. Second, it was found that out-of-the-money options tend to have wider relative prediction intervals. This suggests that the models are also less confident in predicting prices for these particular options. Finally, wider prediction intervals were obtained for options with shorter time-to-maturity. Indeed, as the time-to-maturity approaches zero, option prices tend to exhibit sharper variations near the strike price.

Naturally, this framework can be used to measure the uncertainty associated with price predictions for other types of assets. For instance, conformal prediction can be used to estimate prediction intervals associated with time series forecasts (Gibbs and Candès, 2021; Xu and Yie, 2021). Therefore, it could be applied in the context of stock price forecasting where understanding and quantifying uncertainty is essential for making informed investment decisions.

Acknowledgements

I would like to express my gratitude to Raquel M. Gaspar for her valuable comments on an early draft of this paper. This work was supported by Fundação para a Ciência e a Tecnologia [grant number UIDB/05069/2020].

References

Anders, U., Korn, O., Schmitt, C. (1998). Improving the pricing of options: a neural network approach. Journal of Forecasting, 17, 369–388.

Andreou, P. C., Charalambous, C., Martzoukos, S.H. (2010). Generalized parameter functions for option pricing. Journal of Banking & Finance, 34(3), 633–646.

Angelopoulos, A.N., Bates, S. (2023). Conformal prediction: A gentle introduction. Foundations and Trends in Machine Learning 16(4), 494–591.

Barber, R.F., Candès, E. J., Ramdas, A., and Tibshirani, R.J. (2021). The limits of distribution-free conditional predictive inference. Information and Inference: A Journal of the IMA, 10(2), 455–482.

Breiman, L., Friedman, J., Stone, C. J., Olshen, R. A. (1983). Classification and Regression Trees. Wadsworth, Belment CA.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5–32.

Kinjal Chaudhari, Ankit Thakkar (2023). Neural network systems with an integrated coefficient of variation-based feature selection for stock price and trend prediction. Expert Systems with Applications 219, 119527.

Chen, T., Guestrin, E. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. San Francisco, USA.

Cox, J.C., Ross, S.A., Rubinstein, M. (1979). Option pricing: A simplified approach. Journal of Financial Economics, 7(3), 229–263.

Deng, S., Su, J., Zhu, Y., Yu, Y., Xiao, C. (2024) Forecasting carbon price trends based on an interpretable light gradient boosting machine and Bayesian optimization. Expert Systems with Applications, 242, 122502.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of Statistics 29, 1189-1232.

Garcia, R., Gençay, R. (2000) Pricing and hedging derivative securities with neural networks and a homogeneity hint. Journal of Econometrics, 94, 93–115, 2000.

Gibbs, I., Candès, E. (2021) Adaptive conformal inference under distribution shift. Advances in Neural Information Processing Systems 34 (NeurIPS 2021).

Golez, B., Jackwerth, J.C. (2012). Pinning in the S&P 500 futures. Journal of Financial Economics 106(3), 566–585.

Grinsztajn, L., Oyallon, E., Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? In Proceedings of NeurIPS 2022 – Neural Information Processing Systems. New Orleans, USA.

Gülmez, B. (2023). Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm. Expert Systems with Applications 227, 120346.

Healy, J.V., Dixon, M., Read, B.J., Cai, F.F. (2003). Confidence in data mining model predictions: a financial engineering application. In: 29th Annual Conference of the IEEE Industrial Electronics Society pp. 1926–1931. Virginia, USA.

Hutchinson, J.M., Lo, A.W., Poggio, T. (1994) A nonparametric approach to pricing and hedging derivative securities via learning networks. The Journal of Finance 49, 851–889.

Ivaşcu, C.-F. (2021) Option pricing using machine learning. Expert Systems with Applications, 163, 113799.

Kanwal, A., Lau, M.F., Ng, S.P.H., Sim, K.Y., Chandrasekaran, S. (2022). BiCuDNNLSTM-1dCNN — A hybrid deep learning-based predictive model for stock price prediction. Expert Systems with Applications 202, 117123.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the NeurIPS 2017 – Neural Information Processing Systems. Long Beach, USA.

Koenker, R., Bassett Jr., G, (1978). Regression Quantiles. Econometrica 46(1), 33-50.

Lei, J., Robins, J., Wasserman, L., (2013). Distribution-free prediction sets. Journal of the American Statistical Association, 108, 278–287.

Lei, J., Wasserman, L. (2014). Distribution-free prediction bands for non-parametric regression. Journal of the Royal Statistical Society B, 76, 71–96.

Malliaris M., Salchenberger, L. (1993). A neural network model for estimating option prices. Journal of Applied Intelligence, 3(3), 193–206.

Meinshausen, N. (2006). Quantile regression forests. Journal of Machine Learning Research 7, 983–999.

Nakayama, S., Shiota, A., Mitani, Y., Watanabe, M. (2023). GIS based JEPX spot prices forecasting system using solar power generation focusing on lowest prices. Energy Reports 9, 240-244.

Papadopoulos, H., Proedrou, K., Vovk, V., and Gammerman, A. (2002). Inductive confidence machines for regression. In Proceedings of Machine Learning: European Conference on Machine Learning, 2002, 345–356. Helsinki, Finland.

Papadopoulos, H., Gammerman, A., Vovk, V. (2008) Normalized nonconformity measures for regression conformal prediction. In Proceedings of the 26th IASTED International Conference on Artificial Intelligence and Applications, 64–69. Innsbruck, Austria.

Park, H., Kim, N. Lee, J. (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), 5227–5237.

Quinlan, J.R., Induction of decision trees. Machine Learning 1, 81–106.

Ribeiro, G.T., Santos, A.A.P., Mariani, V.C., Coelho, L.S (2021). Novel hybrid model based on echo state neural network applied to the prediction of stock price return volatility, Expert Systems with Applications 184, 115490.

Romano, Y., Patterson, E., Candès, E. (2019) Conformalized quantile regression. In Advances in Neural Information Processing Systems, vol. 32, 3543–3553. Vancouver, Canada.

Rumelhart, D.E., Hinton, G.E., Williams, R.J. (1986). Learning representations by back-propagating errors. Nature 323, 533–536.

Steinwart, I., Christmann, A. (2011). Estimating conditional quantiles with the help of the pinball loss. Bernoulli 17(1), 211–225.

Takeuchi, I., Le, Q.V., Sears, T.D., Smola , A.J. (2006). Nonparametric quantile estimation. Journal of Machine Learning Research 7. 1231–1264.

Tibshirani, R. (1996b). A comparison of some error estimates for neural network models. Neural Computation 8, 152–163.

Vapnik, V. (1995). The nature of statistical leaning theory. New York: Springer-Verlag, 168-209.

Vovk, V., Gammerman, A., Shafer, and G. (2005). Algorithmic Learning in a Random World. Springer.

Vovk, V., Nouretdinov, I., Gammerman, A. (2009) On-line predictive linear regression. Annals of Statistcs 37(3), 1566–1590.

Vovk, V. (2012). Conditional validity of inductive conformal predictors. In Proceedings of the Asian Conference on Machine Learning, 25, pp. 475–490. Singapore.

Wang, P. (2011). Pricing currency options with support vector regression and stochastic volatility model with jumps. Expert Systems with Applications 38(1), 1–7.

Xu, C., Xie, Y. (2021). Conformal prediction interval for dynamic time-series. Proceedings of the 38th International Conference on Machine Learning, PMLR 139:11559-11569.