Robust European Call Option Pricing via Linear Regression

Ahmad W. Bitar

Robust European Call Option Pricing via Linear Regression

Ahmad W. Bitar
Laboratory of Computer Science and Digital Society
Université de Technologie de Troyes
Troyes, France
ahmad.bitar@utt.fr

***Abstract*—The one-period trinomial option pricing model is well known in the literature as it considers three possible movement directions of the asset price. However, by equating the price of the option with the self-financing hedging portfolio at maturity, this yields a linear system of three equations with two unknowns that correspond to the coefficients for the delta hedging portfolio. Hence, the trinomial model is said to be incomplete, that is, there exists an infinite number of equivalent martingale measures. To deal with this incompleteness, this paper aims to price options via some robust linear regression techniques in order to mainly handle the problem of outliers that the least squares fails to consider. The proposed robust techniques are evaluated on numerical data, the results of which demonstrate their effectiveness for European call option pricing.**

***Index Terms*—Asset prices, Trinomial model, European call option pricing, Least squares, Robust linear regression.**

I. Introduction

Option pricing theory [1]–[12] is considered one of the most important topics in finance. One of the biggest influences finance had on operation markets started when Myron Scholes met Fisher Black [13] in the fall of 1968 and developed a parametric non-arbitrage option pricing model that accurately prices options with a closed-form formula [10]–[12]. This pioneering work of Black and Scholes lies on the assumption that the market is complete where the equivalent martingale measure is unique, and hence the option can be replicated by the self-financing hedging portfolio of the underlying basic assets in the market. However, one of the major limitations of this model is its lack of robustness since it considers a very specific dynamics about the asset prices.

To move from a parametric to robust European option pricing, the one-period discrete-time binomial model was the first non-arbitrage discrete pricing contribution that was developed by William Sharpe [14] and then formalized by Cox, Ross, and Rubenstein [15]. More precisely, the price of the asset at maturity, denoted as $S_1 > 0$, can only increase or decrease with respect to its price at the initial date $S_0 > 0$. After equating the price of the option with the self-financing hedging portfolio at maturity, this yields a linear system of two equations with two unknowns that correspond to the coefficients for the delta hedging portfolio. As the number of equations and the number of unknowns in the linear system are equal to each other, the delta hedging coefficients, in general, can be determined by a unique analytical solution. However, this model assumes that the asset price can never remain the same for two consecutive discrete time steps (e.g., $S_1 \neq S_0$ almost surely), which does not reflect well the real market.

To complete the binomial model, Boyle [16], [17] derived the one-period discrete-time trinomial option pricing model to incorporate three possible movement directions of the asset price S_1, with respect to S_0: (1) going up with a factor e^u, $u > 0$; (2) remains the same; or (3) going down with a factor e^d, $d < 0$. Hence, pricing options under the trinomial model represent a more realistic but more complex structure than the binomial one. However, equating the price of the option with the self-financing hedging portfolio at maturity yields an overdetermined linear system of three equations with two unknowns that again correspond to the coefficients for the delta hedging portfolio [16]–[18]. As the number of equations in the linear system exceeds that of unknowns, a solution for the delta hedging coefficients, in general, cannot be guaranteed. Hence, the trinomial market model is said to be incomplete [16]–[20], that is, there exists an infinite number of equivalent martingale measures.

To deal with pricing options under the incompleteness of the trinomial model, one can build a self-financing hedging portfolio which does not perfectly replicate the option price. An estimation of the delta hedging coefficients can be done simply using the least-squares technique [21]–[23]. The latter is derived with the idea of minimizing the sum of the squared "errors", that is, to adjust the unknown delta hedging coefficients such that the sum of the squares of the differences between the original and predicted PAYOFF values is minimized. As a result, the least squares estimate is the regression equivalent to the sample mean, which is well known to be very sensitive to outliers [24]–[29] and therefore often does not provide a good fit to most of the data. By outliers, we mean observations that deviate from the general pattern of the data. In addition, the least squares [21] is derived under the Gaussian assumption, which proves its lack of robustness against the outliers. Hence, the detla hedging coefficients will not be estimated precisely and will badly affect the prediction of the option price.

In this paper, we alleviate the effect of outliers in efficiently predicting the PAYOFF and the option price at the initial date. We achieve this by two different option prediction strategies:

1) The first one would be to simply (1) clean the data by

manually removing the influential observations (outliers) that are separated from the general pattern of the data; and (2) use the least squares estimation on the remaining part of the data (that is, after the outliers have been removed) [28];

2) An alternative strategy would be to estimate the delta hedging portfolio coefficients via some robust linear regression techniques in order to deal with the outliers that the least squares fails to consider. More precisely, by taking advantage of some linear regression M-estimators [28], [29], the delta hedging coefficients can be estimated quite precisely, and thus the true (known) PAYOFF values as well as the true (unknown) option price at the initial date will be efficiently predicted.

It should be noted that the two different possible positive factors e^u and e^d by which the asset price S_1 increases or decreases with respect to S_0, respectively, are usually not known in the real market and therefore a single specific value for each of the parameters u and d should be chosen manually to price the corresponding option. As a result, only three possible observations are always used to construct the linear regression model. However, using a very small number of observations (for example, three) might not be sufficient to accurately predict the option price under any regression model. To ensure reliable prediction results, generating enough observations may alleviate this challenge.

However, it is not surprising that the price of the option greatly depends on the selection of u and d. For example, the option price always increases with an increase in the value of u. We believe that generating enough observations may greatly reduce the effect of the choice of u and d on the option price. Instead of selecting a specific value for each of these two parameters, we aim to price options under a large set of values for u and d at once. This will help us generate enough observations that can lead to an accurate estimate of the delta hedging coefficients, especially with such a robust linear regression M-estimator. In addition, this makes pricing the option much less sensitive to such a specific single selection of both parameters u and d, and thus a more realistic option pricing model.

This paper is organized as follows. In Section II we first present a brief overview of the standard (one-period) trinomial option pricing model as well as the estimation of the delta hedging portfolio coefficients via the least squares. A numerical example is also presented to evaluate the least squares in the presence of outliers. The proposed robust European call option prediction strategies are outlined in sec III. Section IV presents numerical experiments to gauge the effectiveness of the proposed prediction strategies. Finally, Section V gives concluding remarks and some directions for future work.

II. The standard one-period trinomial pricing model and the least squares

The one-period discrete-time trinomial market model considers two primary assets: (1) a risk-free asset with price $D_1 = D_0 e^r$, with a constant risk-free rate $r > 0$ and $D_0 > 0$; and (2) a risky asset with price S_1, characterized by three jump behavior as follows:

$$S_1 = \begin{cases} S_0 e^u & \text{with probability } p_u \\ S_0 & \text{with probability } p_0 = 1 - p_u - p_d \,, \\ S_0 e^d & \text{with probability } p_d \end{cases}$$

where $u > 0$, $d < 0$, p_u, $p_d \in (0,\, 1)$, and $p_u + p_d < 1$.

Interestingly, for the market to be arbitrage-free, the following condition is necessary [23]:

$$u > r, \quad \frac{e^u - e^r}{e^u - 1} > p_0 > 0 \,. \tag{1}$$

However, once condition (1) is satisfied, this implies that

$$p_u > 0 \,, \quad p_d > 0 \,, \quad \text{and} \quad p_u + p_d < 1 \,.$$

Hence, the couple $(p_u,\, p_d)$ defines a probability if and only if the condition (1) is satisfied.

Now, assume that an investor's portfolio is constructed by $\Delta_S \in \mathbb{R}$ risky asset and $\Delta_D \in \mathbb{R}$ risk-free asset, where Δ_S and Δ_D represent the unknown delta hedging coefficients that are linearly independent and need to be estimated efficiently. The self-financing portfolio can be defined as

$$C_1 - C_0 = \Delta_S \, (S_1 - S_0) + \Delta_D \, (D_1 - D_0) \,.$$

More precisely, one has:

At initial date :

$$C_0 = \Delta_S \, S_0 + \Delta_D \, D_0 \,.$$

At maturity :

$$\begin{aligned} C_1 &= \Delta_S S_1 + \Delta_D D_1 \\ &= \begin{cases} \Delta_S \, (S_0 e^u) + \Delta_D \, (D_0 e^r) = y_1 \\ \Delta_S \, (S_0) + \Delta_D \, (D_0 e^r) = y_2 \,, \\ \Delta_S \left(S_0 e^d\right) + \Delta_D \, (D_0 e^r) = y_3 \end{cases} \end{aligned}$$

where $y_1 = (S_0 e^u - K)_+$, $y_2 = (S_0 - K)_+$, and $y_3 = \left(S_0 e^d - K\right)_+$, represent the price of the call option at maturity (that is, PAYOFF) with a strike $K > 0$ when the asset price S_1 increases, remains the same, or decreases with respect to S_0, respectively. As we can see, this yields an overdetermined linear system of three equations with only two unknowns (Δ_S and Δ_D). We denote the corresponding vector of dependent variables by $\mathbf{y} = [y_1,\, y_2,\, y_3]^T \in \mathbb{R}^3_+$. In general, the vector $\mathbf{y}$ does not span the column space of the matrix of independent variables $\mathbf{X} = \begin{bmatrix} - - \mathbf{x}_1^T - - \\ - - \mathbf{x}_2^T - - \\ - - \mathbf{x}_3^T - - \end{bmatrix} = \begin{bmatrix} S_0 e^u & D_0 e^r \\ S_0 & D_0 e^r \\ S_0 e^d & D_0 e^r \end{bmatrix} \in \mathbb{R}^{3\times 2}_+$, and hence, a solution for $\mathbf{\Delta} = [\Delta_S,\, \Delta_D]^T \in \mathbb{R}^2$ does not exist if one forces $\mathbf{y}$ to be exactly equal to $\mathbf{X\Delta}$, that is, $\mathbf{y} = \mathbf{X\Delta}$.

A. European call option pricing via the least squares

As it is generally impossible to solve the linear system $\mathbf{y} = \mathbf{X\Delta}$ when the number of equations exceeds that of unknowns, it is possible, however, to project $\mathbf{y}$ into the column space of $\mathbf{X}$. Consider the following linear regression model:

$$\mathbf{y} = \mathbf{X\Delta} + \boldsymbol{\epsilon} \,,$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}\left(\mathbf{0},\, \sigma^2 \mathbf{I}_3\right)$ with $\sigma > 0$, $\mathbf{I}_3$ is an identity matrix of size 3×3, and $\mathbf{y} \sim \mathcal{N}\left(\mathbf{X}\boldsymbol{\Delta},\, \sigma^2 \mathbf{I}_3\right)$. Maximizing the log-likelihood with respect to $\boldsymbol{\Delta}$ under the Gaussian assumption gives:

$$\hat{\boldsymbol{\Delta}} = \underset{\boldsymbol{\Delta}}{argmin} \left\{ \sum_{i=1}^{3} \left(y_i - \mathbf{x}_i^T \boldsymbol{\Delta}\right)^2 \right\}$$
$$= \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{y} \;=\; \widehat{\begin{bmatrix} \Delta_S \\ \Delta_D \end{bmatrix}} \in \mathbb{R}^2 .$$

The least squares predictor of the true (known) PAYOFF can be given as:

$$\hat{\mathbf{y}} = \left(\mathbf{X}\hat{\boldsymbol{\Delta}}\right)_+ .$$

Once $\boldsymbol{\Delta}$ is estimated, the least squares predictor of the true (unknown) call option price at the initial date (that is, C_0) is:

$$\hat{C}_0 = \hat{\Delta}_S S_0 + \hat{\Delta}_D D_0 .$$

B. Problem of outliers with the least Squares

Fix $r = 2\%$, $D_0 = \$1$, $S_0 = \$100$, $K = S_0$, $u = 0.1$, and $d = -u$. We aim to compare three different option pricing prediction models:

1) Least squares on $\{y_1, y_2, y_3\}$: the least squares is applied on the whole set of 3 observations $\{y_1, y_2, y_3\}$;
2) Least squares on $\{y_1, y_2\}$: we assume that the observation y_3 badly affects the performance of the least squares on predicting the vector $\mathbf{y}$. We propose to manually remove the observation y_3, and then apply the least squares on the remaining part of the data $\{y_1, y_2\}$;
3) Least squares on $\{y_1, y_3\}$: we manually remove y_2 and then apply the least squares on the remaining part of the data $\{y_1, y_3\}$.

Figure 1 (left side) shows the difference between $\mathbf{y}$ and $\hat{\mathbf{y}}$ for the three models. We can observe that the predicted PAYOFF observations with model "Least squares on $\{y_1, y_2\}$" coincide with the true ones, and the least squares line with model "Least squares on $\{y_1, y_2, y_3\}$" is skewed towards observation y_3. Therefore, applying least squares after removing the observation y_3 potentially decreases the PAYOFF prediction error compared to the other two models.

III. Main contribution

We present in Subsection III-A a method to construct an option pricing model that considers $n \geq 1$ possible values for each of the parameters $u > 0$ and $d < 0$, and thus allows us to construct a linear regression model with $2n + 1$ observations. In addition, we evaluate the least squares in a regression with $2n+1$, $n \geq 1$, observations to detect the nature of the outliers. Once these outliers are identified, one can simply remove them and then re-apply the least squares on the remaining part of the data. Finally, in Subsection III-B, we aim to exploit some robust linear regression M-estimators [28], [29] to accurately estimate the (unknown) vector $\boldsymbol{\Delta}$ with an automatic removal of outliers.

A. Increasing the number of observations to $2n + 1$, $n \geq 1$

In order to estimate the option price at the initial date independently of any such a specific single value for each of the parameters $u > 0$ and $d < 0$, we propose to replace u and d by a set of $n \geq 1$ possible values $\{u_l\}_{l\in[1,\,n]}$ with $u_1 > u_2 > \cdots > u_n > r$ and $\{d_l\}_{l\in[1,\,n]}$ with $d_n < d_{n-1} < \cdots < d_1 < 0$, respectively, with $n \in \mathbb{N}\backslash\{0\}$. As a result, one has to expect $2n + 1$ different values for S_1 and the possibility to construct an overdetermined linear system of $2n + 1$ equations with only 2 unknowns. We have:

$$S_1 = \begin{cases} S_0 e^{u_1} & \text{with } p_{u_1} > 0 \\ \quad\vdots & \\ S_0 e^{u_n} & \text{with } p_{u_n} > 0 \\ S_0 & \text{with } p_0 > 0 \\ S_0 e^{d_1} & \text{with } p_{d_1} > 0 \\ \quad\vdots & \\ S_0 e^{d_n} & \text{with } p_{d_n} > 0 \end{cases}, \; C_1 = \begin{cases} \Delta_S \left(S_0 e^{u_1}\right) + \Delta_D D_1 = y_1 \\ \Delta_S \left(S_0 e^{u_2}\right) + \Delta_D D_1 = y_2 \\ \qquad\vdots \\ \Delta_S \left(S_0 e^{u_n}\right) + \Delta_D D_1 = y_n \\ \boxed{\Delta_S S_0 + \Delta_D D_1 = y_{n+1}} \\ \Delta_S \left(S_0 e^{d_1}\right) + \Delta_D D_1 = y_{n+2} \\ \Delta_S \left(S_0 e^{d_2}\right) + \Delta_D D_1 = y_{n+3} \\ \qquad\vdots \\ \Delta_S \left(S_0 e^{d_n}\right) + \Delta_D D_1 = y_{2n+1} \end{cases},$$

where $p_0 = 1 - p_{u_1} - \cdots - p_{u_n} - p_{d_1} - \cdots - p_{d_n} > 0$, $y_1 = \left(S_0 e^{u_1} - K\right)_+$, $y_2 = \left(S_0 e^{u_2} - K\right)_+$, $y_n = \left(S_0 e^{u_n} - K\right)_+$, $y_{n+1} = \left(S_0 - K\right)_+$, $y_{n+2} = \left(S_0 e^{d_1} - K\right)_+$, $y_{n+3} = \left(S_0 e^{d_2} - K\right)_+$, and $y_{2n+1} = \left(S_0 e^{d_n} - K\right)_+$. We can observe that the standard (one-period) trinomial option pricing model is a special case when $n = 1$.

What happens to the least squares with $2n + 1$ observations?

Fix $r = 2\%$, $\epsilon = 10^{-6}$, $S_0 = \$100$, $K = S_0$, and $D_0 = \$1$. Instead of specifying a single specific value for u as we did in the example in Subsection II-B, we predict the PAYOFF with the least squares under a set of equally spaced 5 possible values $\{u_l\}_{l\in[1,\,5]}$, with $u_1 > \cdots > u_5 > r$. We choose $u_5 = r + \epsilon$ and $u_1 = 1$. We also fix $d_5 = -u_1$, $d_4 = -u_2$, $d_3 = -u_3$, $d_2 = -u_4$, and $d_1 = -u_5$. Hence, we have 11 possible PAYOFF values, that is, $\mathbf{y} = [y_1, y_2, \cdots, y_{11}]^T \in \mathbb{R}_+^{11}$. Figure 1 (right side) shows the difference between the true PAYOFF values and the predicted ones by comparing the following two option pricing prediction models:

1) Least squares on $\{y_i,\, i \in [1,\, 11]\}$: we directly apply the least squares on the whole set of 11 observations;
2) Least squares on $\{y_i,\, i \in [1,\, 11] \,|\, y_i > 0\}$: we assume all the observations which are equal to zero correspond to the influential observations (outliers), and thus, the least squares is only applied on the remaining part of the data (that is, on the set of observations that are > 0).

From Figure 1 (right side), we can clearly observe that the model "Least squares on $\{y_i,\, i \in [1,\, 11] \,|\, y_i > 0\}$" achieves the lowest PAYOFF prediction error and so its predicted PAYOFF observations coincide with the true ones.

B. Estimation of $\boldsymbol{\Delta}$ via robust linear regression M-estimators

Recall that the least squares is derived under the Gaussian assumption which does not take into consideration the outliers. In order not to consider any specific kind of distribution (e.g.,

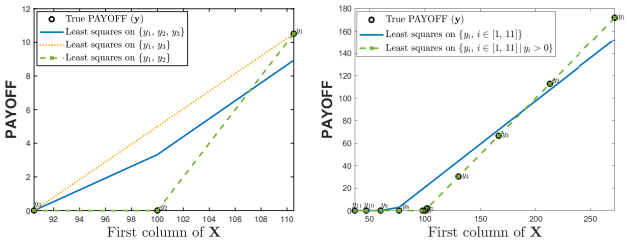

Fig. 1: The difference between $\mathbf{y}$ and $\hat{\mathbf{y}}$ when $K = S_0$.

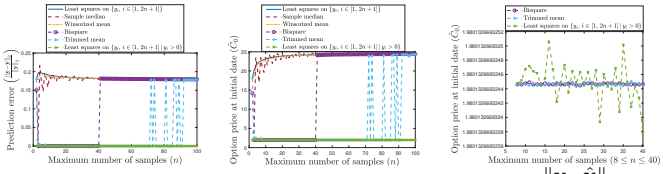

Fig. 2: From left to right: the prediction error $\frac{\|\mathbf{y}-\hat{\mathbf{y}}\|_2}{\|\mathbf{y}\|_2}$, the predicted European call option price at the initial date ($\hat{C}_0$), the predicted $\hat{C}_0$ for bisquare, trimmed mean, and "Least squares on $\{y_i,\, i \in [1,\, 2n+1] \,|\, y_i > 0\}$" when $8 \le n \le 40$.

Gaussian), we assume that the y_i's, $i \in [1,\, 2n+1]$, $n \ge 1$, are independent and not identically distributed with any density of the form $\frac{1}{\sigma} f_0\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right)$. Maximizing the log-likelihood with respect to $(\boldsymbol{\Delta},\, \sigma)$ is thus equivalent to the following (minimization) optimization problem [28]:

$$\begin{aligned} &\max_{\boldsymbol{\Delta},\sigma} \left\{ -(2n+1)\, log\,(\sigma) + \sum_{i=1}^{2n+1} log\left(f_0\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right)\right)\right\} \\ &\quad \equiv \min_{\boldsymbol{\Delta},\sigma} \left\{ (2n+1)\, log\,(\sigma) + \sum_{i=1}^{2n+1} \rho\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right)\right\}, \end{aligned}$$

where $-log\left(f_0\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right)\right) = \rho\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right)$ for any non-constant function $\rho\,(.)$. One can clearly observe that the least squares is a special case when $\rho\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right) = \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right)^2$. Note that the function $\rho\,(.)$ must be chosen to respect several conditions (see definition 2.1 on page 31 of [28] for details).

IV. Experiments

We compare six different option pricing prediction models:

1) Least squares on $\{y_i,\, i \in [1,\, 2n+1]\}$: we directly apply the least squares on the whole set of $2n+1$ observations;
2) Sample median [28]–[31]: $\rho\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right) = \left|\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right|$;
3) Winsorized mean [28]:
$$\rho\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right) = \begin{cases} \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right)^2 & \text{if } \left|\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right| \le \kappa \\ 2\kappa \left|\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right| - \kappa^2 & \text{if } \left|\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right| > \kappa \end{cases};$$
4) Bisquare [28]:
$$\rho\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right) = \begin{cases} 1 - \left(1 - \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma\kappa}\right)^2\right)^3 & \text{if } \left|\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right| \le \kappa \\ 1 & \text{if } \left|\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right| > \kappa \end{cases};$$
5) Trimmed mean [28], [29]:
$$\rho\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right) = \begin{cases} \frac{1}{2}\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right)^2 & \text{if } \left|\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right| \le \kappa \\ \frac{1}{2}\kappa^2 & \text{if } \left|\frac{y_i - \mathbf{x}_i^T \boldsymbol{\Delta}}{\sigma}\right| > \kappa \end{cases};$$
6) Least squares on $\{y_i,\, i \in [1,\, 2n+1] \,|\, y_i > 0\}$: All observations that are equal to zero are eliminated, and the least squares is only applied on the remaining part of the data (that is, on the set of observations that are > 0). When only one observation is strictly positive among the $2n+1$ observations (that is, only $y_1 > 0$), the least squares will then be applied on $\{y_1,\, y_2\}$.

Fix $r = 2\%$, $D_0 = \$1$, and we assume that the asset price at the initial date is $S_0 = \$100$. A call on the asset is available with a strike price $K = S_0$, expiring at maturity. We use CVX in MATLAB 2021a to estimate $\boldsymbol{\Delta}$ for the sample median, and the *robustfit* package to simultaneously estimate both $\boldsymbol{\Delta}$ and σ for the winsorized mean, bisquare, and trimmed mean. To choose the best parameter value of κ for the winsorized mean, bisquare, and trimmed mean, we decide to vary κ between 1 and 5 by step of 10^{-4}. The value of κ that minimizes the criterion $\frac{\|\mathbf{y}-\hat{\mathbf{y}}\|_2}{\|\mathbf{y}\|_2}$ will be selected as optimal.

Fix $\epsilon = 10^{-6}$, $u_1 = 1$, and $u_n = r + \epsilon$. The six option pricing prediction models are evaluated on a large set of $\{u_l\}_{l \in [1,\, n]}$ with $u_1 > u_2 > \cdots > u_n > r$ and $\{d_l\}_{l \in [1,\, n]}$ with $d_n < d_{n-1} < \cdots < d_1 < 0$, both containing equal spaced $n > 1$ values. We consider $d_n = -u_1$, $d_{n-1} = -u_2$, $\cdots$, $d_1 = -u_n$. Figure 2 shows, as a function of $n \in \{2, \cdots, 100\}$ in steps of 1, the PAYOFF prediction error and the prediction results of C_0. We can observe that the bisquare and the trimmed mean achieve a negligible PAYOFF prediction error only when $n \in [3,\, 40]$ and $n \in [8,\, 71]$, respectively. In these two specific ranges of n, the predicted option price $\hat{C}_0$ is about \$1.98013. It is true that "Least squares on $\{y_i,\, i \in [1,\, 2n+1] \,|\, y_i > 0\}$" always achieve the lowest PAYOFF prediction error for all values of n, it can be seen, however, that it presents higher variability in the values of $\hat{C}_0$ compared to the bisquare and trimmed mean (see the third plot on the right side of Figure 2).

V. Discussion and Future Work

Two different European call option pricing prediction strategies via the trinomial model are developed mainly to alleviate the problem of outliers that least squares does not take into account. The first strategy applies the least squares but after the removal of the outliers; whereas the second one exploits some robust linear regression M-estimators. Both strategies are evaluated on numerical experiments, the results of which demonstrate their effectiveness in pricing European call options. Regarding future enhancements, a likely first step would be to improve the performance of the bisquare and the trimmed mean when $n > 40$ and $n > 71$, respectively. A second step would be to estimate $\boldsymbol{\Delta}$ using deep learning techniques (e.g., standard neural networks) when n is quite large. Other promising avenues for further research include (1) the extension of the proposed work to the multi-period case [23], [32], and (2) the improvement of financial risk management by buying European options on assets and exchange rates [33] at the same time to resist any potential fall of the domestic currency value.

REFERENCES

[1] S. Calogero, *A First Course in Options Pricing Theory*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2023. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611977646

[2] L. Bachelier, "Théorie de la spéculation," *Annales scientifiques de l'École Normale Supérieure*, vol. 3e série, 17, pp. 21–86, 1900. [Online]. Available: http://www.numdam.org/articles/10.24033/asens.476/

[3] C. M. Sprenkle, "Warrant prices as indicators of expectations and preferences," *Yale economic essays*, vol. 1, no. 2, pp. 178–231, 1961.

[4] R. Rosett, "Estimating the utility of wealth from call options transactions," *Doctoral dissertation, University of Rochester*, 1963.

[5] A. Ayres, "Risk aversion in the warrants market," *Indus. Management Review*, pp. 497–505, 1967.

[6] A. J. Boness, "Elements of a theory of stock-option value," *Journal of Political Economy*, vol. 72, no. 2, pp. 163–175, 1964. [Online]. Available: http://www.jstor.org/stable/1828962

[7] P. A. Samuelson, *Rational Theory of Warrant Pricing*. Cham: Springer International Publishing, 2015, pp. 195–232.

[8] P. A. Samuelson and R. C. Merton, "A complete model of warrant pricing that maximizes utility," 1969. [Online]. Available: https://api.semanticscholar.org/CorpusID:153453511

[9] R. Merton, "Continuous-time speculative processes: Appendix to paul a. samuelson's "mathematics of speculative price"," *SIAM Rev.*, vol. 15, pp. 34–38, 01 1973.

[10] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *Journal of Political Economy*, vol. 81, no. 3, pp. 637–654, 1973. [Online]. Available: http://www.jstor.org/stable/1831029

[11] R. C. Merton, "Theory of rational option pricing," *The Bell Journal of Economics and Management Science*, vol. 4, no. 1, pp. 141–183, 1973. [Online]. Available: http://www.jstor.org/stable/3003143

[12] T. Crack, *Basic Black-Scholes: Option Pricing and Trading*. T.F. Crack, 2009. [Online]. Available: https://books.google.fr/books?id=E24CPwAACAAJ

[13] M. Scholes. (2018) A conversation with myron scholes. Youtube. [Online]. Available: https://www.youtube.com/watch?v=zooUe_oDSpo

[14] W. Sharpe, *Investments*, ser. Prentice-Hall International editions. Prentice-Hall, 1978. [Online]. Available: https://books.google.fr/books?id=4UgWAQAAMAAJ

[15] J. C. Cox, S. A. Ross, and M. Rubinstein, "Option pricing: A simplified approach," *Journal of Financial Economics*, vol. 7, no. 3, pp. 229–263, 1979.

[16] P. P. Boyle, "Option valuation using a three jump process," 1986. [Online]. Available: https://api.semanticscholar.org/CorpusID:150764610

[17] P. Boyle, "A lattice framework for option pricing with two state variables," *The Journal of Financial and Quantitative Analysis*, vol. 23, no. 1, pp. 1–12, 1988. [Online]. Available: http://www.jstor.org/stable/2331019

[18] Y. Kuang, C. Lin, and N. Yang, "Completing market with one step trinomial model: A european call approach," in *Proceedings of the 2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022)*. Atlantis Press, 2022, pp. 2384–2388. [Online]. Available: https://doi.org/10.2991/aebmr.k.220307.390

[19] J. Ruf, "Lecture notes for 'discrete-time models for the pricing of options'," 2021.

[20] Y. Braouezec, "How fundamental is the one-period trinomial model to european option pricing bounds. a new methodological approach," *Finance Research Letters*, vol. 21, pp. 92–99, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1544612316302823

[21] S. Stigler, *The History of Statistics: The Measurement of Uncertainty Before 1900*. Harvard University Press, 1986. [Online]. Available: https://books.google.fr/books?id=M7yvkERHIIMC

[22] H. Takahashi, "A note on pricing derivartives in an incomplete market," *Graduate School of Economics, Hitotsubashi University, Discussion Papers*, 01 2000.

[23] B. Johan, "The trinomial asset pricing model," *GInstitutionen for matematiska vetenskaper, Goteborgs universitet, Thesis report*, 01 2016.

[24] S. Weisberg, *Applied Linear Regression*, ser. Wiley Series in Probability and Statistics. Wiley, 2005. [Online]. Available: https://books.google.fr/books?id=xd0tNdFOOjcC

[25] D. Belsley, E. Kuh, and R. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, ser. Wiley Series in Probability and Statistics. Wiley, 2005. [Online]. Available: https://books.google.fr/books?id=GECBEUJVNe0C

[26] ——, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, ser. Wiley Series in Probability and Statistics. Wiley, 1980. [Online]. Available: https://books.google.fr/books?id=ALjuAAAAMAAJ

[27] *Wiley Series in Probability and Mathematical Statistics*. John Wiley Sons, Ltd, 1988, pp. i–vi. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470316764.oth1

[28] R. A. Maronna, *Robust Statistics: Theory and Methods (with R)*. John Wiley Sons: Wiley Series in Probability and Statistics, 2019. [Online]. Available: https://doi.org/10.1002/9781119214656

[29] P. J. Huber, "Robust Estimation of a Location Parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73 – 101, 1964. [Online]. Available: https://doi.org/10.1214/aoms/1177703732

[30] I. Barrodale and F. D. K. Roberts, "An improved algorithm for discrete l1 linear approximation," *SIAM Journal on Numerical Analysis*, vol. 10, no. 5, pp. 839–848, 1973. [Online]. Available: http://www.jstor.org/stable/2156318

[31] S. Portnoy and R. Koenker, "The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators," *Statistical Science*, vol. 12, no. 4, pp. 279 – 300, 1997. [Online]. Available: https://doi.org/10.1214/ss/1030037960

[32] A. Pascucci and W. Runggaldier, *Financial Mathematics: Theory and Problems for Multi-period Models*, ser. UNITEXT. Springer Milan, 2012. [Online]. Available: https://books.google.fr/books?id=WD6LkQEACAAJ

[33] A. W. Bitar, "Linear regression for currency European call option pricing in incomplete markets," Dec. 2024, working paper or preprint. [Online]. Available: https://utt.hal.science/hal-04815308