



HAL
open science

Robust European Call Option Pricing via Linear Regression

Ahmad W. Bitar

► **To cite this version:**

| Ahmad W. Bitar. Robust European Call Option Pricing via Linear Regression. 2024. hal-04754957

HAL Id: hal-04754957

<https://utt.hal.science/hal-04754957v1>

Preprint submitted on 26 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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 to 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, and 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] is considered one of the most important topics in finance. It has initially began with the French mathematician Louis Bachelier in 1900 after he developed an option pricing formula with the assumption that the asset prices increments are independent and follow a Brownian motion [2]. Since then, several notable contributions for pricing options, under the difficulty of estimating certain non-observable parameters, have been developed with Sprenkle in 1961 [3], Rosett in 1963 [4], Ayres in 1963 [5], Boness in 1964 [6], Samuelson in 1965 [7], Samuelson et al. in 1969 [8], and Merton in 1972 [9]. However, one of the biggest influence finance had on operation markets started when Myron Scholes met Fisher Black in the fall of 1968 and began working together to discover one of the most influential mathematical model that accurately prices options¹. This Black and Scholes option pricing model appeared in the journal of political economy in 1973 [11], supported by another article on the theory of rational option pricing that was published by Robert Merton in the same year [12].

A European call option is basically the right but not the obligation to buy an asset only at maturity and at a fixed price (known as the strike price) specified at the initial date of the contract. At maturity, there is no need for any mathematical formulation to price an option since its value would simply be equal to (asset price at maturity – strike price)₊ and which is

well known by the PAYOFF. By replacing the asset price at maturity by its current price (at any time before maturity), the strike price by its present value, and by determining the probability that the call option will finishes in the money (after making certain assumptions about how the asset price will behave), Black and Scholes came up with a parametric non-arbitrage option pricing model that accurately prices options with a closed-form formula that does not depend on the risk aversion of the investor. 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 on 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 in 1978 by William Sharpe and then formalized in 1979 by Cox, Ross, and Rubenstein [13]. More precisely, the asset price at maturity, denoted as $S_1 \geq 0$, can only goes up or down with respect to its price at initial date (that is, $S_0 \geq 0$). After equating the price of the option to 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 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) and which does not reflect well the real market.

To complete the binomial model, the one-period discrete-time trinomial option pricing model was derived by Phelim Boyle in 1986 [14] 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 represents a more realistic but more complex structure than the binomial one. However, equating the price of the option to the self-financing hedging portfolio at maturity yields an over-determined linear system of three

¹See the online conversation [10] with Myron Scholes who briefly explained his work with Fisher Black.

equations with two unknowns that correspond, again, to the coefficients for the delta-hedging portfolio. As the number of equations in the linear system exceeds that of unknowns, a solution for the delta-hedging coefficients, in general, can not be guaranteed. Hence, the trinomial market model is said to be incomplete, 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 replicates the option price. An estimation to the delta-hedging coefficients can simply be done via the least squares technique [15], [16]. 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 estimates is the regression equivalent to the sample mean which is well known to be very sensitive to outliers², and therefore, often fails to provide good fits to the bulk of the data. In addition, the least squares is derived under the Gaussian assumption, and which proves its lack of robustness against the outliers. Hence, the delta-hedging coefficients will not be estimated precisely and which has a major bad effect on the prediction of the option price.

In this paper, we mainly aim to alleviate the effect of outliers in order to efficiently predict the PAYOFF as well as the option price at initial date. We achieve this by two different option prediction strategies:

- 1) The first one 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 exploiting some linear regression M -estimators [17], [18], the delta-hedging coefficients can be estimated quite precisely (that is, their estimated values are too close to those of the true unknown ones), and thus, the true (known) PAYOFF values as well as the true (unknown) option price at initial date will be predicted efficiently;
- 2) An alternative strategy would be to simply (1) clean the data by manually removing the observations (outliers) that are separated from the bulk of the data; and (2) use the least squares estimation on the remaining data after the outliers have been removed.

It is worthy to note that the two different possible positive factors e^u and e^d by which the asset price S_1 goes up or down with respect to S_0 , respectively, are usually not known in the real market, and thus, a single specific values for both the parameters u and d should be chosen manually in order 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 little number of observations (ex. three) might be not sufficient to accurately predict the

²By outliers we mean the observations that deviate from the general pattern of the data. For example, one and/or several points that deviate from the linear regression line where all the remaining points are located.

option price under any regression model. Hence, to ensure reliable prediction results, generating enough observations may alleviate this challenge.

It is not surprising that the option price greatly depends on the selection for both u and d . For example, the option price always increases with the increase of the value for u . We believe that generating enough observations may greatly reduce the effect of the choice for both u and d on the option price. Hence, 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 to generate enough observations that can lead to an accurate estimation of the delta-hedging coefficients especially with such a robust linear regression M -estimator. In addition, this renders pricing the option much less sensitive to such a specific single selection of both the two parameters u and d , and thus, a more realistic option pricing model.

This paper is structured along the following lines. First comes in section II a brief overview of the trinomial option pricing model on assets as well as of the estimation of the delta-hedging portfolio coefficients via the least squares technique. The proposed two option prediction strategies for a robust delta-hedging estimation are outlined in sec III. Section IV presents the experiments to gauge the effectiveness of the proposed prediction strategies for European call option pricing. Finally, section V gives concluding remarks and some directions for future work.

Notations: throughout this paper, we depict vectors in lowercase boldface papers and matrices in uppercase boldface papers. The notation $(\cdot)^T$, $(\cdot)_+$, $|\cdot|$, and $\widehat{(\cdot)}$ stand for the transpose, positive part, absolute value, and estimated/predicted value, respectively. A variety of norms on vectors will be used. For instance, Δ is a vector, and Δ_j is the j^{th} element. The vector l_2 and l_1 - norms are defined by $\|\Delta\|_2 = \sqrt{\sum_j \Delta_j^2}$ and $\|\Delta\|_1 = \sum_j |\Delta_j|$, respectively.

II. THE STANDARD ONE-PERIOD TRINOMIAL PRICING MODEL AND THE LEAST SQUARES

In this section, we first overview the standard one-period trinomial option pricing model in subsection II-A, and in which we briefly explain how can one generate an over-determined linear system of three equations with two unknowns that correspond to the delta-hedging portfolio coefficients. In order to estimate those unknown coefficients, we present the least squares technique in subsection II-B. Finally subsection II-C proves, with a simple example, the sensitivity of the least squares estimation to even a single outlier.

A. The (standard) one-period trinomial option pricing model

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 risk-free rate $r \geq 0$ and $D_0 > 0$; and

(2) a risky asset with price S_1 , characterized by three jump behavior as follows:

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

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 [16]:

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

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 well a probability if and only if condition (1) is satisfied.

Now, assume 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 which 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 :

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

where $y_1 = (S_0 e^u - K)_+$, $y_2 = (S_0 - K)_+$, and $y_3 = (S_0 e^d - K)_+$, represent the value of the call option at maturity (that is, the PAYOFF) when the asset price S_1 goes up, remains the same, or goes down with respect to S_0 , respectively. As we can see, this yields an over-determined 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

$$\text{variables } \mathbf{X} = \begin{bmatrix} - & -\mathbf{x}_1^T & - \\ - & -\mathbf{x}_2^T & - \\ - & -\mathbf{x}_3^T & - \end{bmatrix} = \begin{bmatrix} S_0 e^u & D_0 e^r \\ S_0 & D_0 e^r \\ S_0 e^d & D_0 e^r \end{bmatrix} \in \mathbb{R}_+^{3 \times 2},$$

and hence, a solution for $\Delta = [\Delta_S, \Delta_D]^T \in \mathbb{R}^2$ does not exist if one forces \mathbf{y} to be exactly equal to $\mathbf{X}\Delta$, that is, $\mathbf{y} = \mathbf{X}\Delta$.

B. Estimation of Δ 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:

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

where $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_3)$ with \mathbf{I}_3 is an identity matrix of size 3×3 , and $\mathbf{y} \sim \mathcal{N}(\mathbf{X}\Delta, \sigma^2 \mathbf{I}_3)$.

The likelihood under the Gaussian assumption is defined as:

$$L(y_1, y_2, y_3; \Delta, \sigma) = \prod_{i=1}^3 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_i - \mathbf{x}_i^T \Delta)^2}.$$

Maximizing the log-likelihood w.r.t. Δ under the Gaussian assumption gives the usual estimation known as the ordinary least squares estimator. More precisely, we have:

$$\begin{aligned} \hat{\Delta} &= \underset{\Delta}{\operatorname{argmin}} \left\{ \sum_{i=1}^3 (y_i - \mathbf{x}_i^T \Delta)^2 \right\} \\ &= \underset{\Delta}{\operatorname{argmin}} \left\{ \|\mathbf{y} - \mathbf{X}\Delta\|_2^2 \right\}, \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = \widehat{\begin{bmatrix} \Delta_S \\ \Delta_D \end{bmatrix}} \in \mathbb{R}^2. \end{aligned} \quad (2)$$

Hence, the least squares estimation of the true (known) PAY-OFF can be given as:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\Delta}. \quad (3)$$

As the option price should always be positive, thus, it would be more realistic to consider the rectified linear unit of $\hat{\mathbf{y}}$. That is:

$$(\hat{\mathbf{y}})_+.$$

Once Δ is estimated, the least squares estimator of the true (unknown) call option price at initial date (that is, C_0) can be given as:

$$\hat{C}_0 = \hat{\Delta}_S S_0 + \hat{\Delta}_D D_0. \quad (4)$$

C. Problem of outliers with the least Squares: some numerical investigations

Let fix $r = 2\%$, $D_0 = \$1$, $S_0 = \$100$, and $K = S_0$. We assume that both u and d are known and so we aim to test on different values for $u \in \{0.1, 0.3, 0.5, 0.8, 1, 2, 3, 5, 10\}$, and we choose $d = -u$. Table I presents the true PAYOFF vector $\mathbf{y} = [y_1, y_2, y_3]^T$, the predicted PAYOFF vector via the least squares $\hat{\mathbf{y}} = [\hat{y}_1, \hat{y}_2, \hat{y}_3]^T$ and its rectified linear unit version $(\hat{\mathbf{y}})_+$, the prediction error $\frac{\|\mathbf{y} - (\hat{\mathbf{y}})_+\|_2}{\|\mathbf{y}\|_2}$, as well as the prediction via the least squares of the option price at initial date \hat{C}_0 , for the nine different values for u . We can clearly observe that the least squares behaves badly for all the values of u except when u is large enough where the prediction error is very small, and so the option price at initial date (that is, C_0) can be predicted efficiently. However, the option price increases progressively up to a certain limit (for example around \$50.98) with the increase of the value for u , and thus, the option price will be very large when u is large enough.

| u | S_1 | \mathbf{y} | $(\hat{\mathbf{y}})_+$ | $\frac{\ \mathbf{y} - (\hat{\mathbf{y}})_+\ _2}{\ \mathbf{y}\ _2}$ | \hat{C}_0 (in \$) |
|-----|------------------------------|----------------------|----------------------------|--|---------------------|
| 0.1 | [110.5170, 100, 90.4837] | [10.5170, 0, 0] | [8.9363, 3.3277, 0] | 0.3503 | 4.3178 |
| 0.3 | [134.9858, 100, 74.08182] | [34.9858, 0, 0] | [30.7929, 9.8528, 0] | 0.3060 | 10.8429 |
| 0.5 | [164.8721, 100, 60.6530] | [64.8721, 0, 0] | [58.8285, 16.0078, 0] | 0.2637 | 16.9978 |
| 0.8 | [222.5540, 100, 44.9328] | [122.5540, 0, 0] | [115.0617, 24.1670, 0] | 0.2064 | 25.1570 |
| 1 | [271.8281, 100, 36.7879] | [171.8281, 0, 0] | [164.0932, 28.7605, 0] | 0.1733 | 29.7505 |
| 2 | [738.9056, 100, 13.5335] | [638.9056, 0, 0] | [633.8339, 42.5468, 0] | 0.0670 | 44.5073 |
| 3 | [2008.5536, 100, 4.9787] | [1908.5536, 0, 0] | [1906.3057, 47.3987, 0] | 0.0248 | 48.3887 |
| 5 | [14841.3159, 100, 0.6738] | [14741.3159, 0, 0] | [14740.9835, 49.6608, 0] | 0.0033 | 50.6509 |
| 10 | [2202646.5794, 100, 0.00454] | [2202546.5794, 0, 0] | [2202546.5772, 49.9977, 0] | 2.2699×10^{-6} | 50.9877 |

TABLE I
ESTIMATION OF THE OPTION PRICE VIA THE LEAST SQUARES METHOD.

Figure 1 exhibits the difference between \mathbf{y} and $\hat{\mathbf{y}}$, as well as the difference between \mathbf{y} and $(\hat{\mathbf{y}})_+$, for $u = 0.1$ and $d = -u$. From figure 1, we can observe that the entire least squares regression red dashed line is skewed towards the observation y_3 . Hence, the least squares has considered y_3 as an outlier. In other words, the bulk of the data is only formed by y_1 and y_2 , whereas y_3 is the observation that deviates from it.

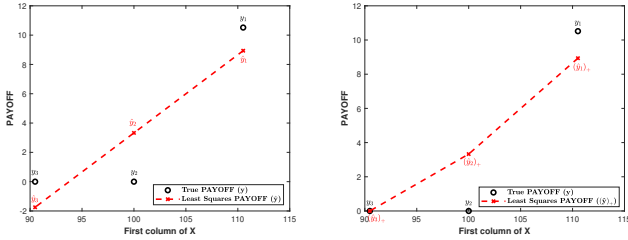


Fig. 1. Performance of the least squares hedging portfolio at maturity.

III. ROBUST ESTIMATION OF THE DELTA-HEDGING PORTFOLIO COEFFICIENTS

In this section, we develop two different prediction strategies for the European call option price in order to alleviate the effect of outliers that the least squares fails to consider. The first strategy is presented in subsection III-B in which we exploit the idea behind the robust M -estimators for linear regression analysis [17], [18]; whereas the second one is presented in subsection III-C and which simply aims to apply the least squares after manually cleaning the data from the outliers that are identified. However, we already mentioned in the introduction that training a regression model with only three observations (ex. the matrix \mathbf{X} has only three rows) might be not sufficient to accurately estimate the unknown delta-hedging coefficients. That is why, we start this section (that is, in subsection III-A) by presenting a method that allows us to construct a linear regression model with $2n + 1$, $n \geq 1$, observations.

A. Increasing the number of observations from 3 to $2n + 1$, $n \geq 1$

In order to estimate the option price at initial date, that is, C_0 , independently from any such a specific single value for each of the parameters $u > 0$ and $d < 0$, we propose to

replace the single possible values for u and d by a n -vector $\mathbf{u} = [u_1, u_2, \dots, u_n]^T$ and $\mathbf{d} = [d_1, d_2, \dots, d_n]^T$, with $n \in \mathbb{N}$, $n \geq 1$, $u_1 > u_2 > \dots > u_n > r > 0$, and $d_n < d_{n-1} < \dots < d_1 < 0$. As a result, one has to expect $2n + 1$ different values for S_1 . That is:

$$S_1 = \begin{cases} S_0 e^{u_1} & \text{with probability } p_{u_1} < 1 \\ \vdots \\ S_0 e^{u_n} & \text{with probability } p_{u_n} < 1 \\ S_0 & \text{with probability } p_0 < 1, \\ S_0 e^{d_1} & \text{with probability } p_{d_1} < 1 \\ \vdots \\ S_0 e^{d_n} & \text{with probability } p_{d_n} < 1 \end{cases}$$

with

$$p_0 = 1 - p_{u_1} - \dots - p_{u_n} - p_{d_1} - \dots - p_{d_n} > 0.$$

Based on these different $2n + 1$ possibilities for S_1 , we have now the possibility to construct an over-determined linear system of $2n + 1$ equations with only 2 unknowns. That is:

$$C_1 = \begin{cases} \Delta_S (S_0 e^{u_1}) + \Delta_D D_1 = y_1 \\ \Delta_S (S_0 e^{u_2}) + \Delta_D D_1 = y_2 \\ \vdots \\ \Delta_S (S_0 e^{u_n}) + \Delta_D D_1 = y_n \\ \boxed{\Delta_S S_0 + \Delta_D D_1 = y_{n+1}}, \\ \Delta_S (S_0 e^{d_1}) + \Delta_D D_1 = y_{n+2} \\ \Delta_S (S_0 e^{d_2}) + \Delta_D D_1 = y_{n+3} \\ \vdots \\ \Delta_S (S_0 e^{d_n}) + \Delta_D D_1 = y_{2n+1} \end{cases}$$

where $y_1 = (S_0 e^{u_1} - K)_+$, $y_2 = (S_0 e^{u_2} - K)_+$, $y_n = (S_0 e^{u_n} - K)_+$, $y_{n+1} = (S_0 - K)_+$, $y_{n+2} = (S_0 e^{d_1} - K)_+$, $y_{n+3} = (S_0 e^{d_2} - K)_+$, and $y_{2n+1} = (S_0 e^{d_n} - K)_+$. We can observe that when $n = 1$, we return back to the standard

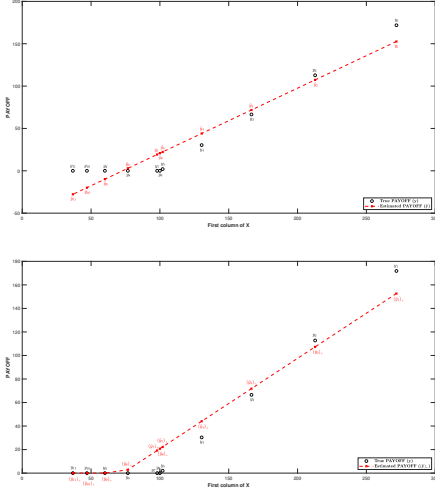


Fig. 2. From top to bottom: the predicted PAYOFF vector via the least squares $\hat{\mathbf{y}} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{11}]^T \in \mathbb{R}_+^{11}$, its rectified linear unit version vector $(\hat{\mathbf{y}})_+ = [(\hat{y}_1)_+, (\hat{y}_2)_+, \dots, (\hat{y}_{11})_+]^T \in \mathbb{R}_+^{11}$.

(one-period) trinomial option pricing model already described in subsection II-A.

What happens to the least squares estimation of Δ with $2n+1$, $n \geq 1$, observations?

Let fix $r = 2\%$, $S_0 = \$100$, $K = S_0$, and $D_0 = \$1$. Instead of specifying a single specific value for u as we did in the example in subsection II-C, we predict the PAYOFF with the least squares under a set of 5 possible evenly spaced values for $u > 0$ at once, that is, a vector $\mathbf{u} = [u_1, u_2, \dots, u_5]^T$, $u_1 > u_2 > \dots > u_5 > r$. We choose $u_5 = r + \epsilon$, $\forall \epsilon > 0$, and $u_1 = 1$. We also fix $\mathbf{d} = [-u_5, -u_4, \dots, -u_1]^T$. Hence, we have 11 true (known) possible PAYOFF values, that is, $\mathbf{y} = [y_1, y_2, \dots, y_{11}]^T \in \mathbb{R}_+^{11}$. Figure 2 exhibits the difference between the true PAYOFF values and the predicted ones via the least squares. We can clearly observe that the observations y_7, y_8, y_9, y_{10} , and y_{11} deviate from the pattern of the data formed by the remaining 6 observations (that is, y_1, y_2, y_3, y_4, y_5 , and y_6), and thus, can be identified as outliers. Therefore, it becomes obvious that the least squares regression line will be highly skewed towards these identified outliers as can be seen from the red dashed-line in figure 2.

B. Estimation of Δ via robust regression M -estimators

Recall that the least squares is derived under the Gaussian assumption which does not take into consideration the outliers. In order to not consider any specific kind of distribution (ex. the Gaussian assumption), we are going to assume that the y_i 's, $i \in [1, 2n+1]$, $n \geq 1$, are independent and not identically distributed with any density of the form $\frac{1}{\sigma} f_0 \left(\frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right)$. The

likelihood under this density function can be written as [17], [18]:

$$L(y_1, \dots, y_{2n+1}; \Delta, \sigma) = \prod_{i=1}^{2n+1} \frac{1}{\sigma} f_0 \left(\frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right).$$

Maximizing the log-likelihood w.r.t. (Δ, σ) is thus equivalent to the following (minimization) optimization problem:

$$\begin{aligned} \max_{\Delta, \sigma} & \left\{ -(2n+1) \log(\sigma) + \sum_{i=1}^{2n+1} \log \left(f_0 \left(\frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right) \right) \right\} \\ \equiv \min_{\Delta, \sigma} & \left\{ (2n+1) \log(\sigma) + \sum_{i=1}^{2n+1} \rho \left(\frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right) \right\}, \end{aligned} \quad (5)$$

where $-\log \left(f_0 \left(\frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right) \right) = \rho \left(\frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right)$ for any non-constant function $\rho(\cdot)$. One can clearly observe that the least squares is a special case when $\rho \left(\frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right) = \left(\frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right)^2$.

By taking the derivative of the objective function in (5) with respect to Δ and σ and set it to zero, we obtain [17]:

$$\sum_{i=1}^{2n+1} \rho' \left(\frac{y_i - \mathbf{x}_i^T \hat{\Delta}}{\hat{\sigma}} \right) \mathbf{x}_i = \mathbf{0},$$

and

$$\frac{1}{(2n+1)} \sum_{i=1}^{2n+1} \left(\frac{y_i - \mathbf{x}_i^T \hat{\Delta}}{\hat{\sigma}} \right) \rho' \left(\frac{y_i - \mathbf{x}_i^T \hat{\Delta}}{\hat{\sigma}} \right) = 1,$$

respectively.

Note that the function $\rho(\cdot)$ must be chosen to respect the following conditions [17]:

- $\rho(x)$ should be a non-decreasing function of $|x|$;
- $\rho(0) = 0$;
- $\rho(x)$ is increasing for $x > 0$ such that $\rho(x) < \rho(\infty)$;
- if ρ is bounded, it is also assumed that $\rho(\infty) = 1$.

In this paper, we mainly focus on five types of $\rho(\cdot)$ -functions (see table II) in order to estimate the delta-hedging coefficients, and thus, the prediction of the PAYOFF as well as the true (unknown) call option price at initial date C_0 .

- 1) The *sample mean* estimator: recall that the least squares estimates, already described in subsection II-B, is the regression equivalent to the sample mean which is well known to be very sensitive to outliers.
- 2) The *sample median* estimator: To deal with the outliers in the data, the first intuition that comes into our mind is to work with the sample median [17], [18] which is much more robust to outliers than the sample mean. The estimator $\hat{\Delta}$ that minimizes problem (5) with this $\rho(\cdot)$ function is known as the L_1 -estimator of Δ . As $2n+1$ is always odd, the derivative of the objective function in (5) with respect to $\mathbf{x}_i^T \Delta$ exists and which is everywhere except for the observations y_1, \dots, y_{2n+1} , and it is equal to $-\sum_{i=1}^{2n+1} \text{sign}(y_i - \mathbf{x}_i^T \Delta)$ which can never be zero. Thus, the minimizer must occur at one of the points y_1, \dots, y_{2n+1} where the objective function in (5) is indeed not differentiable. However, it is a continuous function and is decreasing when $\mathbf{x}_i^T \Delta < \frac{y_{2n+2}}{2}$ and

| | |
|-----------------|--|
| Estimators | $\rho \left(\frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right)$ |
| Sample mean | $\left(\frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right)^2$ |
| Sample median | $\frac{y_i - \mathbf{x}_i^T \Delta}{\sigma}$ |
| Winsorized mean | $\begin{cases} \left(\frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right)^2 & \text{if } \left \frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right \leq \kappa \\ 2\kappa \left \frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right - \kappa^2 & \text{if } \left \frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right > \kappa \end{cases}$ |
| Bisquare | $\begin{cases} 1 - \left(1 - \left(\frac{y_i - \mathbf{x}_i^T \Delta}{\sigma \kappa} \right)^2 \right)^3 & \text{if } \left \frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right \leq \kappa \\ 1 & \text{if } \left \frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right > \kappa \end{cases}$ |
| Trimmed mean | $\begin{cases} \frac{1}{2} \left(\frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right)^2 & \text{if } \left \frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right \leq \kappa \\ \frac{1}{2} \kappa^2 & \text{if } \left \frac{y_i - \mathbf{x}_i^T \Delta}{\sigma} \right > \kappa \end{cases}$ |

TABLE II
THE DIFFERENT $\rho(\cdot)$ -FUNCTIONS.

increasing when $\mathbf{x}_i^T \Delta > y_{2n+2}$. Hence, the minimizer is given by $\mathbf{x}_i^T \hat{\Delta} = y_{2n+2}$. However, unlike the least squares, there is no explicit expression for the minimizer of (5) since the l_1 -norm is not differentiable at zero. However, one can find recursive algorithms that are able to solve it. For example, in our experiments later, we use a Matlab software for disciplined convex programming (e.g., CVX Matlab) to estimate the delta-hedging coefficients.

- 3) The *Winsorized mean* estimator: For the first impression, one can directly think to propose an estimator that combines the robustness of the sample median (under the non-Gaussian assumption) and the low-variance of the sample mean (under exact normality). Hence, for a fair compromise between the mean and median, the function $\rho(\cdot)$ can be chosen to belong to the family of Huber functions. Figure 3 exhibits both the $\rho(\cdot)$ - and $\rho'(\cdot)$ -functions under this family of Huber functions. Note that the solution $\mathbf{x}_i^T \hat{\Delta}$ is closely related to Winsorizing [18].

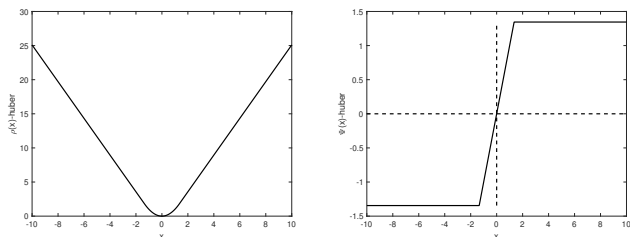


Fig. 3. From left to right: the huber ρ - and ρ' - functions (with $\kappa = 1.345$).

- 4) The *bisquare* (re-descending) estimator: From figure 3, we can observe that the function $\rho(\cdot)$ aims to tend to infinity at infinity, whereas its derivative, that is, the function $\rho'(\cdot)$, is continuous but always shifted by $\pm\kappa = 1.345$. These findings demonstrate that both the

gross as well as the moderately large outliers are going to be treated in the same manner. In this regard, thinking about using a function $\rho'(\cdot)$ that re-descend smoothly to zero at infinity (see figure 4) may be quite helpful since in this case the gross outliers will be completely rejected and the moderately large outliers will be completely ignored.

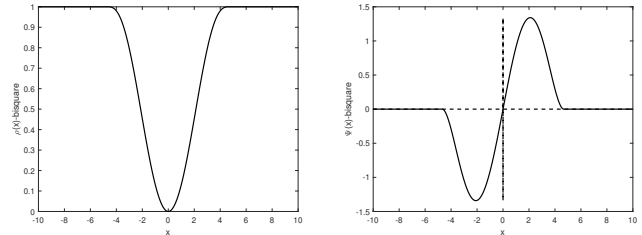


Fig. 4. From left to right: the bisquare ρ - and ρ' - functions (with $\kappa = 4.685$).

- 5) The *trimmed mean* estimator: In addition to the sample median and winsorized mean, one can think of discarding a proportion of the largest and smallest values. This is well-known as the trimming procedure [18] and which can be much more robust to outliers than the sample mean and sample median.

C. Strategy2: Estimation of Δ via least squares after the manual removal of the outliers

As we have seen from the example in subsection III-A, it is possible to identify the nature of the observations (outliers) that deviate from the bulk of the data. Hence, it would be very beneficial to first manually remove these outliers from the data, and then apply the least squares to estimate the delta-hedging coefficient vector Δ . It is obvious to mention that in the case when $K = S_0$, and for any $n \geq 1$, the PAYOFF values y_1, \dots, y_{n+1} will always belong on the regression line, and so they can form the true bulk of the data. Whereas the remaining PAYOFF values, that is, y_{n+2}, \dots, y_{2n+1} are equal to zero, and so they can be considered as the observations that deviate from the general pattern of the data.

As we are now able to identify the nature of the outliers, it becomes easy to manually remove them from the data and then apply the least squares only on the remaining observations (that is, after the outliers have been removed).

IV. EXPERIMENTS

In this section, we mainly aim to evaluate both the prediction strategies of option pricing already described in subsection III-B and subsection III-C. The evaluations are done on two different scenarios for n :

- $n = 1$ (the standard one-period trinomial option pricing (see subsection II-A)): a single value for each of $u > 0$ and $d < 0$ is specified to price the call option. The value for d is always fixed to be equal to $-u$;
- $n > 1$ (our proposed extension of the one-period trinomial option pricing model (see subsection III-A)): a vector of $\mathbf{u} = [u_1, \dots, u_n]^T$ and $\mathbf{d} = [d_1, \dots, d_n]^T$

are specified to price the option. The vector \mathbf{d} is always fixed to be equal to $[-u_n, -u_{n-1}, \dots, -u_1]^T$.

In all the experiments, we fix $r = 2\%$, $D_0 = \$1$, and we assume that the asset price at initial date is $S_0 = \$100$. A call on the asset is available with a strike price $K = S_0$, expiring at maturity. To keep it simple, we assume that the underlying asset pays no dividend during the life of the call, neither counterparty to the transaction is at risk of default, and all the transaction costs, margin requirements and taxes are completely ignored. All our experiments³ are conducted on MATLAB 2021a.

Note that we use the *robustfit* package in MATLAB to simultaneously estimate both Δ and σ for the sample median, winsorized mean, bisquare, and trimmed mean estimators. This package uses an iteratively reweighted least squares to compute the delta-hedging coefficients.

A. Tuning selection of κ

To choose the best parameter value for κ for every robust regression technique (e.g. winsorizing, bisquare, and trimmed mean), we decide to vary κ between 1 and 5 by step of 10^{-4} . The parameter value κ that minimizes the criterion $\frac{\|\mathbf{y} - (\hat{\mathbf{y}})_+\|_2}{\|\mathbf{y}\|_2}$ will be selected as the optimal value.

B. European option pricing

Let fix $\epsilon = 10^{-6}$.

- When $n = 1$: we choose to evaluate both the option pricing prediction strategies in subsection III-B and III-C on a set of 100 single evenly spaced values for u between $r + \epsilon$ and 5. Figure 5 exhibits, as a function of u , the error results in the prediction of \mathbf{y} , the prediction results of C_0 , as well as the different obtained optimal values of κ for the bisquare, winsorized mean and trimmed mean. As we can observe from figure 5, the robust regression M -estimators behave poorly for all the different values of u . However, Strategy2 (that is, when applying the least squares after the manual removal of the outliers) achieves the lowest prediction error for all the values of u and presents a stable \hat{C}_0 always varying around \$1.98013.
- When $n > 1$: for any $u_n \geq r + \epsilon$, both the option pricing prediction strategies in subsection III-B and III-C are evaluated on an interval $\mathbf{u} = [u_1, \dots, u_n]^T$ containing $n > 1$ evenly spaced values. Figure 6 exhibits, as a function of $n \in \{2, \dots, 100\}$ in steps of 1, and for $u_1 = 1$, the error results in the prediction of \mathbf{y} , the prediction results of C_0 , as well as the different obtained optimal values of κ for the bisquare, winsorized mean and trimmed mean. From the top row of figure 6, it is important to note that the bisquare behaves well (that is, a low error in the prediction of \mathbf{y}) only when n is between 2 and 39 where the prediction error is negligible (about 1.33×10^{-15}). Whereas the trimmed mean behaves well when n is between 8 and 70. In these two specific ranges

³The MATLAB code of the proposed work and all the experiments is available upon request.

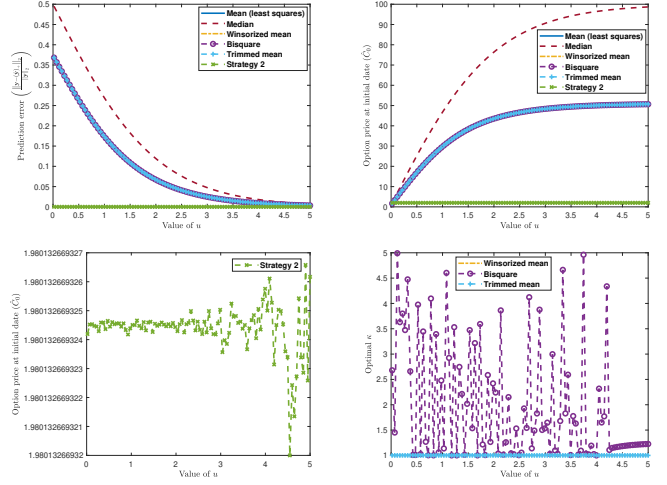


Fig. 5. Evaluations when $n = 1$. From top left to bottom right: the prediction error $\frac{\|\mathbf{y} - (\hat{\mathbf{y}})_+\|_2}{\|\mathbf{y}\|_2}$, the predicted European call option pricing at initial date (\hat{C}_0), the predicted option price \hat{C}_0 for Strategy2 only, the optimal values of κ for {winsorized mean, bisquare, and trimmed mean}.

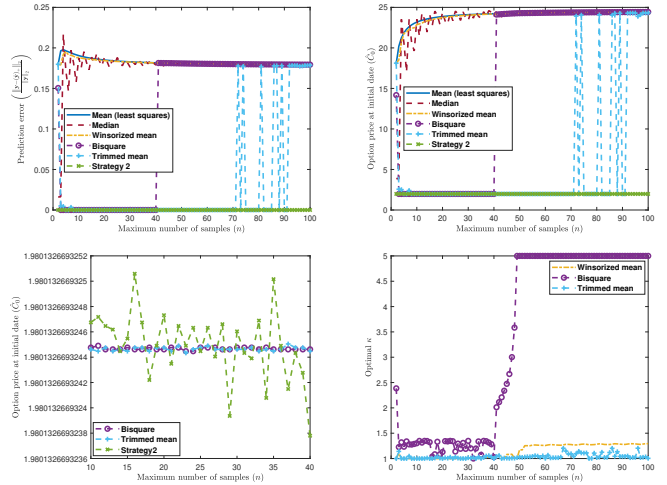


Fig. 6. Evaluations when $n > 1$. From top left to bottom right: the prediction error $\frac{\|\mathbf{y} - (\hat{\mathbf{y}})_+\|_2}{\|\mathbf{y}\|_2}$, the predicted European call option pricing at initial date (\hat{C}_0), the predicted option price \hat{C}_0 for {bisquare, trimmed mean, and strategy2}, the optimal values of κ for {winsorized mean, bisquare, and trimmed mean}.

of n , the predicted option price \hat{C}_0 is about \$1.98013. It is true that Strategy2 always perform the best for all the possibilities for n , it can be seen, however, that it presents higher variance in the values of \hat{C}_0 compared to the bisquare and trimmed mean (see the bottom left plot of figure 6). Figure 7 exhibits the difference between \mathbf{y} and $(\hat{\mathbf{y}})_+$ for the bisquare (when $n = 39$), trimmed mean (when $n = 70$), and Strategy2 (when $n = 100$). As we can observe, the PAYOFF vector \mathbf{y} can be predicted very efficiently.

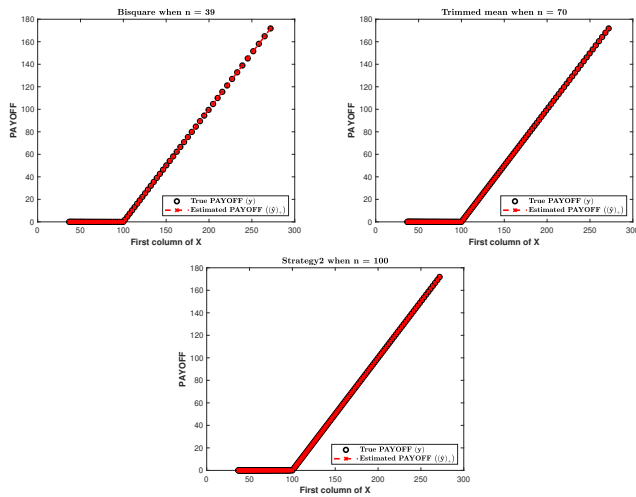


Fig. 7. Difference between the true known PAYOFF y and the predicted one $(\hat{y})_+$.

V. DISCUSSION AND FUTURE WORK

A. Conclusion

The trinomial European call option pricing model is well known to be incomplete, and thus, the least squares linear regression technique can simply be used to build a self-financing hedging portfolio that does not perfectly replicates the call option. However, the least squares is quite sensitive to even a single outlier in the data, and thus, the prediction of the PAYOFF and the option price at initial date can well deviate from their true values. This paper has briefly outlined two different option pricing prediction strategies that alleviate the problem of outliers. The first strategy exploits some robust linear regression M -estimators such as the median, winsorized mean, bisquare, and trimmed mean. Whereas the second strategy simply applies the least squares but after the removal of the outliers. Both the strategies have been evaluated on numerical experiments, and the results of which demonstrate their effectiveness for European call option pricing.

B. Some Directions for Future Work

For future enhancements, a likely first step would be to extend the proposed work to the multi-period case. Other promising avenues for further research include the improvement of the financial risk management by buying European options on assets as well as on exchange rates at the same time. More precisely, when the option is in the money, that is, $S_1 > K$, it would be very beneficial for any buyer to exercise the option especially if K is much lower than S_1 . However, one of the several reasons that can prevent the buyer to exercise may be the potential fall of the domestic currency value, mainly caused by the huge increase of the exchange rate with respect to the foreign currency. In this regard, it would be quite important for any buyer to buy an option on the underlying asset and also another option on the exchange rate at the same time in order to prevent any huge fluctuations

of the exchange rate that can not be for the benefit of the buyer.

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] M. Scholes. A conversation with myron scholes. Youtube.
- [11] 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>
- [12] 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>
- [13] "Option pricing: A simplified approach," *Journal of Financial Economics*, vol. 7, no. 3, pp. 229–263, 1979.
- [14] P. P. Boyle, "Option valuation using a three jump process," 1986. [Online]. Available: <https://api.semanticscholar.org/CorpusID:150764610>
- [15] H. Takahashi, "A note on pricing derivatives in an incomplete market," *Graduate School of Economics, Hitotsubashi University, Discussion Papers*, 01 2000.
- [16] B. Johan, "The trinomial asset pricing model," *GInstitutionen for matematiska vetenskaper, Goteborgs universitet, Thesis report*, 01 2016.
- [17] 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>
- [18] 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>