

Nonparametric Predictive Inference for European Option Pricing based on the Binomial Tree Model

Ting He^{*1}, Frank P.A. Coolen^{†1}, and Tahani Coolen-Maturi^{†2}

¹Department of Mathematical Sciences, Durham University, Durham DH1 3LE

²Durham University Business School, Durham University, Durham DH1 3LB

e-mail: ting.he@durham.ac.uk; frank.coolen@durham.ac.uk;
tahani.maturi@durham.ac.uk

ABSTRACT

In finance, option pricing is one of the main topics. A basic model for option pricing is the Binomial Tree Model, proposed by Cox, Ross, and Rubinstein in 1979 (CRR). This model assumes that the underlying asset price follows a binomial distribution with a constant upward probability, the so-called risk-neutral probability. In this paper, we propose a novel method based on the binomial tree. Rather than using the risk-neutral probability, we apply Nonparametric Predictive Inference (NPI) to infer imprecise probabilities of movements, reflecting more uncertainty while learning from data. To study its performance, we price the same European options utilizing both the NPI method and the CRR model and compare the results in two different scenarios, firstly where the CRR assumptions are right, and secondly where the CRR model assumptions deviate from the real market. It turns out that our NPI method, as expected, cannot perform better than the CRR in the first scenario, but can do better in the second scenario.

Key words: CRR Binomial Tree Model, European Option, Imprecise Probability, Nonparametric Predictive Inference, Option Pricing

1 Introduction

In the finance literature, one of the classic option pricing models is the Binomial Tree Option Pricing Model. The Binomial Tree Model, proposed by Cox, Ross, and Rubinstein in 1979, referred to as the CRR model, is a discrete-time model which has been proven to converge to the Black-Scholes formula when time increments approach to zero (Cox et al., 1979). This model uses the option information, a stock with no dividends having initial stock price S_0 which will either go up by the factor u or go down by the factor d , strike price K_c for the call option and K_p for the put option and m future time steps. If also assume in a frictionless and complete market without any other costs in which the underlying asset price follows a binomial distribution with a constant probability q , which is risk-neutral and normally different from the real market probability p . The pricing formula for a call option is,

$$E_c^{CRR}[S_m - K_c]^+ = \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{m}{k} q^k * (1 - q)^{(m-k)} \quad (1)$$

where k_c^* is such that $u^{k_c^*} d^{m-k_c^*} S_0 - K_c = 0$, and $\lceil k_c^* \rceil$ denotes the smallest integer greater than or equal to k_c^* . For a put option,

$$E_p^{CRR}[K_p - S_m]^+ = \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{m}{k} q^k * (1 - q)^{(m-k)} \quad (2)$$

k_p^* is such that $K_p - u^{k_p^*} d^{m-k_p^*} S_0 = 0$ and where $\lfloor k_p^* \rfloor$ denotes the largest integer less than or equal to k_p^* .

However, this model is based on some strong assumptions making it far from realistic, including that the upward movement probability of the underlying asset is risk-free and constant, and the market is complete without any arbitrage opportunity. All these assumptions are unlikely to be satisfied in the real world. There are many papers challenging those unrealistic assumptions and presenting new option pricing models. For example, Jackwerth and Rubinstein (1996) used a nonparametric method to deduce risk-neutral probabilities from option prices. GMPOP, short for generalized multi-period option pricing model, is a binomial tree model with subjective probability in the real world to price options, but this subjective probability is still constant (Arnold and Crack, 2003).

Imprecise probability generalizes probability theory, for the circumstances that information is too limited to conclude a precise probability for an event of interest. So imprecise probability reflects more uncertainty about the event. Imprecise probabilities have been introduced to describe financial markets and to solve finance problems. Berleant et al. (2008) provide criteria and a measure for portfolio selection problems by utilizing imprecise probabilities. Imprecise probabilities also help with decision making in case of imprecise risk (Jaffray and Jeleva, 2008). Muzzioli and Reynaerts (2008) proposed a model to price American options with imprecise probabilities. The Nonparametric Predictive Inference (NPI) method, a data-based imprecise probability method has been developed for a range of problems in OR, including queueing (Coolen-Schrijner and Coolen, 2007), replacement problems (Coolen-Schrijner et al., 2009), and many applications in reliability (Coolen et al., 2014) and statistics (Coolen, 1998); (Coolen et al., 2013). NPI has been applied to finance prediction, providing a relatively straightforward way to study future stock return when little further information is available (Baker et al., 2017). Because of the attractive properties of the NPI method (Coolen, 2011), its implementation in option pricing is

appealing. Unlike the CRR model, where the probability of stock movement is constant and precise, the probabilities from the NPI method are in the form of an interval with lower and upper bounds, gained through studying the observed data within a frequentist statistics framework, which makes it an appealing forecasting method (Coolen, 2011). Another advanced property of NPI is that it keeps learning from data. When predicting multiple future observations, NPI considers all the predicted observations as observed data and uses the new imprecise probabilities learnt from the predicted data and historical data to forecast the next future observation (Coolen, 1998). Thanks to utilizing imprecise probabilities from NPI method, outcomes of the predictions exhibit more variation than those of the CRR model. To study core pricing procedure based on the NPI method, we neglect the discounted factor at this stage focusing only on the basic model.

Section 2 provides a brief introduction to the NPI method. We propose NPI for the European option pricing model in Section 3, and in Section 4 our NPI method is compared to the CRR model in order to study its performance. Some conclusions and extensions are discussed in Section 5.

2 Nonparametric Predictive Inference

Nonparametric Predictive Inference (NPI) is an inferential framework based on the assumption $A_{(n)}$ presented by Hill (Hill, 1968), which directly provides probabilities for future observations by using few model assumptions and observed values of related random quantities. Suppose that there exists a sequence of real-valued, continuous and exchangeable random quantities, X_1, \dots, X_n, X_{n+1} , without any tie between any two of them. Assume that X_1, \dots, X_n be ordered and their realizations denoted as $x_{(1)} < \dots < x_{(n)}$ and let $x_{(0)} = -\infty$ and $x_{(n+1)} = \infty$ for ease of notation. These ordered observed data partition the real line into $n + 1$ open intervals $I_j = (x_{(j-1)}, x_{(j)})$, where $j = 1, 2, \dots, n + 1$. For the future observation X_{n+1} on the basis of n observed values, the assumption $A_{(n)}$ (Hill, 1968) is

$$P(X_{n+1} \in I_j) = \frac{1}{n+1} \text{ for } j = 1, 2, \dots, n + 1$$

So the probability for the event that the next observation falls in the interval $I_j = (x_{(j-1)}, x_{(j)})$ is $\frac{1}{n+1}$, for each interval I_j . $A_{(n)}$ does not assume any knowledge of the distribution of random quantities of interest. By introducing imprecise probability theory, $A_{(n)}$ provides optimal bounds for the probability of any event of interest involving X_{n+1} , namely lower and upper probabilities in imprecise probability theory (Walley, 1991) and interval probability theory (Weichselberger, 2001), following from De Finetti's fundamental theorem of probability (De Finetti, 1974). NPI is a frequentist statistical method which has strong consistency properties (Augustin and Coolen, 2004).

The NPI method has been developed for Bernoulli data (Coolen, 1998), each with a 'success' or 'failure' result. Y_1^n represents the number of successful trials in n observed trials. Let Y_{n+1}^{n+m} represent the number of successful observations in m future trials. Here the n and m observations are exchangeable. The upper probability for the event $Y_{n+1}^{n+m} \leq k^*$, given data $Y_1^n = s$, for $s \in \{0, \dots, n\}$, is (Coolen, 1998)

$$\overline{P}(Y_{n+1}^{n+m} \leq k^* | Y_1^n = s) = \binom{n+m}{m}^{-1} \sum_{k=0}^{k^*} \binom{s+k-1}{k} \binom{n-s+m-k}{m-k} \quad (3)$$

The NPI upper probability for the event $Y_{n+1}^{n+m} \geq k^* + 1$, given $Y_1^n = s$, for $s \in \{0, \dots, n\}$, is

$$\bar{P}(Y_{n+1}^{n+m} \geq k^* + 1 | Y_1^n = s) = \binom{n+m}{m}^{-1} \sum_{k=k^*+1}^m \binom{s+k}{k} \binom{n-s+m-k-1}{m-k} \quad (4)$$

We can deduce the corresponding NPI lower probabilities by the conjugacy property $\bar{P}(A) = 1 - \underline{P}(A^c)$, where A^c is the complementary event to A ,

$$\underline{P}(Y_{n+1}^{n+m} \leq k^* | Y_1^n = s) = 1 - \bar{P}(Y_{n+1}^{n+m} \geq k^* + 1 | Y_1^n = s) \quad (5)$$

$$\underline{P}(Y_{n+1}^{n+m} \geq k^* + 1 | Y_1^n = s) = 1 - \bar{P}(Y_{n+1}^{n+m} \leq k^* | Y_1^n = s) \quad (6)$$

Similarly, NPI can infer the lower and upper expectations of a function $g(Y(m))$ given observed data, here to simplify the formulas we use $Y(m)$ to represent the number of successes in m future trails instead of Y_{n+1}^{n+m} . The NPI method for Bernoulli data (Coolen, 1998) provides a set \mathcal{P} of classical, precise, probability distributions for which the presented lower and upper probabilities are optimal bounds. In imprecise probability theory, this set \mathcal{P} is called a structure. The lower and upper expected values for a real-valued function g of $Y(m)$ can be derived by

$$\underline{E}(g(Y(m))) = \inf_{p \in \mathcal{P}} E^p(g(Y(m))) \quad (7)$$

$$\bar{E}(g(Y(m))) = \sup_{p \in \mathcal{P}} E^p(g(Y(m))) \quad (8)$$

where E^p is the expected value corresponding to a specific precise probability distribution $p \in \mathcal{P}$. Then for these purposes, we need to use the probability functions that can give us the boundaries of the expected values rather than the probability bounds. However, for European options the probability functions giving optimal expected values and lower and upper probabilities are same, and this will be explained in the next section. Since stock price movements for each time step in the simple Binomial tree model can be represented as Bernoulli data, NPI for Bernoulli data is suitable to infer imprecise probabilities and expected payoffs for call and put options, as presented in the following section.

3 NPI for European Option Pricing

We introduce the assumptions of our NPI European option pricing method. As shown in Figure 1, the underlying asset has two possible outcomes at the next time step, either going up to uS_0 or going down to dS_0 , with $u > 1$, $d < 1$ and S_0 the initial stock price at the start time without paying any dividends during the period considered. We assume that the n historical data are sufficient to analyze option prices, and among n observed data there are s successful observations, i.e. s times the stock price went up and $n - s$ times it went down. To simplify our model, we assume there is no effect of discounted factor, assuming the time of trading is close to the maturity so the influence of any discount factor could be neglected.

Referring to Figure 1 and the option definition, for each type of option only paths with positive end values are taken into account, because an option is a right for the buyer and the buyer would like to exercise the option if the payoff is positive. For call options, only paths which have payoff $S_m - K_c$

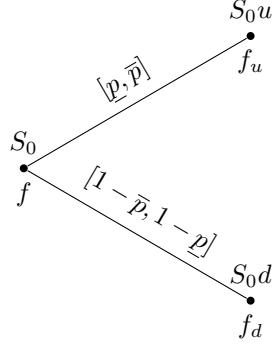


Figure 1: Stock and option prices one-step tree

greater than zero, noted as $[S_m - K_c]^+$, are taken into account, where S_m is the stock price at maturity and K_c is the strike price, then

$$S_m - K_c = u^{Y(m)} d^{m-Y(m)} S_0 - K_c > 0 \quad (9)$$

$$Y(m) > \frac{\ln K_c - \ln S_0 - m \ln d}{\ln u - \ln d} =: k_c^* \quad (10)$$

The NPI lower and upper probabilities for a call option, all stock prices at the m step higher than the strike price of this call option, are calculated according to the NPI method for Bernoulli data.

$$\underline{P}(Y(m) \geq \lceil k_c^* \rceil) = \binom{n+m}{m}^{-1} \sum_{k=\lceil k_c^* \rceil}^m \binom{s+k-1}{k} \binom{n-s+m-k}{m-k} \quad (11)$$

$$\overline{P}(Y(m) \geq \lceil k_c^* \rceil) = \binom{n+m}{m}^{-1} \sum_{k=\lceil k_c^* \rceil}^m \binom{s+k}{k} \binom{n-s+m-k-1}{m-k} \quad (12)$$

For put options, we consider paths with payoffs $K_p - S_m > 0$, where K_p is the strike price. On the basis of this definition, the payoff of a put option is $[K_p - S_m]^+$, then

$$K_p - S_m = K_p - u^{Y(m)} d^{m-Y(m)} S_0 > 0 \quad (13)$$

$$Y(m) < \frac{\ln K_p - \ln S_0 - m \ln d}{\ln u - \ln d} =: k_p^* \quad (14)$$

Following the same steps as we did for call options, we find paths considered for put option, and the interested event should be $Y(m) \leq \lfloor k_p^* \rfloor$.

$$\underline{P}(Y(m) \leq \lfloor k_p^* \rfloor) = \binom{n+m}{m}^{-1} \sum_{k=0}^{\lfloor k_p^* \rfloor} \binom{s+k}{k} \binom{n-s+m-k-1}{m-k} \quad (15)$$

$$\overline{P}(Y(m) \leq \lfloor k_p^* \rfloor) = \binom{n+m}{m}^{-1} \sum_{k=0}^{\lfloor k_p^* \rfloor} \binom{s+k-1}{k} \binom{n-s+m-k}{m-k} \quad (16)$$

Actually, rather than lower and upper probabilities, we are more interested in lower and upper

expected values, and they are given by Equations (7) and (8), where the real-valued function $g(Y(m))$ is equal to $[S_m - K_c]^+$ for a call option or $[K_p - S_m]^+$ for a put option, because S_m is a random variable depending on m . According to the trading actions, expected boundary payoffs are renamed, $\underline{E}(g(Y(m)))$ denotes the maximum payoff an investor would buy and $\overline{E}(g(Y(m)))$ denotes the minimum payoff an investor would be willing to sell for. As we have already computed the lower and upper probabilities, for call options (Equations (11) and (12)), as well as for put options (Equations (15) and (16)), and the real-valued function g , then formulas for European option expected payoffs can be generated:

Minimum selling payoff of call option

$$\begin{aligned}\overline{E}_c[S_m - K_c]^+ &= \binom{n+m}{m}^{-1} \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] (\overline{P}(Y(m) \geq k) - \overline{P}(Y(m) \geq k+1)) \\ &= \binom{n+m}{m}^{-1} \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{s+k}{k} \binom{n-s+m-k-1}{m-k}\end{aligned}\quad (17)$$

Here for each term with k from $\lceil k_c^* \rceil$ to m , we assign a probability $\overline{P}(Y(m) \geq k) - \overline{P}(Y(m) \geq k+1)$. This ensures that we give the maximum possible probability to the largest possible value for k , then the maximum possible remaining probability to the second largest value for k , and so on.

Minimum selling payoff of put option

$$\begin{aligned}\overline{E}_p[K_p - S_m]^+ &= \binom{n+m}{m}^{-1} \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] (\overline{P}(Y(m) \leq k) - \overline{P}(Y(m) \leq k-1)) \\ &= \binom{n+m}{m}^{-1} \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{s+k-1}{k} \binom{n-s+m-k}{m-k}\end{aligned}\quad (18)$$

For downward paths with k from 0 to $\lfloor k_p^* \rfloor$, each path we assign a probability $\overline{P}(Y(m) \leq k) - \overline{P}(Y(m) \leq k-1)$, which ensures that we give the maximum possible probability to the lowest possible value for k , then then maximum possible remaining probability to the second lowest value for k , and so on.

Using similar derivations, we can formulate the lower expected value for a call option or a put option.

Maximum buying payoff of call option

$$\underline{E}_c[S - K_c]^+ = \binom{n+m}{m}^{-1} \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{s+k-1}{k} \binom{n-s+m-k}{m-k}\quad (19)$$

Maximum buying payoff of put option

$$\underline{E}_p[K_p - S_m]^+ = \binom{n+m}{m}^{-1} \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{s+k}{k} \binom{n-s+m-k-1}{m-k}\quad (20)$$

Therefore, for each type of option there is an interval of expected payoffs with bounds as the maximum

buying payoff and the minimum selling payoff. As we calculated, for call and put option we get an interval of the expected values, which means with limited information any value in this interval is reasonable to the NPI investor, and any value outside this interval is appealing to a NPI investor. When the NPI investor is offered a payoff higher than the minimum selling payoff, it is overvalued according to NPI outcomes. Similarly, the NPI investor would see any value less than the maximum buying payoff as undervalued, while the value in between expected value bounds does not trigger any trading action. The expected NPI boundary payoffs also follow a similar put-call relationship as the traditional put call parity, and formulas are $\overline{E}_c[S_m - K]^+ - \underline{E}_p[K - S_m]^+ = \overline{E}(S_m) - K$ and $\overline{E}_p[K - S_m]^+ - \underline{E}_c[S_m - K]^+ = K - \underline{E}(S_m)$, where $\overline{E}(S_m)$ and $\underline{E}(S_m)$ are the lower and upper expectations of S_m from the NPI method. In terms of the portfolio theory, this relationship has the same meaning as the classic one, which a portfolio containing opposite trading position for a call and a put option, longing a put and shorting a call or longing a call and shorting a put, has the same value as a portfolio consisted of a zero-coupon bond and a stock, longing a bond and shorting a stock or longing a stock and shorting a bond.

4 Performance Study

In this section we study the performance of the NPI method for European options in comparison to the CRR model. If there are only two investors in the option market, the CRR person and the NPI person, we would like to see the expected profit or loss of the NPI person when he trades with the CRR person. We consider two scenarios, first the CRR model perfectly captures the future market trend, meaning the stock price at the maturity will equal to the expected stock price from the CRR model. In the second scenarios, the CRR model is wrong about the future trend, which the real market probability is different from the risk-neutral probability. Whereas in both these scenarios the NPI method predicts based on the historical data.

In these two scenarios, we would like to compare payoffs from the NPI method and CRR model based on the same option with the same underlying stock. The key factor in the comparison is s when the maturity of the option m is fixed, the other factors of the binomial tree and the CRR model are fixed, including the number of historical data n . Different values of s will lead to different NPI payoffs, so compared to the CRR payoff this will result in different trading actions. The CRR payoff formulated by Equation (1) for a call option or Equation (2) for a put option is a constant value, while NPI payoffs vary with s . There are three trading cases according to s . When the CRR payoff is in between the maximum buying and the minimum selling NPI payoffs, there is no trading between the two investors. Otherwise, the NPI person will either sell an option or buy it depending on whether the CRR payoff is lower than the maximum buying NPI payoff or higher than the minimum selling NPI payoff. We ignore time discounting at this stage, so we can regard the expected payoffs as the expected price, and we denote lower and upper expected prices as \overline{Pr} and \underline{Pr} , as well as expected CRR price Pr^{CRR} , equal to expected option payoffs \overline{E} , \underline{E} and E^{CRR} , respectively. As we would like to know the worst result that the NPI person will have eventually, we assume the NPI person will quote the maximum buying price or the minimum selling price and any trade occurs at these prices.

4.1 Scenario 1: Assuming that the CRR person is correct

In this scenario, we assume that the CRR person is correct, meaning the upward movement probability in the real market p is equal to the risk-neutral probability q used by the CRR person, and the option payoff has the value same as the expected payoff from the CRR model. Equations (17) and (19) are applied to compute the NPI call option bound payoffs, and Equations (18) and (20) for the NPI put option bound payoffs. In terms of the CRR model, the expected payoffs are evaluated with Equations (1) and (2). As discussed in Section 3, each of the expected minimum selling payoff and the expected maximum buying payoff have an intersection point with the expected CRR payoff, and we note each two intersection points as s_1, s_2 for call option ($s_1 \leq s_2$) and s_3, s_4 for put option ($s_3 \leq s_4$). In this case the value s_q , the number of success historical data under the constraint $\frac{s_q}{n} = q$, is in the interval of two intersection points, $s_1 \leq s_q \leq s_2$ for call option and $s_3 \leq s_q \leq s_4$ for put option, for if $s = s_q$ the CRR expected value is between the lower and the upper expected values. Therefore, for call option there exist inequalities $\frac{s_1}{n} \leq q \leq \frac{s_2}{n}$ and for put option there exist inequalities $\frac{s_3}{n} \leq q \leq \frac{s_4}{n}$. After the analytic study for NPI payoff patterns, we learn for call option the maximum buying payoff and the minimum selling payoff increase as s increases whereas for put option they decrease as s increases. Then different trading actions of the NPI person according to different s are presented below.

For call option:

Case 1.1: $s \geq s_2$

In this case, because of $\frac{s}{n} \geq \frac{s_2}{n} \geq q$, the NPI person would be more optimistic than the CRR person about underlying stock future price, and the expected maximum buying price \underline{Pr}_c is higher than the fair price Pr_c^{CRR} from the CRR model, so the NPI person would like to buy a call option. Because in this scenario the CRR person is right, the loss of the NPI person in this case depends on this option exercise.

Under the situation that at the maturity this call option will be exercised the loss of the NPI person could be formulated as below. Obviously, if the NPI maximum buying price \underline{Pr}_c is quoted, so he needs to pay the buying price as a payment for this call option and gain the profit from the payoffs $S_T - K_c$. Then the loss L for the NPI investor in this case under this situation is:

$$\begin{aligned}
 L(n, m, s : s \geq s_2 | Pr = \underline{Pr}_c) &= \underline{Pr}_c - S_T + K_c \\
 &= \underline{E}_c[S_m - K_c]^+ - E_c^{CRR}[S_m - K_c]^+ \\
 &= \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \left[\binom{m+n}{m}^{-1} \binom{s+k-1}{k} \binom{n-s+m-k}{m-k} - \binom{m}{k} q^k (1-q)^{m-k} \right] \\
 &= \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{m}{k} \left[\binom{n+m}{s+k}^{-1} \binom{n}{s} \frac{s}{s+k} - q^k (1-q)^{m-k} \right]
 \end{aligned} \tag{21}$$

Here $\underline{Pr}_c - Pr_c^{CRR} = \underline{E}_c[S_m - K_c]^+ - E_c^{CRR}[S_m - K_c]^+$. Due to that the prediction of the CRR person is totally right, the payoff at the maturity $S_T - K_c$ will be equal to the CRR expected payoff $E_c^{CRR}[S_m - K_c]^+$.

In the other circumstance is that the NPI investor won't exercise this call option. So he will lose this

call option price \underline{Pr}_c .

$$\begin{aligned}
L(n, m, s : s \geq s_2 | Pr = \underline{Pr}_c) &= \underline{Pr}_c \\
&= \underline{E}_c[S_m - K_c]^+ \\
&= \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{m+n}{m}^{-1} \binom{s+k-1}{k} \binom{n-s+m-k}{m-k}
\end{aligned} \tag{22}$$

Case 1.2: $s_2 > s > s_1$

In this case, there is no action between the two investors, for the market price is higher than the NPI investor's expected maximum buying price and lower than his expected minimum selling price, $\underline{Pr}_c < Pr_c^{CRR} < \overline{Pr}_c$;

$$L(n, m, s : s_2 > s > s_1) = 0 \tag{23}$$

Case 1.3: $s \leq s_1$

In this case, the expected CRR price Pr_c^{CRR} is higher than the minimum selling price of the NPI person \overline{Pr}_c , so the NPI person would like to sell a call option. If we want to learn the loss of the NPI person L , two situations happen according to this option exercise. First, when the call option will be exercised eventually, longing a call option is the wise action. As we assume in this scenario assumptions of CRR person is totally correct, which means the opposite action taken by the NPI person is wrong, so under this assumption, the NPI person will face a huge amount to loss when he decides to sell a call option. In this case, there will be two part constituting this profit and loss, one part is the payoffs spread $S_T - K$, where S_T is the actual stock price at the maturity. The other part is the profit gained by selling this call option \overline{Pr}_c , and under our assumptions we use the expected payoff instead of the price $\overline{Pr}_c = \overline{E}_c[S_m - K_c]^+$. Because of the assumption about the perfection of the CRR model, that the stock price at the maturity equals to the expected stock price, then the payoff spread at the maturity is equal to the expected CRR payoff, $S_T - K = E_c^{CRR}[S_m - K_c]^+$. The formula below calculates the loss the NPI person will face when he quotes at the minimum selling price:

$$\begin{aligned}
L(n, m, s : s_1 \geq s | Pr = \overline{Pr}_c) &= S_T - K_c - \overline{Pr}_c \\
&= E_c^{CRR}[S_m - K_c]^+ - \overline{E}_c[S_m - K_c]^+ \\
&= \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \left[\binom{m}{k} q^k (1-q)^{m-k} - \binom{m+n}{m}^{-1} \binom{s+k}{k} \binom{n-s+m-k-1}{m-k} \right] \\
&= \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{m}{k} \left[q^k (1-q)^{m-k} - \binom{n+m}{s+k}^{-1} \binom{n}{s} \frac{n-s}{n-s+m-k} \right]
\end{aligned} \tag{24}$$

For the other situation, when this call option won't be exercised, selling a call option definitely is a good choice. Because doing that, the NPI person will earn the call option price. The loss of the NPI

person in this situation is negative as follows:

$$\begin{aligned}
L(n, m, s : s_1 \geq s | Pr = \overline{Pr}_c) &= -\overline{Pr}_c \\
&= -\overline{E}_c[S_m - K_c]^+ \\
&= -\sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{m+n}{m}^{-1} \binom{s+k}{k} \binom{n-s+m-k-1}{m-k}
\end{aligned} \tag{25}$$

Expected loss of the NPI person for a call option

As we have calculated losses of the NPI person comparing to the totally correct CRR person according to each different s cases, we would like to evaluate the expected loss. In view of what we have discussed above, s follows the binomial distribution $s \sim Bin(n, q)$ in this example, so the expected loss $L(q)$ can be formulated as:

If the call option will be exercised, the expected loss of the NPI person is

$$\begin{aligned}
&E_c[L(q)] \\
&= \sum_{s=0}^{s_1} L(n, m, s : s < s_1 | Pr = \overline{Pr}_c) \binom{n}{s} q^s (1-q)^{n-s} + \sum_{s=s_2}^n L(n, m, s : s > s_2 | Pr = \underline{Pr}_c) \binom{n}{s} q^s (1-q)^{n-s} \\
&= \sum_{s=0}^{s_1} \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{m}{k} \binom{n}{s} \left[q^{k+s} (1-q)^{n+m-s-k} - \binom{n+m}{s+k}^{-1} \binom{n}{s} \frac{n-s}{n-s+m-k} q^s (1-q)^{n-s} \right] \\
&+ \sum_{s=s_2}^n \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{m}{k} \binom{n}{s} \left[\binom{n+m}{s+k}^{-1} \binom{n}{s} q^s (1-q)^{n-s} \frac{s}{s+k} - q^{k+s} (1-q)^{n+m-s-k} \right]
\end{aligned} \tag{26}$$

If not, the expected loss of the NPI person is

$$\begin{aligned}
&E_c[L(q)] \\
&= \sum_{s=0}^{s_1} L(n, m, s : s < s_1 | Pr = \overline{Pr}_c) \binom{n}{s} q^s (1-q)^{n-s} + \sum_{s=s_2}^n L(n, m, s : s > s_2 | Pr = \underline{Pr}_c) \binom{n}{s} q^s (1-q)^{n-s} \\
&= -\sum_{s=0}^{s_1} \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{m+n}{m}^{-1} \binom{s+k}{k} \binom{n-s+m-k-1}{m-k} \binom{n}{s} q^s (1-q)^{n-s} \\
&+ \sum_{s=s_2}^n \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{m+n}{m}^{-1} \binom{s+k-1}{k} \binom{n-s+m-k}{m-k} \binom{n}{s} q^s (1-q)^{n-s}
\end{aligned} \tag{27}$$

For put option:

Case 1.4: $s \geq s_4$

The CRR price is higher than the minimum NPI selling price, so the NPI person would like to sell this put option and gain the put option price \overline{Pr}_p . If this put option has a negative payoff at the maturity,

then in this case the loss L can be calculated as:

$$\begin{aligned}
L(n, m, s : s \geq s_4 | Pr = \overline{Pr}_p) &= -\overline{Pr}_p \\
&= -\overline{E}_p[K_p - S_m]^+ \\
&= -\sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{n+m}{m}^{-1} \binom{s+k-1}{k} \binom{n-s+m-k}{m-k}
\end{aligned} \tag{28}$$

Here Pr is the price the NPI person quoted in the market equal to the minimum NPI selling price, and like what happened in the call option, rather than actual prices \overline{Pr}_p we use the minimum selling payoff $\overline{E}_p[K_p - S_m]^+$, for we try to avoid influences by discounted factor at the start of our study.

However, if this put option has a positive payoff, selling a put option is not smart, for it will lead some loss from this put option exercise by the CRR person, then the NPI person needs to pay the payoff $K_p - S_T$. The loss of the NPI person is larger:

$$\begin{aligned}
L(n, m, s : s \geq s_4 | Pr = \overline{Pr}_p) &= K_p - S_T - \overline{Pr}_p \\
&= E_p^{CRR}[K_p - S_m]^+ - \overline{E}_p[K_p - S_m]^+ \\
&= \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \left[\binom{m}{k} q^k (1-q)^{m-k} - \binom{n+m}{m}^{-1} \binom{s+k-1}{k} \binom{n-s+m-k}{m-k} \right] \\
&= \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{m}{k} \left[q^k (1-q)^{m-k} - \binom{n+m}{s+k}^{-1} \binom{n}{s} \frac{s}{s+k} \right]
\end{aligned} \tag{29}$$

The payoff of this put option at the maturity is identical to the expected value of this put option from the CRR model, $K_p - S_T = E_p^{CRR}[K_p - S_m]^+$, under the assumption of the CRR model perfection.

Case 1.5: $s_4 > s > s_3$

In this case, the CRR price is in the interval of NPI prices, so there is no transaction when this case is encountered. Therefore, in this case there is no loss.

$$L(n, m, s : s_4 > s > s_3) = 0 \tag{30}$$

Case 1.6: $s \leq s_3$

The CRR price is lower than the maximum NPI buying price, so he will buy a put option from the CRR person paying this put option price \underline{Pr}_p . If the NPI person buys a right to exercise from the market which he won't at the maturity, all he will lose is the put option price if he quotes at the maximum NPI buying price.

$$\begin{aligned}
L(n, m, s : s \leq s_3 | Pr = \underline{Pr}_p) &= \underline{Pr}_p \\
&= \underline{E}_p[K_p - S_m]^+ \\
&= \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{m+n}{m}^{-1} \binom{s+k-1}{k} \binom{n-s+m-k}{m-k}
\end{aligned} \tag{31}$$

Under the same assumption as the first scenario of this put option, the price is taken place by the payoff $Pr_p = E_p[K_p - S_m]^+$.

But if this put option will be exercised, the NPI person could get the payoff of this put option $K_p - S_T$. The loss in this situation will be:

$$\begin{aligned}
L(n, m, s : s \leq s_3 | Pr = \underline{Pr}_p) &= \underline{Pr}_p - K_p + S_T \\
&= E_p[K_p - S_m]^+ - E_p^{CRR}[K_p - S_m]^+ \\
&= \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \left[\binom{m+n}{m}^{-1} \binom{s+k-1}{k} \binom{n-s+m-k}{m-k} - \binom{m}{k} q^k (1-q)^{m-k} \right] \\
&= \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{m}{k} \left[\binom{m+n}{s+k}^{-1} \binom{n}{s} \frac{n-s}{n-s+m-k} - q^k (1-q)^{m-k} \right]
\end{aligned} \tag{32}$$

Expected loss of the NPI person for a put option

After calculating the loss in each case, we would like to explore the value of the expected loss for this put option with the same underlying stock according to $s \sim Bin(n, q)$, and the formula is listed below:

The put option payoff is negative, then the expected loss for this put option is

$$\begin{aligned}
E_p[L(s)] &= \sum_{s=0}^{s_3} L(n, m, s : s \leq s_3 | Pr = \underline{Pr}_p) \binom{n}{s} q^s (1-q)^{n-s} + \sum_{s=s_4}^n L(n, m, s : s \geq s_4 | Pr = \overline{Pr}_p) \binom{n}{s} q^s (1-q)^{n-s} \\
&= \sum_{s=0}^{s_3} \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{m+n}{m}^{-1} \binom{s+k-1}{k} \binom{n-s+m-k}{m-k} \binom{n}{s} q^s (1-q)^{n-s} \\
&\quad - \sum_{s=s_4}^n \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{n+m}{m}^{-1} \binom{s+k-1}{k} \binom{n-s+m-k}{m-k} \binom{n}{s} q^s (1-q)^{n-s}
\end{aligned} \tag{33}$$

Otherwise, the expected loss for this put option is

$$\begin{aligned}
E_p[L(s)] &= \sum_{s=0}^{s_3} L(n, m, s : s \leq s_3 | Pr = \underline{Pr}_p) \binom{n}{s} q^s (1-q)^{n-s} + \sum_{s=s_4}^n L(n, m, s : s \geq s_4 | Pr = \overline{Pr}_p) \binom{n}{s} q^s (1-q)^{n-s} \\
&= \sum_{s=0}^{s_3} \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{m}{k} \binom{n}{s} \left[\binom{m+n}{s+k}^{-1} \binom{n}{s} \frac{n-s}{n-s+m-k} q^s (1-q)^{n-s} - q^{k+s} (1-q)^{n-s+m-k} \right] \\
&\quad + \sum_{s=s_4}^n \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{m}{k} \binom{n}{s} \left[q^{k+s} (1-q)^{n-s+m-k} - \binom{n+m}{s+k}^{-1} \binom{n}{s} \frac{s}{s+k} q^s (1-q)^{n-s} \right]
\end{aligned} \tag{34}$$

An interesting characteristic is disclosed in the NPI expected loss formulas for both call and put option in this scenario. We started with imprecise NPI prices, but ended up getting a precise expected loss, because for each s , the trading action for the NPI person is determined compared to the CRR price, so only one NPI price is taken into account for each case, and the loss becomes an explicit value as action

price is settled. This explicit value of expected NPI loss is convenient for us to compare the two pricing methods.

Example 1

After discussing three trading cases for each type of option according to s , we would like to compare payoffs in an example. To start with the comparison, we need to define some input values in the example. For the binomial tree, the initial stock price $S_0 = 20$, and at every next step this stock price will either go up with the upward factor $u = 1.1$ or go down with the downward factor $d = 0.9$. We set the same strike price $K_c = K_p = 21$ for both the call option and the put option. We set a risk neutral probability q equal to 0.65, which is exactly identical to the real market probability of movements, $q = p = 0.65$. Since we assume the CRR model is totally right in this scenario, then the proportion of upward movements $\frac{s}{n}$ of historical data should follow the CRR prediction. To do this analytical study, understanding patterns of payoffs according to s and calculating the expected profit or loss of the NPI investor, if total historical data n is equal to 50, s will follow binomial distribution, $s \sim Bin(50, 0.65)$. In this example, we will first plot the patterns of all payoffs with fixed maturity, then we would like to know the expected loss of the NPI person with varying m . Finally because the NPI method is a method based on historical data, we want to check if n is increasing the expected loss of the the NPI person will get better.

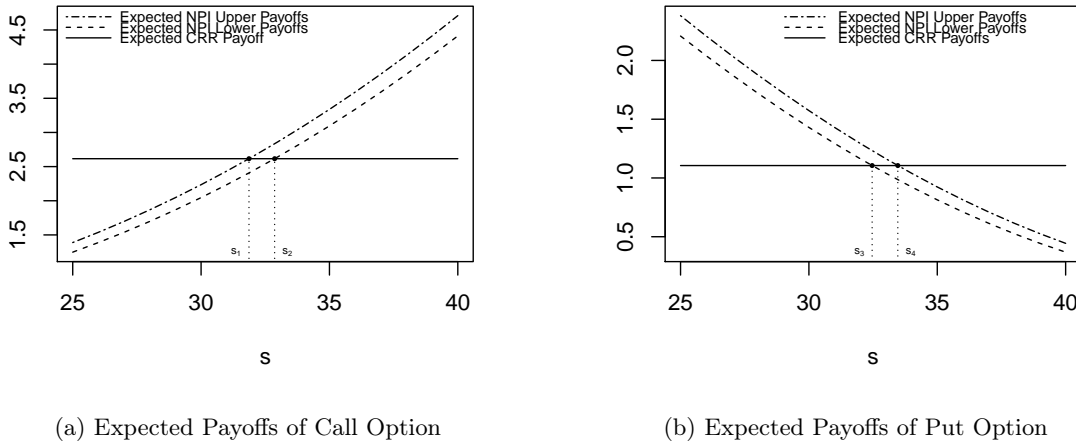


Figure 2: Expected Payoff of European Options from both the NPI method and the CRR model in Scenario 1

We would like to compare the two models with $m = 4$. Payoff patterns for call and put option are plotted in Figure 2, and it clearly displays that three cases for each type option we mentioned above happen in this example. In this example, the values of intersection points between NPI payoffs and the CRR payoff are gained using Newton's method. For a call option, the intersection s_1 value of the NPI upper payoff and the CRR payoff equals to 31.86541, and the intersection s_2 of the NPI lower payoff and the CRR payoff equals to 32.8654. For put option, s_3 equals to 32.46275 and s_4 equals to 33.46276, which are points of the NPI lower payoff intersecting with the CRR payoff and the NPI upper payoff intersecting with the CRR payoff. According to s values of intersection, we could tell when $\frac{s}{n}$ is equal to values near to q , the CRR payoff falls in the interval of NPI payoffs, no trading action exists in this

circumstance. When s falls outside the intersection interval $[s_1, s_2]$ or $[s_3, s_4]$, the NPI person and the CRR person will trade with each other, then NPI person will either gain profit from the CRR person or lost his money. Because we assume the CRR person is always right, so we expect NPI person to face an amount of loss during their trades. As the loss of the NPI person for different cases can be estimated, and we know the distribution s follows, the expected loss in this scenarios for the NPI person is available.

As we know, if the NPI person has decided to invest in an option based on a specific underlying stock, all input values are fixed except the number m of future steps. Then the influence of varying m toward expected NPI losses is in interest. Due to in this example, the call option will be exercised but the put option not, so Equations (26) and (33) are the formulas to calculate the expected loss of the NPI person, and we reveal expected NPI losses with various m in Figure 3.

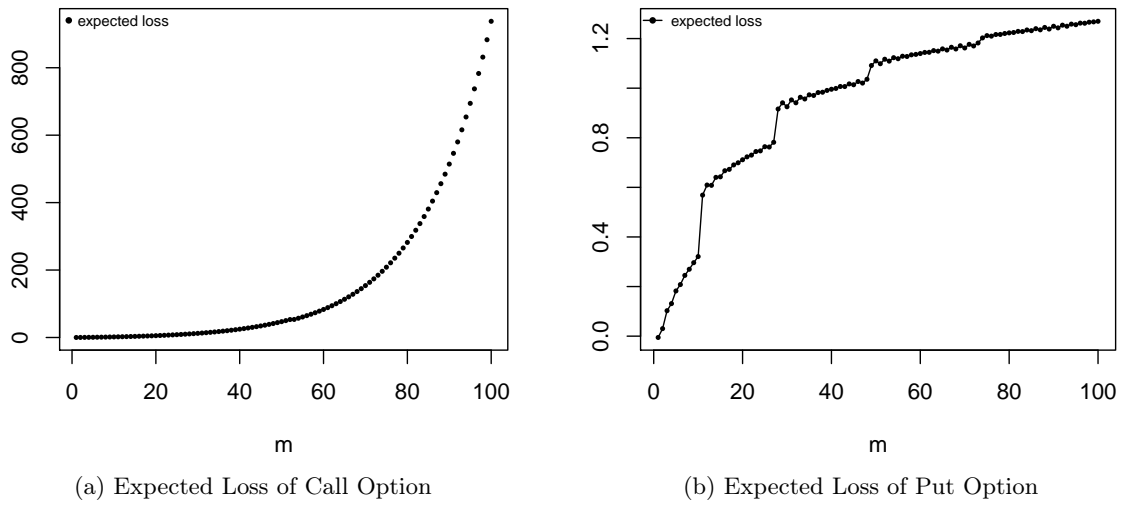


Figure 3: Expected Loss for the NPI person in Scenario 1

From Figure 3, there is no doubt no matter call or put option the NPI person decides to invest in and how long the maturity is, the NPI person is always expected to face an amount of loss. What is more, the expected loss manifests that it is wise to take part in short-term investment rather than long-term one for the NPI person, because the NPI expected loss increases more than linearly as m increases. The reason that the expected loss pattern shows the convexity as m increases is as m increase the pattern of the NPI payoff for call option gets more and more convex, and the part $s \geq s_2$ of the NPI expected payoff for call option takes a big part of the expected loss. However, there are some local disorders in both expected loss patterns, especially for the put option, the expected NPI loss is in a stairs type raise. The reason of these local disorders is payoffs' intersections movements when m increases, shown in Figure 4. In this figure, we plotted s integer value of varying intersection points, s_1 and s_2 for the call option and s_3 s_4 for the put option, along with increasing m . As we illustrated in expected loss formulas, the expected loss consists of differently weighted losses in different cases according to s , and intersection movements affect probabilities of each part losses, leading to local disorders. Another characteristic of intersections determine trends and locations of these disorders, when it comes to type s_1 , s_2 , s_3 , and s_4 in NPI pricing formulas, these values are supposed to be integers, and s_1 and s_3 are transferred to the first integer less than their values, while s_2 and s_4 become the first integer greater than their values. It

is clear in Figure 4. From the pattern in the figure, it is clear when we change intersection values into their corresponding integers, for call option there is a value step down of s_1 and s_2 which explains why there is a sudden decrease gap in the expected loss pattern for call option. For put option s_3 and s_4 increase values in steps with m , resulting in a stairs type expected loss increase. All these characteristics of s intersection values could explain local disorders of the expected loss.

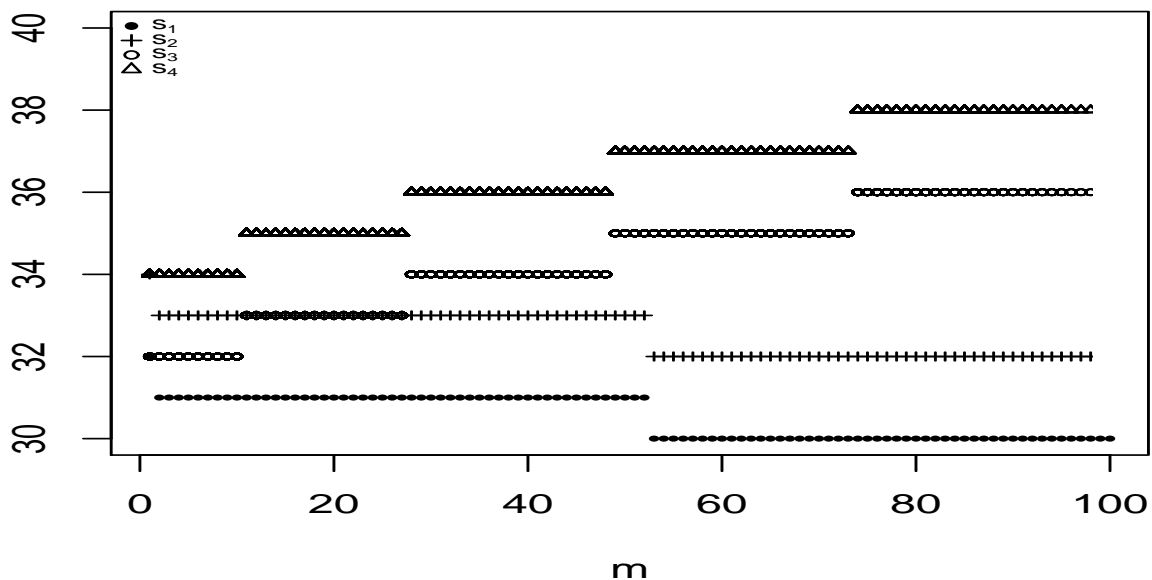


Figure 4: Intersection s move with varying m

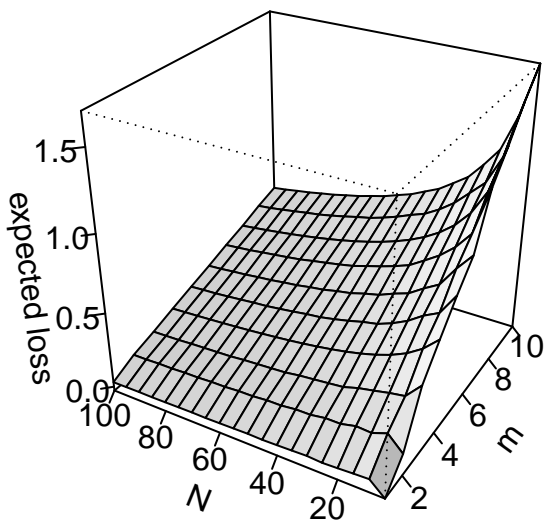


Figure 5: Influence on the expected loss with increasing historical data (call option): m is the number of future time steps and the number of the historical data $n = N * m$

Although under assumptions of this scenario the NPI person will always pay for his wrong prediction, this situation can be improved if more historical information can be reached. A 3D plot for call option, Figure 5 the expected loss of the NPI person with increasing historical data n and varying maturity m , supports this statement. It denotes that for each maturity m , as we increase n , the expected loss decreases, except when n and m are both very small. When n is small the interval between the maximum buying price and minimum selling price is very wide, and when m is very small the patterns of the NPI prices resemble a straight line. Therefore, a small amount of loss from the $s \geq s_2$ and a small amount of profit $s \leq s_1$ cancel each other out. However, as m is not too small, we can minimize the expected loss by increasing n . This is because, when n is small each increment of probability in each time step, for instance $\frac{s}{n+1}$ moving to $\frac{s+1}{n+2}$, changes greatly. Whereas for larger n , lower and upper probabilities at every step are more stable and approaching to q for calculating the expected loss of the NPI method in this scenario $s \sim Bin(n, q)$. Illustration from financial aspect also makes sense, which when an investor has more trustable historical information, his prediction is more accurate compared to the market, and there is less chance he will lose money. Overall, under the assumption that the CRR person knows every information to forecast a right price, the NPI person would not be expected to perform better than the CR person, and the longer the NPI person in this game the more expected losses he will give upon. However, because our method keeps learning from historical data, if there are more historical data available the loss will decrease.

4.2 Scenario 2: Assuming that the CRR person is wrong

In this scenario, there are also three possible trading actions the NPI person may take according to the value of s . Same as in Scenario 1, for call and put option there are two intersections between the NPI expected prices and the CRR expected price, and these intersections' positions depend on q . We note intersections for call option as s_5 s_6 ($s_5 \leq s_6$) and for put option as s_7 s_8 ($s_7 \leq s_8$). The relationship happening in the Scenario 1 still valid, $\frac{s_5}{n} \leq q \leq \frac{s_6}{n}$ for call option $\frac{s_7}{n} \leq q \leq \frac{s_8}{n}$. However, in this scenario the real market probability p is different from the risk-neutral probability q , and we assume the historical data can reflect the market at some level, $s \sim Bin(n, p)$. Thus, even though there still exist three cases of trading according to the value of s , the case that has the highest chance to occur is $\frac{s}{n}$ around p but q . We expect the NPI person will get some profit since the CRR person is wrong. The profit of the NPI person in three cases for each type option is listed below:

For call option:

Case 2.1: $s \leq s_5$

Then, when $s \leq s_5$ the NPI person would like to sell a call option to the CRR person and gain the call option price \overline{Pr}_c as the profit. If this call option will have a positive payoff, then selling a call option will cause some loss from this call option exercise by the CRR person, $S_T - K_c$. Then the profit of the

NPI person is:

$$\begin{aligned}
Pro(n, m, s : s \leq s_5 | Pr = \overline{Pr}_c) &= \overline{Pr}_c - S_T + K_c \\
&= \overline{E}_c[S_m - K_c]^+ - E_c^{CRR}(p) \\
&= \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \left[\binom{m+n}{m}^{-1} \binom{s+k}{k} \binom{n-s+m-k-1}{m-k} - \binom{m}{k} p^k (1-p)^{m-k} \right] \\
&= \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{m}{k} \left[\binom{n+m}{s+k}^{-1} \binom{n}{s} \frac{n-s}{n-s+m-k} - p^k (1-p)^{m-k} \right]
\end{aligned} \tag{35}$$

Here we neglect the difference between option payoffs and option premiums $\overline{Pr}_c = \overline{E}_c[S_m - K_c]^+$ because of the assumption that the contract settle date is close to the expiration date. As the call option payoff at the maturity is hard to estimate, we used the expected value from the CRR model with the probability p .

Another situation is this call option has a negative payoff, then the NPI person will earn this call option price without worrying about the CRR person will exercise it at the maturity. And the profit for this case can be formulated as follows:

$$\begin{aligned}
Pro(n, m, s : s \leq s_5 | Pr = \overline{Pr}_c) &= \overline{Pr}_c \\
&= \overline{E}_c[S_m - K_c]^+ \\
&= \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{m+n}{m}^{-1} \binom{s+k}{k} \binom{n-s+m-k-1}{m-k}
\end{aligned} \tag{36}$$

Case 2.2: $s_5 < s < s_6$

There is no transaction in this case, for the CRR price falls in the interval between the minimum selling price and the maximum buying price.

$$Pro(n, m, s : s_6 > s > s_5) = 0 \tag{37}$$

Case 2.3: $s \geq s_6$

When $s \geq s_6$ occurs, the CRR expected price is lower than the NPI maximum buying price, so the NPI person will buy this call option. So if at the maturity this call option will be exercised, then this trading is effective. The NPI person needs to pay this call option price but win the payoff at the maturity.

$$\begin{aligned}
Pro(n, m, s : s \geq s_6 | Pr = \underline{Pr}_c) &= S_T - K_c - \underline{Pr}_c \\
&= E_c^{CRR}(p) - \underline{E}_c[S_m - K_c]^+ \\
&= \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \left[\binom{m}{k} p^k (1-p)^{m-k} - \binom{m+n}{m}^{-1} \binom{s+k-1}{k} \binom{n-s+m-k}{m-k} \right] \\
&= \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{m}{k} \left[p^k (1-p)^{m-k} - \binom{n+m}{s+k}^{-1} \binom{n}{s} \frac{s}{s+k} \right]
\end{aligned} \tag{38}$$

When this call option won't be exercised at the maturity, although the CRR person made the wrong

prediction for the stock upward movement probability, the historical data is worse than the CRR prediction misleading the NPI person to a wrong decision, buying a call option, and this will cause an amount of loss. The loss is the price premium of the call option which is maximum buying price from our NPI method.

$$\begin{aligned}
L(n, m, s : s \geq s_6 | Pr = \underline{Pr}_c) &= \underline{Pr}_c \\
&= \underline{E}_c[S_m - K_c]^+ \\
&= \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{m+n}{m}^{-1} \binom{s+k-1}{k} \binom{n-s+m-k}{m-k}
\end{aligned} \tag{39}$$

Expected profit of the NPI person for a call option

After listing the profit and loss formulas for three cases, it is time to calculate the expected profit of the NPI person. In this scenario, $s \sim Bin(n, p)$ and the intersections depend on the expected value from the CRR model, so both p and q will influence the expected profit of the NPI person.

When the call option will be exercised, then the expected profit of the NPI person is

$$\begin{aligned}
&E_c[Pro(p, q)] \\
&= \sum_{s=0}^{s_5} Pro(n, m, s : s \leq s_5 | Pr = \overline{Pr}_c) \binom{n}{s} p^s (1-p)^{n-s} \\
&\quad + \sum_{s=s_6}^n Pro(n, m, s : s \geq s_6 | Pr = \underline{Pr}_c) \binom{n}{s} p^s (1-p)^{n-s} \\
&= \sum_{s=0}^{s_5} \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{m}{k} \binom{n}{s} \left[p^s (1-p)^{n-s} \binom{n+m}{s+k}^{-1} \binom{n}{s} \frac{n-s}{n-s+m-k} - p^{s+k} (1-p)^{n-s+m-k} \right] \\
&\quad + \sum_{s=s_6}^n \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{m}{k} \binom{n}{s} \left[p^{s+k} (1-p)^{n-s+m-k} - p^s (1-p)^{n-s} \binom{n+m}{s+k}^{-1} \binom{n}{s} \frac{s}{s+k} \right]
\end{aligned} \tag{40}$$

Note that this formula depends on q because values of intersections s_5 and s_6 are calculated according to q . When this call option won't be exercised at the maturity, the expected profit of the NPI person is

$$\begin{aligned}
&E_c[Pro(p, q)] \\
&= \sum_{s=0}^{s_5} Pro(n, m, s : s \leq s_5 | Pr = \overline{Pr}_c) \binom{n}{s} p^s (1-p)^{n-s} \\
&\quad - \sum_{s=s_6}^n L(n, m, s : s \geq s_6 | Pr = \underline{Pr}_c) \binom{n}{s} p^s (1-p)^{n-s} \\
&= \sum_{s=0}^{s_5} \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{m+n}{m}^{-1} \binom{s+k}{k} \binom{n-s+m-k-1}{m-k} \binom{n}{s} p^s (1-p)^{n-s} \\
&\quad - \sum_{s=s_6}^n \sum_{k=\lceil k_c^* \rceil}^m [u^k d^{m-k} S_0 - K_c] \binom{m+n}{m}^{-1} \binom{s+k-1}{k} \binom{n-s+m-k}{m-k} \binom{n}{s} p^s (1-p)^{n-s}
\end{aligned} \tag{41}$$

For put option:

Case 2.4: $s \leq s_7$

The CRR price is lower than the maximum NPI buying put option price which will result in the NPI person buy this put option from the CRR person. If this put option won't be exercised, then in this case the NPI person will lose his put option fee.

$$\begin{aligned}
L(n, m, s : s \leq s_7 | Pr = \underline{Pr}_p) &= \underline{Pr}_p \\
&= \underline{Pr}_p \\
&= \underline{E}_p[K_p - S_m]^+ \\
&= \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{m+n}{m}^{-1} \binom{s+k}{k} \binom{n-s+m-k-1}{m-k}
\end{aligned} \tag{42}$$

Here $\underline{Pr}_p = \underline{E}_p[K_p - S_m]^+$ means we assume the discounted procedure can be neglect in this step. When the put option will be exercised at the maturity, the NPI person will pay the put option price and gain the payoff of this put option $K_p - S_T$.

$$\begin{aligned}
Pro(n, m, s : s \leq s_7 | Pr = \underline{Pr}_p) &= K_p - S_T - \underline{Pr}_p \\
&= E_p^{CRR}(p) - \underline{Pr}_p \\
&= E_p^{CRR}(p) - \underline{E}_p[K_p - S_m]^+ \\
&= \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \left[\binom{m}{k} p^k (1-p)^{(m-k)} - \binom{m+n}{m}^{-1} \binom{s+k}{k} \binom{n-s+m-k-1}{m-k} \right] \\
&= \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{m}{k} \left[p^k (1-p)^{(m-k)} - \binom{n+m}{s+k}^{-1} \binom{n}{s} \frac{n-s}{n-s+m-k} \right]
\end{aligned} \tag{43}$$

As we could not know the real put option payoff in this scenario, we assume the real payoff is approximately equal to the value calculated by the CRR model with risk-neutral probability p .

Case 2.5: $s_7 < s < s_8$

In this case, s falls in between of s_7 and s_8 , where no trading action will occur, for the CRR price is higher than the maximum buying price and lower than the minimum selling price.

$$Pro(n, m, s : s_7 < s < s_8) = 0 \tag{44}$$

Case 2.6: $s \geq s_8$

In this case, the CRR expected price is higher than the minimum selling price, then the NPI person will sell this put option. If the CRR person is not able to exercise this put option at the maturity, the

NPI person will gain the price without any payment, then the profit in this case is:

$$\begin{aligned}
& \text{Pro}(n, m, s : s \geq s_8 | Pr = \overline{Pr}_p) = \overline{Pr}_p \\
& = \overline{E}_p[K_p - S_m]^+ \\
& = \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{m+n}{m}^{-1} \binom{s+k-1}{k} \binom{n-s+m-k}{m-k}
\end{aligned} \tag{45}$$

But when the stock price is not optimistic and the CRR person will exercise the put option, the NPI person will take a wrong action, selling the put option, which violates to the real market. The NPI person will face a payment as put option payoffs which is larger than the profit he could earn from selling put option price.

$$\begin{aligned}
& L(n, m, s : s \geq s_8 | Pr = \overline{Pr}_p) = K_p - S_T - \overline{Pr}_p \\
& = E_p^{CRR}(p) - \overline{E}_p[K_p - S_m]^+ \\
& = \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \left[\binom{m}{k} p^k (1-p)^{(m-k)} - \binom{m+n}{m}^{-1} \binom{s+k-1}{k} \binom{n-s+m-k}{m-k} \right] \\
& = \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{m}{k} \left[p^k (1-p)^{(m-k)} - \binom{n+m}{s+k}^{-1} \binom{n}{s} \frac{s}{s+k} \right]
\end{aligned} \tag{46}$$

Expected profit of the NPI person for a put option

Eventually, we could get the expected profit of put option computing as below:

When the stock price is optimistic and this put option won't be exercised, the expected profit of the NPI person is

$$\begin{aligned}
& E_p[\text{Pro}(p, q)] \\
& = \sum_{s=0}^{s_7} L(n, m, s : s \leq s_7 | Pr = \underline{Pr}_p) \binom{n}{s} p^s (1-p)^{n-s} \\
& \quad + \sum_{s=s_8}^n \text{Pro}(n, m, s : s \geq s_8 | Pr = \overline{Pr}_p) \binom{n}{s} p^s (1-p)^{n-s} \\
& = - \sum_{s=0}^{s_7} \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{m+n}{m}^{-1} \binom{s+k}{k} \binom{n-s+m-k-1}{m-k} \binom{n}{s} p^s (1-p)^{n-s} \\
& \quad + \sum_{s=s_8}^n \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{m+n}{m}^{-1} \binom{s+k-1}{k} \binom{n-s+m-k}{m-k} \binom{n}{s} p^s (1-p)^{n-s}
\end{aligned} \tag{47}$$

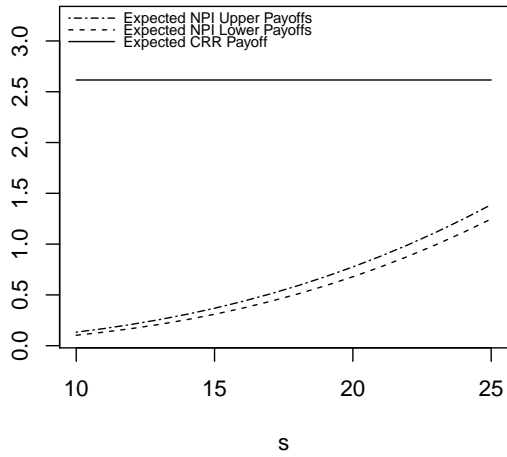
When this put option will be exercised, the expected profit of the NPI person is

$$\begin{aligned}
& E_p[Pro(p, q)] \\
&= \sum_{s=0}^{s_7} Pro(n, m, s : s \leq s_7 | Pr = \underline{Pr}_p) \binom{n}{s} p^s (1-p)^{n-s} \\
&\quad - \sum_{s=s_8}^n L(n, m, s : s \geq s_8 | Pr = \overline{Pr}_p) \binom{n}{s} p^s (1-p)^{n-s} \\
&= \sum_{s=0}^{s_7} \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{m}{k} \left[p^k (1-p)^{(m-k)} - \binom{n+m}{s+k}^{-1} \binom{n}{s} \frac{n-s}{n-s+m-k} \right] \binom{n}{s} p^s (1-p)^{n-s} \\
&\quad - \sum_{s=s_8}^n \sum_{k=0}^{\lfloor k_p^* \rfloor} [K_p - u^k d^{m-k} S_0] \binom{m}{k} \left[p^k (1-p)^{(m-k)} - \binom{n+m}{s+k}^{-1} \binom{n}{s} \frac{s}{s+k} \right] \binom{n}{s} p^s (1-p)^{n-s}
\end{aligned} \tag{48}$$

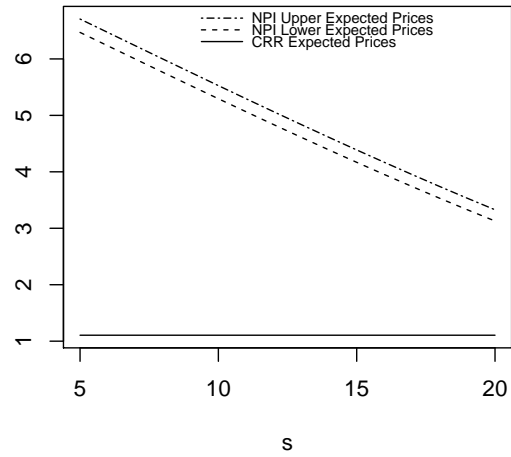
Example 2

We have learnt about how much profit or loss the NPI person will face in every case as well as the expected profit for him. Then we would like to illustrate the comparison in an example. As we will discuss the NPI method for European options versus the CRR model under the other assumption, the CRR model deviates from the real market value in the future, we input the risk-neutral probability $q = 0.65$, while the real market probability of upward movement is $p = 0.25$. This means based on information from the market, the stock's future is not bright, and its price will drop. However, the CRR person overvalues this stock believing its price will raise, whereas the NPI person has highly chance to predict it right base on the historical data. All other inputs in this simulation stays the same, $S_0 = 20, K_c = K_p = 21, u = 1.1, d = 0.9, n = 50$. Since q does not change, payoffs calculated from Equations (1) and (2) stay the same. But NPI results computed with Equations (17) and (19) for call option and Equations (18) and (20) for put option will be different, for the analytical study of payoff patterns according to s and expected profit or loss calculation in this scenario $s \sim Bin(50, 0.25)$ meaning $\frac{s}{n}$ will highly likely be around p rather than q . As Example 1, we would like to study the performance in three way; first we would like to learn the pattern of each expected value from both pricing method with fixed m , then knowing the expected profit of the NPI person with varying m is what we intend to do, finally we want to check the influence of n on the expected profit of the NPI person.

As an example, we predict option payoffs in four future steps, $m = 4$, and plot them in Figure 6. Actually, the whole patterns of NPI payoffs and the CRR payoff with s from 0 to 50 are the same as in Scenario 1, and intersections between our NPI payoffs and the CRR payoff are the same. But to distinguish from intersections in Scenario 1, we use different notations, then $s_5 = s_1 = 31.86541$, $s_6 = s_2 = 32.8654$, $s_7 = s_3 = 32.46275$ and $s_8 = s_4 = 33.46276$. The only different part is the area that s has a high chance to fall in, which is the part of payoffs we mainly focus on, shown in two graphs standing for each option type in Figure 6. On the basis of the figure and what we mentioned about whole patterns, we could make clear that for call option the NPI person have higher chance to sell the call option as well as for put option the NPI person is more willing to buy the put option from the CRR person, and both two actions gaining profit. There also exist possibilities that the historical data is even more wrong than the CRR model, issuing in the NPI person will buy a call option or sell a put option,



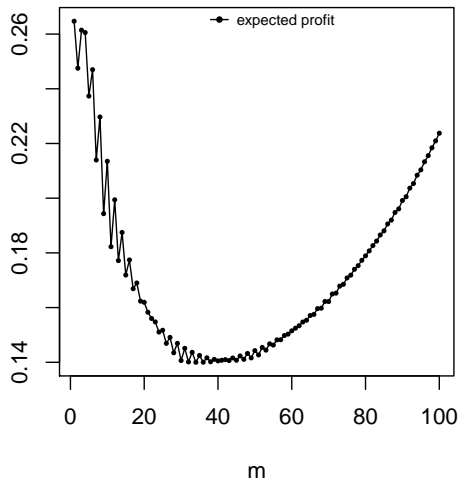
(a) Expected Payoffs of Call Option



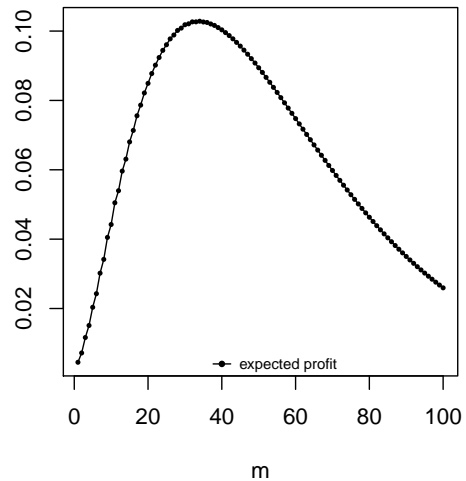
(b) Expected Payoffs of Put Option

Figure 6: Payoffs from the NPI method for European options and CRR model in Scenario 2

although the likelihood of that happening is quite low. Overall, we look forward to see in this scenario NPI person will gain some profit, and confirm this guess by plotting expected profit for both call and put option in Figure 7.



(a) Expected Profit of Call Option



(b) Expected Profit of Put Option

Figure 7: Payoffs from the NPI method for European options and CRR model in Scenario 2

Same as what happened in Scenario 1, after figuring out trading actions in all kinds of s cases, we know the exact quote price for each case, finally leading us to a precise expected profit or loss. Based on expected profit formulas Equations (41) and (48), it is easier for an investor to choose which maturity he wants to invest in with our NPI method. After all, once an option to a specific underlying asset has been settled the only factor which will influence the price is the maturity. We plot expected profit with varying m for call option and put option option in Figure 7.

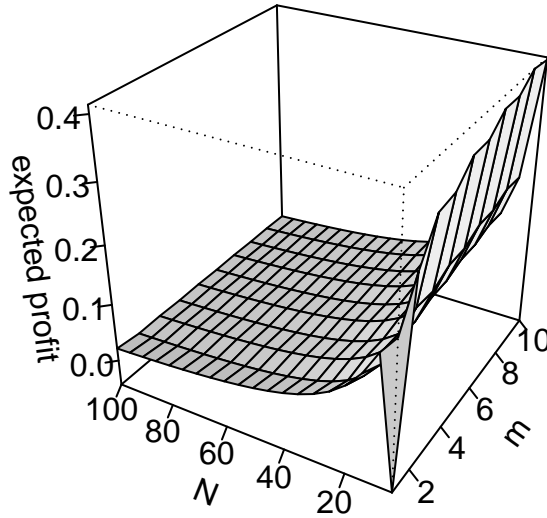


Figure 8: Influence on the expected profit with increasing historical data (put option): explanation m is the number of future time steps and the number of historical data $n = N * m$

It turns out, under the assumption that the CRR person does a very wrong prediction with opposite trading direction, the NPI person is expected to gain some positive profit no matter he invests which type of option. The fluctuation and the convex shape of the NPI profit for the call option is caused by price of this call option. As m increases, the pattern of NPI expected prices for this call option according to s becomes more convex. The concave shape of expected profits for the put option is caused by the competition between the payoff $K_p - S_T$ and the option price, for they have opposite moving directions when m increases. In our example, the stock price will end up going down, so playing with a put option is a safer choice with less profit, because the more risk exists the higher return an investor can get. In general, according to the predicted direction of stock price movement, buying a relevant option is better and safer than selling an opposite direction option, which is already commonly agreed in the real market. When we increase n in this scenario the profit of the NPI person will reduce, except when n and m are both small. The reason of no profit with small n and m is the same as Example 1, and we plotted the NPI expected profit for put option to confirm our point in Figure 8. Like what we have discussed in Scenario 1, enlarging historical data makes the forecasting from the NPI method close to market behaviour with more stable movement probabilities in the binomial tree, then the difference between the NPI prediction and the CRR prediction narrows down. From the perspective of financial market, the more effective prediction closing to the real market an investor gets, the less opportunity he has to beat the market, for he will never take an action if his prediction is the same as the real market.

Example 3

In this example, we want to see if the CRR model predicts a wrong probability but with the same direction as the real market, $p > 0.5$ and $q > 0.5$ with $q \neq p$ or $p < 0.5$ and $q < 0.5$ with $q \neq p$. All

inputs in this example are same as in Example 2, except $q = 0.45$, and because q value is changed, the intersections with fixed m are different from those in Example 2. However, the most interesting problem is the expected profit of the NPI person in this example.

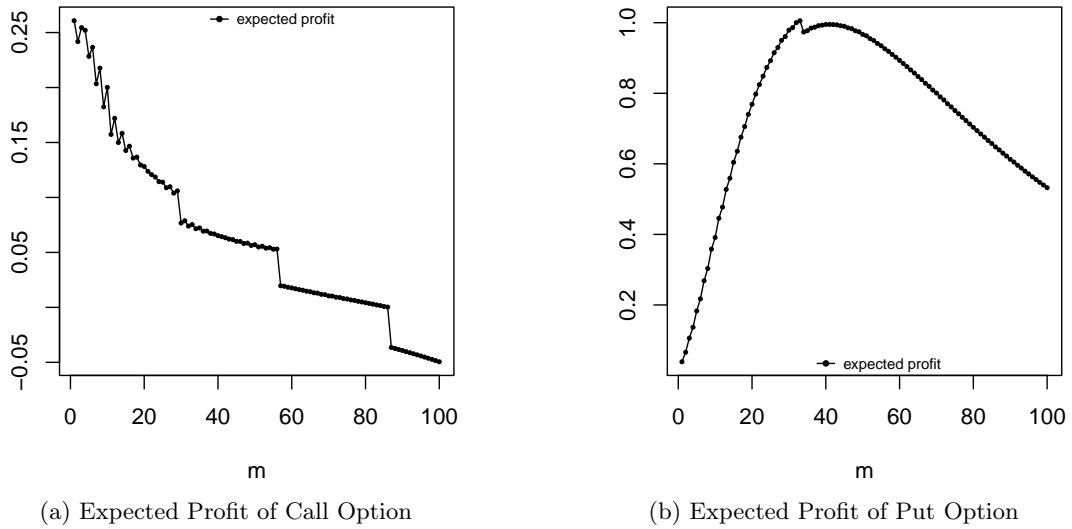


Figure 9: Payoffs from the NPI method for European options and CRR model in Scenario 2

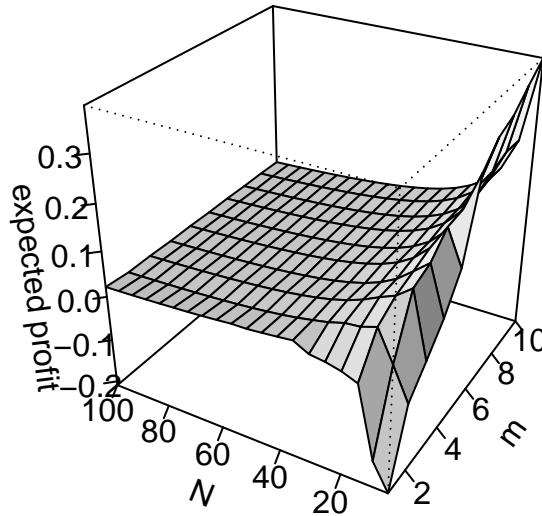


Figure 10: Influence on the expected profit with increasing historical data (put option): explanation m is the number of future time steps and the number of historical data $n = N * m$

Even though in this example the CRR prediction and the real market have the same direction, the stock price will go downwards, so we should still use Equations (41) and (48) to calculate the expected

profit of the NPI person. The results are plotted in Figure 10, showing that when n and m are both small the NPI person will face some loss, because the historical information is not enough for the NPI person to act effectively and small m makes the NPI call option payoff pattern according s less steeper for the area holding greatest probability. However, if we increase n , both profit and loss will approach to zero as we have already discussed in Examples 1 and 2.

We have performed a more detailed study of the expected profit of the NPI person for varying values of q , given p . In Table 1, we set the real market probability $p = 0.25$, and $S_0 = 20, K_c = K_p = 21, u = 1.1, d = 0.9$. We calculated the expected profit according to different q . In this study, on the basis of 40 historical data we discover that for a difference between q and $p = 0.25$ is greater than 0.1, the NPI person is expected to gain some profit by investing in either a European call or put option with maturity identical to 2. As n increases, the absolute value of the expected profit narrows down like what displays in Figures 8 and 10. The reason is when n increases, meaning the NPI person has more information from the market, the interval between the maximum buying price and the minimum selling price gets smaller, approaching to the fair market price, this gives fewer opportunities for the NPI person to beat the market. Therefore, along with increasing n , the NPI person will gain less profit from a two time-step European option with a large difference between p and q , while lose less money with a small difference between p and q . If we increase the maturity to $m = 4$ time steps, with fixed n , the trend of results referring to different levels of differences between p and q is identical to the one when $m = 2$. But for the longer term option, $m = 4$, based on the same fixed corresponding historical data, the NPI person will face more loss when the p and q are close to each other comparing to that when $m = 2$. Thus if there exists more data information given a fixed option maturity, the interval of the difference between q and p leading to a negative profit for the NPI person are smaller than those based on less data information. We have investigated further cases, including other values of p , for the problem of the expected profit according to differences between p and q is similar as discussed above.

Table 1: Expected profit and loss changing with p and q difference ($p = 0.25$)

$p = 0.25$	Expected Profit											
	$m = 2$						$m = 4$					
	$n = 40$		$n = 100$		$n = 200$		$n = 40$		$n = 100$		$n = 200$	
q	call	put	call	put	call	put	call	put	call	put	call	put
0.15	-0.22	0.00	-0.21	-0.10	-0.20	0.01	-0.23	0.00	-0.19	0.03	-0.18	0.02
0.25	-0.17	-0.24	-0.11	-0.17	-0.08	-0.13	-0.19	-0.46	-0.12	-0.34	-0.09	-0.25
0.35	0.13	-0.10	0.21	0.01	0.21	0.02	0.05	-0.18	0.18	0.02	0.19	0.03
0.45	0.25	0.07	0.22	0.04	0.21	0.02	0.25	0.14	0.21	0.07	0.19	0.04
0.55	0.26	0.09	0.22	0.04	0.21	0.02	0.28	0.18	0.22	0.07	0.19	0.04
0.65	0.26	0.09	0.22	0.04	0.21	0.02	0.28	0.18	0.22	0.07	0.19	0.04
0.75	0.26	0.09	0.22	0.04	0.21	0.02	0.28	0.18	0.22	0.07	0.19	0.04

5 Concluding Remarks

The NPI method for European option pricing, a method learning from historical data, relaxes some classic assumptions, one of the most important one is we do not assume probabilities of upward movement

remain constant. After setting up the NPI method for European options, we compared our model with the CRR model. In this analytical study, two extreme scenarios were investigated. Scenario 1 is the CRR person predicts with the same knowledge as that in the market, the CRR price is identical to the market price. In this scenario, the NPI person is not expected to beat the CRR person, but with more historical information the NPI person performs better. Scenario 2 is the opposite of Scenario 1, the CRR person made a mistake during prediction. In this scenario, investing in a corresponding option with the same direction as your prediction is a good move for a NPI investor. As the prediction from the CRR model gets closer to the truth, the advantages of the NPI method dwindle. Because the NPI method for European options generates an interval of prices, by integrating it with the CRR model there exist some decision routes for investors trading based on historical data. Chen, Coolen and Coolen-Maturi study two decision routes in their paper and results are satisfying (Chen et al., 2017). Another point to be acknowledged is owing to the assumption of ignoring the discounted procedure. Therefore, it is more convinced to use our NPI method in short term investment. For long term investment, we need to explore the NPI option pricing method with discounted procedure, and after that we could apply our model to the American option evaluation.

A further topic of interest for further study in our method is whether or not all historical data should be taken into account. It is clearly good to do so if one can safely assume that the future observations will be exchangeable with all the past one. However, if one believes that there has been a considerable change in the data at some point in the past ones, it may be appropriate to restrict the historical data to observations after such a change.

This paper presents in the way that helps us learn the properties of the NPI method for European option pricing. We only consider the basic binomial tree model as a simple first step of this research. This method is only used for ideal model situations, but it underpins a range of more realistic models for which we aim to investigate NPI in the future. We study the NPI method only by comparing its performance in regard to another trader who would use the CRR model, either with perfect or imperfect knowledge. Real world scenarios with multiple traders are also interesting for further research. It is also interesting to investigate the application of the NPI method for European option pricing in real markets where it may also be possible to improve the method by creating hybrid strategies based on multiple pricing methods.

Acknowledgement

The authors are grateful to two reviewers of this paper for comments that led to improved the presentation.

References

Arnold, T., and T.F. Crack, 2003, Option pricing in the real world: a generalized binomial model with applications to real options, *4th Real Options International Conference* <http://hdl.handle.net/10523/1548>.

- Augustin, T., and F.P.A. Coolen, 2004, Nonparametric predictive inference and interval probability, *Journal of Statistical Planning and Inference* **124**, 251–272.
- Baker, R.M., T.A. Coolen-Maturi, and F.P.A. Coolen, 2017, Nonparametric predictive inference for stock returns, *Journal Of Applied Statistics* **44**, 1333–1349.
- Berleant, D., L. Andrieu, J.P. Argaud, F. Barjon, M.P. Cheong, M. Dancre, G. Sheble, and C.C. Teoh, 2008, Portfolio management under epistemic uncertainty using stochastic dominance and information-gap theory, *International Journal of Approximate Reasoning* **49**, 101–116.
- Chen, J., F.P.A. Coolen, and T. Coolen-Maturi, 2017, Application of frequentist imprecise probability method in saymmetric binomial asset and european option risk pricing model, *7th International Conference of the Financial Engineering and Banking Society* .
- Coolen, F.P.A., 1998, Low structure imprecise predictive inference for bayes' problem, *Statistics and Probability Letters* **36**, 349–357.
- Coolen, F.P.A., 2011, *Nonparametric Predictive Inference*, In : *International Encyclopedia of Statistical Science* (M.Lovric,ed., Springer, Berlin).
- Coolen, F.P.A., T. Coolen-Maturi, and A.H. Al-nefaiee, 2014, Nonparametric predictive inference for system reliability using the survival signature., *Journal of Risk and Reliability* **42**, 437–448.
- Coolen, F.P.A., P. Coolen-Schrijner, T. Coolen-Maturi, and F.F. Elkhaffi, 2013, Nonparametric predictive inference for ordinal data, *Communications in Statistics - Theory and Methods* **42**, 3478–3496.
- Coolen-Schrijner, P., and F.P.A. Coolen, 2007, Nonparametric adaptive age replacement with a one-cycle criterion, *Reliability Engineering and System Safety* **92**, 74–84.
- Coolen-Schrijner, P., S.C. Shaw, and F.P.A. Coolen, 2009, Opportunity-based age replacement with a one-cycle criterion, *Journal of the Operational Research Society* **60**, 1428–1438.
- Cox, J.C., S.A. Ross, and M. Rubinstein, 1979, Option pricing: A simplified approach, *Journal of Financial Economics* **7**, 229–263.
- De Finetti, B., 1974, *Theory of Probability* (Wiley, London).
- Hill, B.M., 1968, Posterior distribution of percentiles: Bayes theorem for sampling from a population, *Journal of the American Statistical Association* **63**, 677–691.
- Jackwerth, J.C., and M. Rubinstein, 1996, Recovering probability distributions from option prices, *The Journal of Finance* **51**, 1611–1631.

- Jaffray, J.Y., and M. Jeleva, 2008, Information processing under imprecise risk with an insurance demand illustration, *International Journal of Approximate Reasoning* **49**, 117–129.
- Muzzioli, S., and H. Reynaerts, 2008, American option pricing with imprecise risk-neutral probabilities, *International Journal of Approximate Reasoning* **49**, 140–147.
- Walley, P., 1991, *Statistical Reasoning with Imprecise Probabilities* (Chapman and Hall, London).
- Weichselberger, K., 2001, *Elementare Grundbegriffe einer Allgemeineren Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als Umfassendes Konzept* (Physika, Heidelberg).