VOL. 10, NO. 4, MARCH 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

A NOVEL DECISION TREE APPROACH FOR OPTION PRICING USING A CLUSTERING BASED LEARNING ALGORITHM

J. K. R. Sastry, K. V. N. M. Ramesh and J. V. R. Murthy
KL University, JNTU Kakinada, India
E-Mail: drsastry@kluniversity.in

ABSTRACT

Decision tree analysis involves forecasting future outcomes and assigning probabilities to those events. One of the most basic fundamental applications of decision tree analysis is for the purpose of option pricing. The binomial tree would factor in multiple paths that the underlying asset's price can take as time progresses. The price of the option is calculated using the discrete probabilities and their associated pay-offs at maturity date of the option. In this work we came up with an approach to build a binomial decision tree that can be used to price European, American and Bermudian options and a methodology to train the decision tree using a clustering based learning algorithm that minimizes the mean square error (MSE) between the observed and predicted option prices. The training methodology involves clustering the options based on moneyness and fit a linear equation for each cluster to calculate the confidence that needs to be used in building the binomial decision tree for a particular strike price within the cluster. It is observed that the MSE for option price using the proposed model is less when compared to the Black-Scholes model for the proposed learning algorithm.

Keywords: option pricing, clustering, decision tree, binomial option pricing.

1. INTRODUCTION

A derivative [13] is an agreement between two parties that has a value derived on the underlying asset. There are many kinds of derivatives with most notable being swaps, futures and options. An option [13] is a financial derivative that represents contract sold by one party (option writer) to another party (option holder). The contract offers the buyer the right, but not the obligation, to buy (call) or sell (put) a security or other financial asset at an agreed-upon price (the strike price) during a certain period of time or on a specific date (exercise date). The Black-Sholes formula [2] presented the first pioneering tool for rational valuation of options. There are several assumptions, used to derive the original Black Sholes model, relaxation of which had been reported in the literature: No dividends Relaxed in [15], No taxes nor transaction costs, Constant interest rates relaxed in [15], No penalties for short sales, Continuous market operation relaxed in [16], Continuous share price relaxed in [7], Lognormal terminal stock price return relaxed in [14]. In addition, Black-Sholes model assumes; continuous diffusion of the underlying relaxing which resulted in jump diffusion model [15], constant standard deviation/volatility, and no effect on option prices from supply/demand. These models improve pricing performance and generalize Black-Sholes formula to a class of models referred to as the modern parametric option pricing models. Modern parametric option pricing models which are a generalization to the Black-Sholes model are more complex and have poor out-of-sample performance and use implausible and/or inconsistent implied parameters. They often produce parameters inconsistent with underlying time-series and inferior hedging and retain systematic price bias they were intended to eliminate [3], [4].

Prompted by shortcomings of modern parametric option-pricing, new class of methods was created that do not rely on pre-assumed models but instead try to uncover/induce the model, or a process of computing prices, from vast quantities of historic data. Many of them utilize learning methods of Artificial Intelligence. Non-parametric approaches are particularly useful when parametric solution either; lead to bias, or are too complex to use, or do not exist at all. The purest version of non-parametric option-pricing methods, are model-free methods. They involve no finance theory but estimates option prices inductively using historical or implied variables and transaction data. Although some form of parametric formula usually is involved, at least indirectly, it is not the starting point but a result of an inductive process. There are several methods in this group:

- Model-free option pricing with Genetic Programming (GP)
- Model-free option-pricing with kernel regression
- Model-free option-pricing with Artificial Neural Networks (ANN)

The independence of model-free approaches from any finance theory means prices produced by them may not conform to rational pricing and/or may not capture restrictions implied by arbitrage [10]. To improve model-free approaches in this respect, constraints have to be introduced [5]. There are several ways used to enforce rational pricing into model-free pricing; The Equivalent Martingale Measure (EMM) adjusts prices to reflect a preference-free, risk-neutral market. In risk-neutral economy all assets must earn the same return [6]. Under the risk-adjusted probability distribution, the stock price follows a Martingale (a stochastic process where the best forecast of tomorrow's price is today's) and is arbitrage-free. Non-parametric adjustments to Black-Sholes estimate a portion of the option-pricing non-parametrically while retaining the conventional option-pricing framework to guarantee rational-pricing. Generalized Deterministic Volatility estimates unknown volatility either parametrically or non-parametrically and inserts this estimate into a conventional model. The three sample approaches in this category are:

- Implied Binomial Tree [18], [19].
- Generalized Deterministic volatility functions [9]
- Kernel approach [1], [11].

Generalized volatility approaches have their costs and benefits. The implied tree approach for example can help with estimation of exotic, path dependent options where no analytical formula exists.

Option pricing using binomial decision tree allows to price even American or Bermudian options which can be exercised at any time or at specific dates respectively with in maturity. A method to generate binomial option pricing has been proposed by Cox J *et al.* [8]. However it assumes that the underlying returns are log-normally distributed. A novel approach [21] has been proposed to generate a binomial decision tree which uses chebyschev inequality to determine the underlying future states and also uses absolute returns instead of logarithmic returns in determining the volatility used in generating the binomial decision tree. A methodology to build the binomial decision tree with implied volatility incorporated in it has been proposed [20]. The model allows changing the implied volatility used in building the tree through a parameter called "k". In real world markets, it has been observed that the implied volatility is a function of strike price and maturity of the option. Therefore for a given maturity the implied volatility is a function of strike price. Any option pricing model has to be trained to attain the market prices by using certain training data in which the pricing parameters are adjusted to attain the known market prices.

In this work we came up with a training method that fits a linear equation for calculating the confidence ($1-1/k^2$) to be used in binomial decision tree and implied volatility to be used in the Black-Scholes model on a set of options that are clustered based on moneyness (S/S_k). The mean square error (MSE) in the observed option prices and predicted option prices is calculated. The predicted prices with the proposed model are also compared with the predicted prices from the Black-Scholes Model. It is observed that the MSE obtained using the proposed model is less than the MSE obtained using the Black-Scholes model.

The rest of the paper is organized as follows; section 2 describes the method to generate binomial decision tree and the data model used for clustering algorithm and the parameters used in pricing the option. Section 3 describes training the model using the clustering methodology on options and curve fitting for confidence and various tests that are performed on the pricing model. Section 4 describes the results and discusses the accuracy of the model with and without clustering. Section 5 discusses the future work that can be done in this area.

2. DECISION TREE CONSTRUCTION AND OPTION PRICING

Markets will be either bullish or bearish. The value of an asset increases in a bullish market and decreases in a bearish market. Hence in a binomial decision tree only two states are considered for the movement of the underlying asset one is the up-move and the other is the down move. The magnitude of the up or down moves depends on the volatility of the underlying asset. For a known volatility of a random process the chebyschev inequality [6] states that the probability that the values of the random process lies with-in "k" standard deviations (volatility) from the mean is given by ($1-1/k^2$). Considering the return on an underlying (equity or Index) as a random process with zero mean and variance $\sigma^2 t$, the probability that the underlying value lies between the upper boundary $S_t+k\sigma\sqrt{t}$ and $S_t-k\sigma\sqrt{t}$ is given by equation(1).

$$P[S_t-k\sigma\sqrt{t} \leq S \leq S_t+k\sigma\sqrt{t}] \geq 1-1/k^2 \quad (1)$$

Considering that the probability of an up move as "p" the probability of down move is "1-p". Figure 1 shows the movement of the underlying and its associated probabilities. Considering arbitrage free pricing with risk-free interest rate as "r" the equation (2) has to be satisfied where the expected value of the underlying on a future date should be equal to the continuously compounded value at risk-free rate.

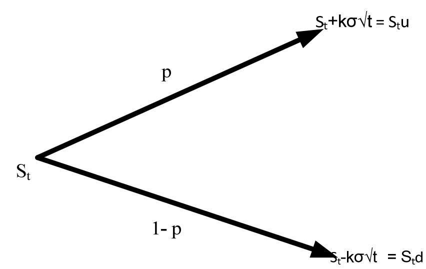

Figure-1. One step binomial tree.

$$p(S_t+k\sigma\sqrt{t}) + (1-p)(S_t-k\sigma\sqrt{t}) = St^*e^{rt} \quad (2)$$

$$p = 0.5 \{ 1 + (S_t * (e^{rt} -1) / k\sigma\sqrt{t}) \} \quad (3)$$

$$u = 1 + k\sigma\sqrt{t} / S_t \quad (4)$$

$$d = 1 - k\sigma\sqrt{t} / S_t \quad (5)$$

The value of the underlying asset at step "n" attained due to "r" up moves and "n-r" down moves is given by equation (6). The corresponding nodal probability is given by equation (7).

$$S_{t+n}(r) = S_t u^r d^{(n-r)} \quad (6)$$

$$P(S_{t+n}(r)) = (n!/(r!*(n-r)!)\, p^r (1-p)^{n-r} \quad (7)$$

Once the decision tree is constructed the price of the option is determined by calculating the pay-off at the leaf nodes of the tree. The pay-off for call and put options are given in equations (8) and (9) respectively.

$$P_c = \max(0,\ S - S_k) \tag{8}$$

$$P_p = \max(0, S_k - S) \tag{9}$$

The call option premium is calculated by using equation (10).

$$C = p*(\max(0, S_t u - S_k)) + (1-p)*(\max(0, S_t d - S_k)) * e^{-rt} \tag{10}$$

Similarly the put option premium is calculated by using equation (11).

$$P = p*(\max(0, S_k - S_t u)) + (1-p)*(\max(0, S_k - S_t d)) * e^{-rt} \tag{11}$$

Thus the value of the call/put option is the discounted value of a weighted average of the expiration date value of the call.

2.1 Data model

The data required for predicting the option price for various strike prices on a particular underlying are the underlying time series to calculate the volatility (σ), the data related to options traded on the underlying at a particular strike and maturity and the risk free rate of interest. The schema of respective data tables is given in Tables-1, 2 and 3.

Table-1. Underlying time series.

S. No.	Attributes
1	Underlying Symbol
2	End Of Day Underlying Price
3	Date

Table-2. Option data.

S. No.	Attributes
1	Underlying Symbol
2	Option Type
3	Strike Price(S_k)
4	Maturity(t) in months
5	Option Premium
6	Current Underlying Price(S_t)

Table-3. Interest rate.

S. No.	Attributes
1	Currency Symbol
2	Time Period
3	Interest Rate (r)

From the underlying time series, the daily returns are calculated using the differences in the historical end of day underlying price between two successive dates i.e., (S_{t+1}-S_t). The returns over a period of latest 252 trading days is used to calculate the annual volatility(σ) of the underlying which is used in building the underlying binomial decision tree.

The option data is used to train and test the proposed model for option pricing that uses the binomial decision tree built using the data from tables 1, 2 and 3 along with other input parameters as mentioned below.

1. Current Underlying Price
2. Volatility
3. Interest Rate
4. Number of Steps
5. Option Maturity
6. Confidence

3. MODEL TRAINING AND TESTING

At any point of time the parameters that are used in option pricing are directly observed from the market except the volatility. The volatility depends on the historical time series that is used to calculate the returns. It is observed in the real world that the volatility used in pricing the options of various moneyness on the same underlying and same maturity varies. Hence a methodology to calculate the volatility based on the moneyness has to be established so that the prices calculated using the model matches the market prices. The methodology proposed fits a linear equation to calculate the confidence ($1-1/k^2$) to be used in the construction of binomial decision tree that minimizes the error between the observed market price and the model predicted price. The value of “k” that is obtained from calculated confidence is a multiplying factor for historical volatility converting it into implied volatility. Various clustering algorithms [12] and a comparison of the same have been presented by Osma Abu Abbas [17]. The proposed clustering algorithm forms clusters based on moneyness bounds. The curve fitting is done for each cluster of options, clustered based on the moneyness and the MSE in each case is calculated.

3.1 Clustering and curve fitting algorithm

1. Collect all options that have same maturity, option type, and underlying and different strike price.
2. Calculate the moneyness using the formula S_t / S_k
3. Compute 5 clusters based on the below conditions
   a. Cluster-1 if ($S_t / S_k < 0.85$)
   b. Cluster-2 if ($0.85 \le S_t / S_k < 0.95$)
   c. Cluster-3 if ($0.95 \le S_t / S_k < 1.05$)
   d. Cluster-4 if ($1.05 \le S_t / S_k < 1.15$)
   e. Cluster-5 if ($1.15 \le S_t / S_k$)
4. Within each cluster pick the maximum, minimum and median strike prices
5. Compute the value confidence ($1-1/k^2$) and implied volatility that needs to be used in the binomial decision tree and Black-Scholes model respectively for the strike prices obtained in the above step so that the option price calculated using the binomial decision tree and Black-Scholes model matches the observed price.

6. Fit a linear equation using the values of (strike price, $1-1/k^2$) and (strike price, implied volatility) taking the pairs of (median, maximum)and (median, minimum) and use them to calculate the confidence and implied volatility that needs to be used for strikes on either side of the median respectively.
7. For any new strike price for which the option price has to be calculated, compute the moneyness and associate the moneyness to a cluster as defined in step 3. Compare the moneyness with the moneyness of the median of a cluster and use one of the linear equations that defines the confidence/implied volatility on either sides of the median based on whether the moneyness is less than or greater than the median moneyness.

The model is trained using the put option data of Index option on S&PCNX Nifty. The data for put options expiring on April 24th with one and two month maturity period i.e with March 24th and February 24th as start date respectively are taken for training and cross validation. The MSE on the predicted values using test data is calculated using equation (12).

$$MSE = 1/N \sum_{j=0}^{N} (observationj - predictionj)^2 \quad (12)$$

The cross-validation (CV) accuracy is measured in terms of mean square error (MSE).

3.2 Curve fitting example

The closing value of Nifty on 24th March 2014 is 6583.5. The median strike price in the cluster-3 composing of at-the-money options is 6600. The confidence which matches the corresponding option price of 91.3 is 95.98. Similarly for option with strike price of 6800 the confidence that matches the corresponding option price of 196.35 is 94.61. Therefore equation (13) governs the confidence for a given strike price greater than median strike price of 6600.. Similarly equation (14) gives the implied volatility to be used in the Black-Scholes model.

$$\text{Confidence} = 95.98+m*(\text{Strike} - 6600) \quad (13)$$

Where

$$m = (94.61-95.98)/ (6800-6600)$$

$$\text{Implied Volatility} = 14.47+m*(\text{Strike} - 6600) \quad (14)$$

Where

$$m = (12.49-14.47)/ (6800-6600)$$

For strike prices less than 6600 the confidence and implied volatility are given by equation (15) and (16) respectively.

$$\text{Confidence} = 95.98+m*(\text{Strike} - 6600) \quad (15)$$

Where

$$m = (96.86-95.98)/ (6400-6600)$$

$$\text{Implied Volatility} = 14.47+m*(\text{Strike} - 6600) \quad (16)$$

Where

$$m = (16.45-14.47)/ (6400-6600)$$

Equations (13) and (15) gives the confidence that have been used in building the binomial decision tree to get the option price for strike prices on either side of the median. Equations (14) and (16) give the implied volatility to be used in the Black-Scholes model to obtain option price.

3.3 Model testing

An option pricing model that follows rational pricing should satisfy the principles given below:

1. Arbitrage free pricing. An option pricing model is said to be arbitrage free if it satisfies the put-call parity for which the equation below is satisfied on non-dividend paying stock.

$$C + S_k * e^{-rt} = P + S_t$$

Where

C = Call Option Price
P = Put Option Price
S_t = Underlying Price
S_k = Strike Price
r = Risk free rate
t = Time to maturity

2. The call/put option price should increase with increase in confidence/ volatility.
3. The price of the call option decreases with increase in strike price.
4. The price of a put option increases with decrease in strike price
5. The price of a call option increases with increase in value of the underlying
6. The price of a put option decreases with increase in value of the underlying.
7. The price of a call option increases with increase of interest rate.
8. The price of a put option increases with decrease of interest rate.
9. The price of a call/put option increases with increase in maturity.

The results satisfying the principles 2 to 9 are presented in Figures-2 to 6. The adherence to the principle 1 is shown in Table-4.

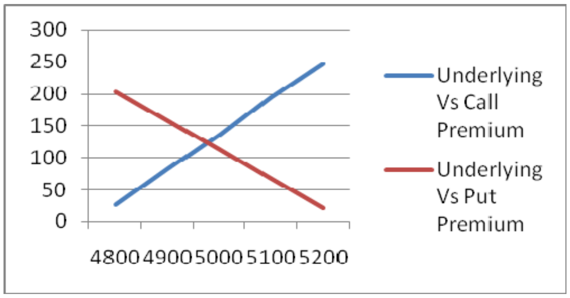

Figure-2. Underlying Vs Premium.

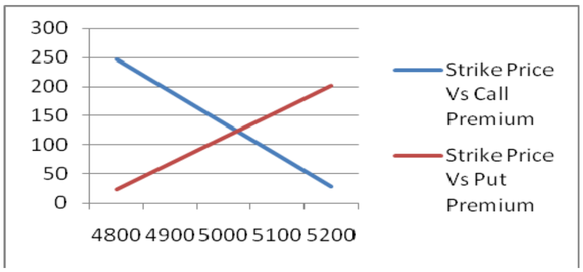

Figure-3. Strike Price Vs Premium.

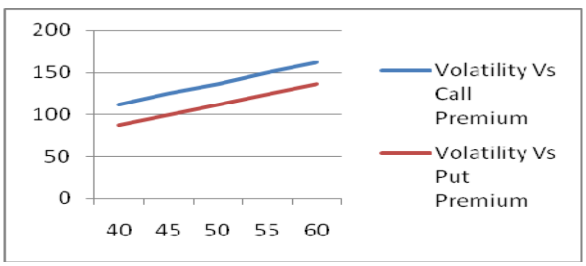

Figure-4. Volatility Vs Premium.

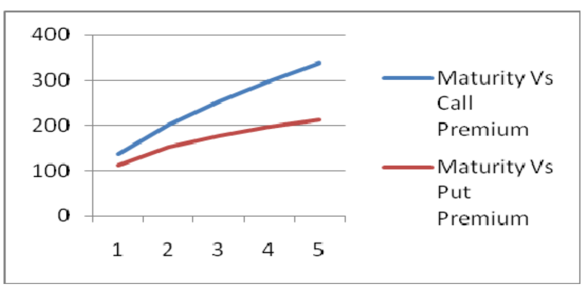

Figure-5. Maturity Vs Premium.

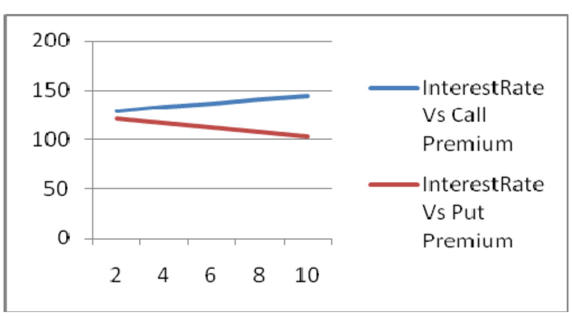

Figure-6. Interest Rate Vs Premium.

Table-4. Put Call Parity for Sk = 5000, r = 6% , t =1month.

Underlying Value	Call Option Premium	Put Option Premium	C + Sk * e-rt	P + St
4800	027.27	202.33	5002.3	5002.3
4900	081.96	157.02	5057.0	5057.0
5000	136.84	111.90	5111.9	5111.9
5100	194.46	066.99	5169.5	5167.0
5200	247.23	022.28	5222.3	5222.3

4. RESULTS AND CONCLUSIONS

The proposed model and the Black-Scholes model are trained using the clustering and curve fitting algorithm to obtain the confidence in case of proposed model and implied volatility in case of Black-Scholes model for corresponding strike prices. The models are cross-validated for strike prices other than the ones used in training the models on the same date. The MSE for the proposed model and Black-Scholes model is calculated using equation (12). It is observed that the MSE with the proposed model is less than the MSE of Black-Scholes model. The MSE between the option prices obtained using the Black-Scholes model and the proposed model is given in Tables-5 and 6.

Table-5. Mean Square Error Vs Model

Cluster	MSE proposed model	MSE Black-Scholes
1	00.09	00.11
2	00.03	00.06
3	35.47	36.45
4	00.02	00.06
5	00.04	00.07

Table-6. Mean Square Error Vs Maturity.

Maturity period in months	MSE proposed model	MSE Black-Scholes
1	35.47	36.45
2	17.35	60.66
3	15.25	77.54

It is observed that the MSE of option prices obtained using the novel proposed model is less than the MSE obtained using Black-Scholes model. Also the proposed model can be used to price even European, American and Bermudian options compared to Black-Scholes model which can price only European options. Unlike the model free approaches, the proposed model satisfies the principles of rational pricing and obeys all the principles of the finance theory and also can be calibrated to the real time data using the proposed clustering algorithm there by making it a tool to calculate the option price for various strikes that match the market prices. The proposed model and the training algorithm are simple to understand and give a decision tree that depicts various paths that the terminal price can take. It is observed that while pricing options and the confidence used in building the binomial decision tree has to be adjusted such that the strike price falls between the maximum and minimum values that the underlying can take, otherwise the price of the option becomes zero as it is assumed that the scenario doesn't occur. Hence as the strike price moves towards zero or infinity, the confidence should be increased so that the strike price falls between the maximum and minimum values that the underlying can take. Hence the confidence rather being constant takes the shape of smile. This explains the reason for volatility smile as the strike price of the options move away from the current price of the underlying.

REFERENCES

[1] Ait-Sahalia Y., Bickel P., and Stoker T. 1998. Goodness-of it tests for regression using kernel methods. Princeton University.

[2] Black F., Scholes M. 1973. The pricing of options and corporate liabilities. Journal of political economy. 81: 637-654.

[3] Bakshi G., Cao C. and Chen Z. 1977. Empirical performance of alternative option-pricing models. The Journal of Finance. Vol.52, No.5, pp.2003-2049.

[4] Bakshi G., Cao C. and Chen Z. 1998. Pricing and hedging long-term options. Journal of Econometrics.

[5] Barucci F., Cherubini U. and Landi L. 1997. Neural networks for contingent claim pricing via the Glarekin method. In Computational Approaches to Economic Problems (H. Amman and B. Rustem, eds.), Kluwer Academic Publishers, Amsterdam. pp. 127-141.

[6] Campbell J., Lo A. and Mackinlay G. 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton. New Jersey, USA.

[7] Cox J.C., Ross S.A. 1976. The Evaluation of Options for alternative processes. Journal of Financial Economics. 3, pp.145-166.

[8] Cox J., Ross S. and Rubinstein M. 1979. Option Pricing: A Simplified Approach". Journal of Financial Economics. 7, No.3, pp.229-263.

[9] Dumas B., Fleming I. and Whaley R. E. 1996. Implied volatility functions: Empirical tests. The Journal of Finance. Vol.53, No.6, pp.2059-2106.

[10] Ghysels F., Patilea V., Renault F. and Torres O. 1997. Nonparametric methods and option-pricing. 97s-19, CIRANO, Montreal.

[11] Gourieroux C., Monfort A. and Tenreiro C. 1994. Kernel M-Estimators: Nonparametric Diagnostics for Structural Models. 9405, CEPREMAP, Paris.

[12] Han J., Kamber M. 2001. Data Mining Concepts and Techniques. Morgan Kaufmann Publishers.

[13] Hull J. 2006. Options Futures and Other Derivatives. 6th Edition, PHI.

[14] Jarrow R. and Rudd A. 1982.Approximate option valuation for arbitrary stochastic processes. Journal of Financial Economics. Vol.10, pp.347-369.

[15] Merton R.C. 1973. Theory of Rational Option Pricing. Bell Journal of Economics and Management Science. Vol.4. Pp. 141-183.

[16] Merton, R. 1976. Option Pricing when Underlying Stock Returns are Discontinuous. Journal of Financial Economics. Vol.3, pp. 125-144.

[17] Osma Abu Abbas. 2008. Comparisons Between Data Clustering Algorithms. The International Arab Journal of Information Technology.Vol.5, No.3.

[18] Rubinstein, M. "Implied binomial trees," The Journal of Finance. Vol.49, pp.771-818, 1994.

[19] Shimko D. 1993. "Bounds of Probability," RISK, Vol.6, No.4, pp.33-37, 1993.

[20] Venkata Naga Malleswara Ramesh K., Venkata Ramana Murthy J., Kodanda Rama Sastry J. 2011. Incorporating Implied Volatility In Pricing Options Using Implied Binomial Tree. 2nd IIMA International Conference on Advanced Data Analysis, Business Analytics and Intelligence.

[21] Venkata Naga Malleswara Ramesh K., Subhash Reddy M. 2007. Estimating Risk and Hedging In Options Trading. 20th Australasian Banking and Finance Conference.