Post-conference Proceedings, Keynote Speech, Numerical Methods for Finance Conference, Dublin, Ireland, 14-15 June 2006

EDDIE for Discovering Arbitrage Opportunities[a]

Edward Tsang[1], Sheri Markose[2], Hakan Er[3] and Alma Garcia[1]

[1] Department of Computer Science and [2] Department of Economics, University of Essex, UK
[3] Department of Business Administration, Akdeniz University, Turkey

Extended Abstract

The prices of the option and futures of a stock both reflect the market's expectation of futures trends of the stock's price. Their prices normally align with each other within a limited window. When they do not, arbitrage opportunities arise: an investor who spots the misalignment will be able to buy (sell) options on one hand, and sell (buy) futures on the other and make risk-free profits. In this paper, we focus on *put-call-futures* parity arbitrage opportunities. The upper bound of a futures bid price, denoted by F_{bt}, is given by:

$$F_{bt}e^{-ra(T-t)} \leq C_{at} - P_{bt} + Xe^{-rb(T-t)} + TC \qquad (1)$$

Here, *T* is the expiration date and *t* is today, i.e. *T–t* is the remaining time to maturity; C_{at} is the option's call premium at the ask, P_{bt} is the option's put premium at the bid. X is the exercise price for the option. *TC* is the transaction cost. *ra* is the interest rate on the borrowing to finance the futures. *rb* is the interest rate to lend. If condition (1) is violated then the arbitrageur will be able to make a risk free profit equal to:

$$F_{bt}e^{-ra(T-t)} - [C_{at} - P_{bt} + Xe^{-rb(T-t)} + TC] > 0 \qquad (2)$$

When condition (1) is violated, a *short arbitrage* profit can be realized by shorting futures and protecting it by a synthetic long futures position by (i) buying a call option, (ii) shorting a futures option, and (iii) borrowing the present discounted value of the futures price and lending the same for the exercise price [9].

Historical data suggest that option and futures prices on the LIFFE Market (London) occasionally do not satisfy condition (1). In the LIFFE tick trade data from January 1991 to June 1998, we identified 8,073 profitable short arbitrage and 7,410 profitable long arbitrage opportunities when no transaction cost is considered. If we assume a transaction cost of £60 per put-call-futures arbitrage operation, then 2,345 (or 29%) of the 8,037 triplets would still be profitable.

The profits in (2) are those that accrue if the arbitrageur could have obtained as quotes the trade prices recorded at these points in time. In reality, due to delay, the arbitrageur may not be able to obtain the quoted prices. Therefore, an arbitrageur may not be able to exploit all the profitable arbitrage opportunities, especially if it reacts passively. Besides, price misalignments are corrected rapidly by the market, so reacting ahead of the others is crucial to securing the risk free profits. Therefore, the challenge is to not only to spot such opportunities, but to discover them ahead of other arbitrageurs. This motivated us to turn our attention to our previous work on forecasting.

EDDIE is a genetic programming tool for forecasting [5][8]. A specialization of EDDIE, which we called EDDIE-ARB, was implemented for forecasting arbitrage opportunities. EDDIE uses constraints to focus its search in promising areas of the space [5]. The task that we gave EDDIE-ARB was to predict arbitrage opportunities five minutes ahead of time.

As a tool, EDDIE enables economists and computer scientists to work together to identify relevant independent variables [8]. The usefulness of EDDIE-ARB as a tool is fully demonstrated in this project: When data was first fed into EDDIE-ARB, no patterns were found. The economists and computer scientists in this project together noticed that certain sub-components were repeatedly generated by the program. In response to that, data was further prepared to help EDDIE-ARB to succeed. For example, "moneyness" (spot price divided by strike price) was introduced, as (a) it is meaningful in economics and (b) this pattern was found by EDDIE-ARB repeatedly. Similarly, "basis" (futures price minus spot price) was introduced to capture mis-pricing in the futures leg of the arbitrage. Scaling was applied to certain variables to avoid the precision problem (which computer scientists are more sensitive to than economists).

---

[a] This paper is based on and extended from [7].

The above preparation alone was not enough to help EDDIE-ARB find patterns reliably. The difficulty of this forecasting problem is that a large percentage of the cases were negative instances. Only about 3% of the instances in the training data represented opportunities. This meant a program that made no positive recommendations (i.e. classifying all cases to be non-profitable) would achieve an accuracy of 97%, even though it had 0% recall. Such forecast would not help us to spot any arbitrage opportunities, and therefore have no commercial value. To tackle this problem, we removed certain negative training instances in order to rebalance the database (we removed those instances which showed no follow-up in the market). When the data set contained around 25% positive instances, EDDIE-ARB started to pick up repeated patterns.

We trained and tested EDDIE on intraday historical tick data on the FTSE-100 European style index option traded on LIFFE from March 1, 1991 to June 18, 1998, and verified it on out-of-sample data from June 30, 1998 to March 30, 1999. The constraints in EDDIE-ARB enabled us to trade precision against recall. For example, the final data set used (i.e. after heavy pre-processing) allowed us to find rules with 99% precision and 53% recall.

Results by EDDIE-ARB were compared with those obtained by a naïve ex ante rule, which only reacted when misalignments were detected. If we assume an operational delay of one minute after opportunities are identified, then expected profit may not be realized by the naïve rule (as explained above). Under this assumption, EDDIE-ARB out-performed the naïve rule on average profit per operation in the test data. However, EDDIE-ARB only picked up a very small percentage of the profitable arbitrage opportunities. As a result, the total amount of profit made by the naïve rule was comparable to EDDIE-ARB's. Our next challenge was therefore to improve EDDIE-ARB's recall rate. Two general methods, namely the Scenario Method and the Repository Method have been developed. Early results suggest that one can collect and combine rules from multiple decision trees to improve precision and recall with these methods [3][4].

This work falls into the research area of *chance discovery*, the discovery of chances through innovation. In chance discovery, the innovation part often involves human input [1][6]. A "chance" here refers to an event or a situation with significant impact on human decision making – a new event/situation that can be conceived either as an opportunity (e.g. in business) or as a risk (such as an earthquake) [2]. Chance discovery extends data-mining, which is often limited to patterns recognition in a given data set. For example, in the EDDIE-ARB project, we identified new attributes which were not present in the original data set. We have in this project established EDDIE-ARB as a promising tool for bringing human users and a computer program together to discover arbitrage opportunities. We have also demonstrated how economists and computer scientists could work together to achieve results that neither party alone was capable of achieving.

References

[1] A. Abe & Y. Ohsawa (eds.), Special Issue on Chance Disocvery. New Generation Computing, 21(1), Berlin: Springer and Tokyo: Ohmsha, 2003

[2] Chance Discovery Consortium, http://www.chancediscovery.com/english/

[3] A.L. Garcia-Almanza and E.P.K. Tsang, Simplifying Decision Trees Learned by Genetic Algorithms, Proceedings, Congress on Evolutionary Computation (CEC) 2006, 7906-7912

[4] A. Garcia & E.P.K. Tsang, The Repository Method for Chance Discovery in Financial Forecasting, Proceedings, 10th International Conference on Knowledge-Based & Intelligent Information & Engineering Systems (KES2006), Bournemouth, UK, 9-11 October 2006

[5] J. Li, FGP: a genetic programming based tool for financial forecasting, PhD Thesis, University of Essex, Colchester, Essex CO4 3SQ, UK, 2001

[6] Y. Ohsawa & P. McBurney, Chance Discovery, Springer Publishers, 2003

[7] E.P.K. Tsang, S. Markose & H. Er, Chance dicsovery in stock index option and future arbitrage, New Mathematics and Natural Computation, World Scientific, Vo.1, No.3, 2005, 435-447

[8] E.P.K.Tsang, P.Yung & J.Li, EDDIE-Automation, a decision support tool for financial forecasting, Journal of Decision Support Systems, Special Issue on Data Mining for Financial Decision Making, Vol.37, No.4, 2004, 559-565

[9] A.L. Tucker, Financial Futures, Options and Swaps, West Publishing Company, St. Paul, MN., 2001