Journal of

Applied Mathematics and Decision Sciences

Special Issue
Intelligent Computational Methods for Financial Engineering

Hindawi Publishing Corporation
http://www.hindawi.com

Volume 2009

Intelligent Computational Methods for Financial Engineering

Journal of
Applied Mathematics and Decision Sciences

Intelligent Computational Methods for Financial Engineering

Guest Editors: Lean Yu, Shouyang Wang, and K. K. Lai

Contents

Hindawi Publishing Corporation
Journal of Applied Mathematics and Decision Sciences
Volume 2009, Article ID 394731, 2 pages
doi:10.1155/2009/394731

Editorial

Intelligent Computational Methods for Financial Engineering

Lean Yu,[1] Shouyang Wang,[1] and K. K. Lai[2]

[1] *Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China*
[2] *Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong*

Correspondence should be addressed to Lean Yu, yulean@amss.ac.cn

Received 17 June 2009; Accepted 17 June 2009

As a multidisciplinary field, financial engineering is becoming increasingly important in today's economic and financial world, especially in areas such as portfolio management, asset valuation and prediction, fraud detection, and credit risk management. For example, in a credit risk context, the recently approved Basel II guidelines advise financial institutions to build comprehensible credit risk models in order to optimize their capital allocation policy. Computational methods are being intensively studied and applied to improve the quality of the financial decisions that need to be made. Until now, computational methods and models are central to the analysis of economic and financial decisions.

However, more and more researchers have found that the financial environment is not ruled by mathematical distributions or statistical models. In such situations, some attempts have also been made to develop financial engineering models using some emerging intelligent computing approaches. For example, an artificial neural network (ANN) is a nonparametric estimation technique which does not make any distributional assumptions regarding the underlying asset. Instead, ANN approach develops a model using sets of unknown parameters and lets the optimization routine seek the best fitting parameters to obtain the desired results. That is, these emerging intelligent solutions can efficiently improve the decisions for financial engineering problems. Having agreed on this basic fact, the guest editors determined that the main purpose of this special issue is not to merely illustrate the superior performance of a new intelligent computational method, but also to demonstrate how it can be used effectively in a financial engineering environment to improve and facilitate financial decision making.

For this purpose, this special issue presents some new progress in intelligent computational methods for financial engineering. In particular, the special issue addresses

how the emerging intelligent computational methods (e.g., ANN, support vector machines, evolutionary algorithm, and fuzzy models, etc.) can be used to develop intelligent, easy-to-use and/or comprehensible computational systems (e.g., decision support systems, agent-based system, and web-based systems, etc.), which is expected to trigger some thoughts and deepen further research.

In this special issue, 12 papers were selected from 26 submissions related to intelligent computational methods for financial engineering from different countries and regions. The authors of the selected papers are from USA, Canada, Australia, Japan, Finland, Tunisia, Sri Lanka, Taiwan, South Africa, and China, respectively. In addition, all the selected papers went through a standard peer review process of the journal and the authors of some papers made necessary revision in terms of reviewing comments. The selected papers include "Optimal Bespoke CDO Design via NSGA-II" by Diresh Jewan, Renkuan Guo, and Gareth Witten, "Modified neural network algorithms for predicting trading signals of stock market indices" by C. D. Tilakaratne, M. A. Mammadov, S. A. Morris, "Selecting the best forecasting implied volatility model using genetic programming" by Wafa Abdelmalek, Sana Ben Hamida, Fathi Abid, "Discrete analysis of portfolio selection with optimal stopping time" by Jianfeng Liang, "A New decision-making method for stock portfolio selection based on computing with linguistic assessment" by Chen Tung Chen, Wei Zhan Hung, "A fuzzy pay-off method for real option valuation" by Mikael Collan, Robert Fuller, Jozsef Mezei, "Valuation for an American continuous-installment put option on bond under vasicek interest rate model" by Guohe Deng, Lihong Huang, "Callable Russian options and their optimal boundaries" by Atsuo Suzuki, Katsushige Sawaki, "Valuation of game options in jump diffusion model and with applications to convertible bonds" by Lei Wang and Zhiming Jin, "Fuzzy real options in brownfield redevelopment evaluation" by Qian Wang, Keith W. Hipel, D. Marc Kilgour, "Discriminant analysis of zero recovery for China's NPL" by Yue Tang, Xiaoguang Yang, Hao Chen, Bo Wang, Muzi Chen, Min Chen, and "Cumulative gains model quality metric" by Thomas Brandenburger, Alfred Furth. The guest editors hope that the papers published in this special issue would be of value to academic researchers and business practitioners and would provide a clearer sense of direction for further research, as well as facilitating use of existing methodologies in a more productive manner.

The guest editors would like to place on record their sincere thanks to Professor Mahyar A. Amouzegar, the Editor-in-Chief of Journal of Applied Mathematics and Decision Sciences, for this very special opportunity provided to us for contributing to this special issue. The guest editors have to thank all the referees for their kind support and help, which has guaranteed that this special issue is of high standard. Finally, the guest editors would like to thank the authors of all the submissions to this special issue for their contribution. Without the support of the authors and the referees, it would have been impossible to make this special issue for our readers. It is hoped that readers can find some topics of interest and benefit to them. The guest editors also hope that this special issue would inspire researchers in the fields of intelligent financial engineering to explore more creative contributions in their research fields.

Lean Yu
Shouyang Wang
K. K. Lai

Hindawi Publishing Corporation
Journal of Applied Mathematics and Decision Sciences
Volume 2009, Article ID 925169, 32 pages
doi:10.1155/2009/925169

Research Article

Optimal Bespoke CDO Design via NSGA-II

Diresh Jewan,[1] Renkuan Guo,[1] and Gareth Witten[2]

[1] *Department of Statistical Sciences, University of Cape Town, Private Bag, Rhodes' Gift, Rondebosch 7701, Cape Town, South Africa*
[2] *Peregrine Quant, PO Box 44586, Claremont, Cape Town, 7735, South Africa*

Correspondence should be addressed to Renkuan Guo, renkuan.guo@uct.ac.za

Received 28 November 2008; Accepted 9 January 2009

Recommended by Lean Yu

This research work investigates the theoretical foundations and computational aspects of constructing optimal bespoke CDO structures. Due to the evolutionary nature of the CDO design process, stochastic search methods that mimic the metaphor of natural biological evolution are applied. For efficient searching the optimal solution, the nondominating sort genetic algorithm (NSGA-II) is used, which places emphasis on moving towards the true Paretooptimal region. This is an essential part of real-world credit structuring problems. The algorithm further demonstrates attractive constraint handling features among others, which is suitable for successfully solving the constrained portfolio optimisation problem. Numerical analysis is conducted on a bespoke CDO collateral portfolio constructed from constituents of the iTraxx Europe IG S5 CDS index. For comparative purposes, the default dependence structure is modelled via Gaussian and Clayton copula assumptions. This research concludes that CDO tranche returns at all levels of risk under the Clayton copula assumption performed better than the sub-optimal Gaussian assumption. It is evident that our research has provided meaningful guidance to CDO traders, for seeking significant improvement of returns over standardised CDOs tranches of similar rating.

1. Introduction

Bespoke CDOs provides tailored credit solutions to market participants. They provide both long-term strategic and tactical investors with the ability to capitalise on views at the market, sector and name levels. Investors can use these structures in various investment strategies to target the risk/return profile or hedging needs. These strategies can vary from leverage and correlation strategies to macro and relative value plays [1].

Understanding the risk/return trade-off dynamics underlying the bespoke CDO collateral portfolios is crucial when maximising the utility provided by these instruments. The single-tranche deal can be put together in a relatively short period of time. This is aided by the development of numerous advance pricing, risk management and portfolio optimisation techniques.

The most crucial tasks in putting together the bespoke CDO is choosing the underlying credits that will be included in the portfolio. Investors often express preferences on individual names, and there is likely to be credit rating constraint and industry concentration limits imposed by the investors and rating agencies [2].

Given these various investor defined requirements, the structurer is required to optimise the portfolio to achieve the best possible tranche spreads for investors. This was a complicated task, however, with the advent of faster computational pricing and portfolio optimisation algorithms, aid structurers in presenting bespoke CDO which conform to the investment parameters.

The proper implementation of the decision steps lies in the solution of the multiobjective, multiconstrained optimisation problem, where investors can choose an optimal structure that matches their risk/return profile. Optimal structures are defined by portfolios that lie on the Pareto frontier on the CDO tranche yield/portfolio risk plane.

Davidson [2] provides an interesting analogy between CDO portfolio optimisation processes and evolutionary cycles espoused by Charles Darwin. In the natural world, life adapts to suit the particulars of its environment. To adapt to a specific environment, a simple but extraordinarily powerful set of evolutionary techniques are employed—reproduction, mutation and survival of the fittest. In this way, nature explores the full range of possible structures to hone in on those that are most perfectly suited to the surrounding environment.

The creation of the CDO collateral portfolio can broadly be seen in similar ways. Given a certain set of investor and/or market constraints, such as the number of underlying credits, the notional for the credits, concentration limits and the weighted average rating factor, credit structurers need to be able to construct a portfolio that is best suited to the market environment. If the portfolio does not suit the conditions, it evolves so that only those that are "fittest," defined by having the best CDO tranche spread given the constraints, will survive. Many of the same techniques used in the natural world can be applied to this constrained portfolio optimisation problem. Evolutionary algorithms have received a lot of attention regarding their potential for solving these types of problems. They possess several characteristics that are desirable to solve real world optimisation problems up to a required level of satisfaction.

Our previous research work focused on developing a methodology to optimise credit portfolios. The Copula Marginal Expected Tail Loss (CMETL) model proposed by Jewan et al. [3], is one that minimises credit portfolio ETL subject to a constraint of achieving expected portfolio returns at least as large as an investor defined level, along with other typical constraints on weights, where both quantities are evaluated in the CMETL framework. Jewan et al. [3] have shown that ETL optimal portfolio techniques, combined with copula marginal (factor copula) distribution modelling of the portfolio risk factors can lead to significant improvements in risk-adjusted returns.

Our research work now investigates a new approach to asset allocation in credit portfolios for the determination of optimal investments in bespoke CDO tranches. Due to the complexity of the problem, advance algorithms are applied solve the constrained multiobjective optimisation problem. The nondominating sort genetic algorithm (NSGA-II) proposed by Deb et al. [4] is applied.

NSGA-II is a popular second generation multiobjective evolutionary algorithm. This algorithm places emphasis on moving towards the true Pareto-optimal region, which is essential in real world credit structuring problems. The main features of these algorithms are the implementation of a fast nondominated sorting procedure and its ability to handle

constraints without the use of penalty functions. The latter feature is essential for solving the multiobjective CDO optimisation problem.

The study uses both Gaussian and Clayton copula models to investigate the effects of different default dependence assumptions on the Pareto frontier. Various real world cases are considered, these include the constrained long-only credits and concentrated credit cases. Two objectives are used to define the CDO optimisation problem. The first is related to the portfolio risk, which is measured by the Expected-tail-loss (ETL). ETL is a convex risk measure and has attractive properties for asset allocation problems. The second objective is the CDO tranche return. This objective requires a CDO valuation model. We apply an extension of the Implied Factor model proposed by Rosen and Saunders [5]. The extension is a result of the application of the Clayton copula assumption. Rosen and Saunders [5] restricts their analysis to the Gaussian case, but use a multifactor framework.

The breakdown of the paper is as follows. The next section briefly discusses the mechanics of bespoke CDOs. It outlines the three improtant decision making steps involved in the structuring process. In section four, a robust and practical CDO valuation framework based on the application of the single-factor copula models given in Section 3, is presented. This is in conjunction with weighted Monte Carlo techniques used in options pricing. The results of the study on the impact of the different copula assumption on the market implied loss distribution is then presented. Finally the analysis on the implied credit tail characteristics under the various copula assumptions is given. Section 5 defines convex credit risk measures in a self contained manner. Sections 4 and 5 establish the theory behind the objective functions used in the CDO optimisation model. In Section 6 the generic model for multiobjective bespoke CDO optimisation is presented. The components of the NSGA-II are discussed and the algorithm outlined. This then paves the way to perform a prototype experiment on a bespoke CDO portfolio constructed from constituents of the iTraxx Europe IG S5 index. The final section highlights the improtant research findings and discusses several areas of future study.

2. Bespoke CDO Mechanics

A bespoke CDO is a popular second-generation credit product. This standalone single-tranche transaction is referred to as a bespoke because it allows the investor to customise the various deal characteristics such as the collateral composition, level of subordination, tranche thickness, and credit rating. Other features, such as substitution rights, may also play an improtant role [1]. In these transactions, only a specific protion of the credit risk is transferred, unlike the entire capital structure as in the case of standardised synthetic CDOs. Most of these transactions involve 100–200 liquid corporate CDS.

While the bespoke CDO provides great flexibility in the transaction parameters, it is crucial that investors understand the mechanics of the deal. A key feature of these transactions is the greater dialogue that exists between the parties during the structuring process, avoiding the "moral hazard" problem that existed in earlier CDO deals [6].

In a typical bespoke CDO transaction, there are three main decision steps for potential investors:

(1) *Credit Selection for the Reference portfolio*: The first step in structuring a bespoke CDO is the selection of the credits for the collateral portfolio. Investors can choose a portfolio of credits different from their current positions. They can also sell

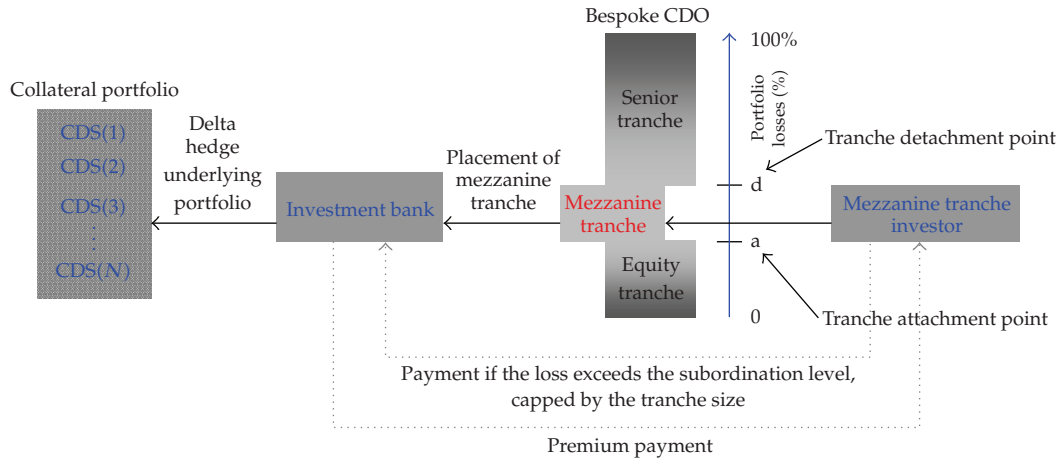


Figure 1: Placement of mezzanine tranche in a typical bespoke CDO transaction.

protection on the subset of names in their current portfolio in case they have overweight views on particular credit and/or sectors.

(2) *Defining the Subordination Level and Tranche Size*: Once a credit portfolio has been selected, investors must choose a subordination level and tranche size. These variables determine the degree of leverage and the required protection premium. Investors that are primarily concerned about the credit rating of the CDO tranche, rated by major rating agencies such as Moody's, S&P and Fitch, could choose the tranche size and subordination level so as to maximise the premium for a selected rating [1]. Usually investors choose a subordination level that will provide a total spread equal to some desired return target. This "tranching" of credit portfolio risk can provide any desired risk/return profile. By choosing the position of a tranche on a capital structure, investors can decouple their views on default risk from their views on market risk. (See Rajan et al. [1, page 203–220] for trading strategies involving bespoke CDO transactions.)

(3) *Substitution of Credits in the Reference portfolio*: The third distinctive feature of bespoke CDOs is the investors' ability to dynamically manage their investment profit and loss by substituting credits in the portfolio.

An optimal transaction would be based on a collateral portfolio that lies on the efficient frontier on the tranche return/risk plan, solution of a multiobjective optimisation problem that satisfies both trade and regulatory constraints.

A typical placement of a bespoke CDO is outlined in Figure 1.

In the above schematic, the investor goes long the credit risk in a mezzanine tranche. The challenge of issuing bespoke CDOs is the ongoing need and expense of the risk management of the tranche position. The common method of hedging these transactions is to manage the risk like an options book. Greeks similar to those related to options can be defined for CDO tranches. The distribution of risk in these transactions is not perfect, leaving dealers exposed to various first and second order risks [7].

3. Factor Copula Models for Default Dependence Modelling

Copula-based credit risk models were developed to extend the univariate credit-risk models for the individual obligors to a multivariate setting which keeps all salient features of the individual credit-risk models while incorporating a realistic dependency structure between the obligor defaults.

Copulas were first introduced by Sklar [8], in the context of probabilistic metric spaces, however their applications to finance have been very recent. The idea was first invoked within finance by Embrechths et al. [9], in connection with the inadequacies of linear correlation as a measure of dependence or association.

In credit portfolio modelling the copula approach and factor models have become an industry-wide standard to describe the asset correlation structure. Construction of mutual asset correlations between entities of the portfolio on common external factors represents a very flexible and powerful modelling tool. This modelling approach can be understood as a combination of the copula and firm-value approach. This factor approach is quite standard in credit risk modelling [10–14]. These models have been widely applied to credit derivatives modelling for essentially two reasons:

(i) factor models represent an intuitive framework and allow fast calculation of the loss distribution function; and

(ii) the full correlation matrix, which represents a challenging issue in large credit portfolios, need not be fully estimated.

Factor models can then be used to describe the dependency structure amongst credits using a so called "credit-versus-common factors" analysis rather than a pairwise analysis.

Credit risk models can be divided into two mainstreams, structural models and reduced form models. In the reduced-form methodology, the default event in these models is treated exogenously. The central idea is to model the default counting process N_c. This is a stochastic process which assumes only integer values. It literally counts default events, with $N_c(t)$ denoting the number of events that have occurred up to time t. In the case of the mth-to-default, the default time is given by the following:

$$\tau_m = \min\left\{t \in [0,T] \mid N_c(t) = m\right\}. \tag{3.1}$$

Using the standard settings for financial product pricing, we fix a complete probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ where the model lives. Let $\mathbb{Q}$ denote the risk neutral probability measure. All subsequently introduced filtrations are subsets of the filtration $\mathcal{F}$ and augmented by the zero-sets of $\mathcal{F}$ [15].

We will consider a latent factor V such that conditionally on V, the default times are independent. The factor approach makes it simple to deal with a large number of names and leads to very tractable results. We will denote by $p_t^{i|V} = \mathbb{Q}(\tau^{(i)} \le t \mid \widehat{\mathcal{H}_t} \vee \sigma(V_t))$, the conditional default probability of the name i, and by $q_t^{i|V} = \mathbb{Q}(\tau^{(i)} > t \mid \widehat{\mathcal{H}_t} \vee \sigma(V_t))$, the corresponding conditional survival probability. Conditionally on V, the joint survival function is given by the following:

$$\mathbb{Q}\left(\tau^{(1)} \ge t, \tau^{(2)} \ge t, \ldots, \tau^{(m)} \ge t \mid \widehat{\mathcal{H}_t} \vee \sigma(V_t)\right) = \prod_{i=1}^{n} \mathbb{E}\left[N_c^{(i)}(t) \mid \widehat{\mathcal{H}_t} \vee \sigma(V_t)\right] \tag{3.2}$$

for m reference entities. The filtration, $\mathscr{H}_t$ represents all available information in the economy at time t. The filtration $\mathscr{H}_t^i$ represents all available information restricted to obligor i and the background process. These filtrations enable us to model the intensity of the default process of an obligor independent of the information about the remaining obligors.

The next two subsections provide a description of the two copula models used in the study. These two models will define the default dependence structure, which is used in a Monte Carlo simulation, to derive the credit portfolio loss distribution. These distributions are then used in pricing, portfolio and risk management.

3.1. Gaussian Copula

A convenient way to take into account the default dependence structure is through the Gaussian copula model. This has become the market standard in pricing multiname credit products. In the firm-value approach a company will default when its "default-like" stochastic process, X, falls below a barrier. Define the stochastic process,

$$X_i = \rho_i V + \sqrt{1-\rho_i^2}\epsilon_i, \tag{3.3}$$

where V and ϵ_i are independent standard Gaussian variants, with $\text{Covar}(\epsilon_i, \epsilon_j) \neq 0$ for all $i \neq j$. When $\rho = 0$, this corresponds to independent default times while $\rho = 1$ is associated with the comonotonic case. X_i can be interpreted as the value of the assets of the company, and V, the general state of the economy. The default dependences come from the factor V. Unconditionally, the stochastic processes are correlated but conditionally they are independent.

The default probability of an entity i denoted by F_i can be observed from market prices of credit default swaps. Under the copula model, each X_i is mapped to $\tau^{(i)}$ using a percentile-to-percentile transformation. In a Gaussian copula model the point $X_i = x$ is transformed into $\tau^{(i)} = t$ with $\tau^{(i)} = F^{-1}(\Phi(x))$. It follows from (3.3) that the conditional default probability is given by:

$$\mathbb{Q}\left(X_i \leq x \mid V\right) = \Phi\left(\frac{x - \rho_i V}{\sqrt{1-\rho_i^2}}\right), \tag{3.4}$$

where $x = \Phi^{-1}(F_i(t))$, and $\mathbb{Q}(X \leq x) = \mathbb{Q}(\tau \leq t)$,

$$\mathbb{Q}\left(\tau^{(i)} \leq t \mid V\right) = \Phi\left(\frac{\Phi^{-1}(F_i(t)) - \rho_i V)}{\sqrt{1-\rho_i^2}}\right). \tag{3.5}$$

We also have that:

$$\tau^{(i)} = F^{-1}\left(\Phi\left(\rho_i V + \sqrt{1-\rho_i^2}\epsilon_i\right)\right), \tag{3.6}$$

this is used when we want to simulate the default times. The Gaussian copula has no upper or lower tail dependence. Embrechths et al. [9] proves this result.

3.2. Clayton Coupla

Most of the existing copula models involve using the Gaussian copula which has symmetric upper and lower tails, without any tail dependence. This symmetry fails to capture the fact that firm failures occur in cascade in difficult times but not in better times. That is, the correlation between defaults increases in difficult times. The Clayton copula encapsulates this idea.

This "class" of Archimedean copulas was first introduced by Clayton [16] from his studies on the epidemiological chronic diseases. Clayton [16] only developed a general form (without imposing any parametric constraints) while Oakes [17] refined the copula in term of its parameterisation. Friend and Rogge [18], Laurent and Gregory [19], Schloegl and O'Kane [20], Schönbucher and Schubert [15], and Schönbucher [21] have been considering this model in a credit risk context, primarily due to the lower tail dependence which is ideal for credit risk applications.

The asset value process in a single factor Clayton copula is given by the following:

$$X_i = \left(1 - \frac{\log(\epsilon)}{V}\right)^{-1/\theta_C}, \tag{3.7}$$

where V is the systematic risk factor, a positive random variable, following a standard Gamma distribution with shape parameter $1/\theta_C$ with $\theta_C > 0$, and scale parameter equal to unity.

Using the definition of default times and the Marshall and Olkin [22] sampling algorithm, the conditional survival probability given by the following:

$$q_t^{i|V} = \exp\left(V\left(1 - F_i(t)^{-\theta_C}\right)\right). \tag{3.8}$$

The default times are given by,

$$\tau^{(i)} = F^{-1}\left(\left(1 - \frac{\log(\epsilon)}{V}\right)^{-1/\theta_C}\right). \tag{3.9}$$

From expression (3.9), it is clear that the stochastic intensities are proprotional to V. Thus the latent variable acts as a multiplicative effect on stochastic intensities. High levels of the latent variable are associated with shorter survival default times. For this reason, V is called a "frailty."

When $\theta_C = 0$, we obtain the product copula and this implies the default times are independent. When $\theta_C \to \infty$, the Clayton copula turns out to be the upper Frëchet bound, corresponding to the case where default times are comonotonic [19]. As the parameter θ_C increases, the Clayton copula increases with respect to the supermodular order. This implies an increasing dependence in default times, and hence has some direct consequences for the pricing of CDO tranches as shown by Laurent and Gregory [19].

4. Implied Factor Copula Models for CDO Valuation

The current market standard for pricing synthetic CDOs is the single-factor Gaussian copula model introduced by Li [23]. Numerous research has shown that a single parameter model is unable to match the price of all quoted standardised CDO tranches [24–32].

It is common practice to quote an implied "correlation skew"—a different correlation which matches the price of each tranche. This assumption is analogous to the Black-Scholes implied volatility in the options market. Implied tranche correlations not only suffer from interpretation problems, but they might not be unique as in the case of mezzanine tranches, and cannot be interpolated to price bespoke tranches [33].

We define bespoke tranches in the following cases:

(i) the underlying portfolio and maturity is the same as the reference portfolio, but the tranche attachment and/or detachment points are different,

(ii) the underlying portfolio is the same as the reference portfolio but the maturities are different, or

(iii) the underlying portfolio differs from the reference portfolio.

The following subsection presents a robust and practical CDO valuation framework based on the application of the single-factor copula models presented in the previous section. The method to recover the credit loss distributions from the factor copula structure is then presented. The implied factor model is then derived. This development is in conjunction with weighted Monte Carlo techniques used in options pricing. The Gaussian model presented here is a special case of the multifactor Gaussian copula model proposed by Rosen and Saunders [5]. The application of the Clayton copula model is an extension of Rosen and Saunders [5] work.

The impact of the different copula assumptions on the loss distribution is also investigated. The credit tail characteristics are analysed. This is imperative to ensure that the copula model used has the ability to capture the default dependence between the underlying credits, and not severely underestimate the potential for extreme losses. The loss analysis is performed on a homogeneous portfolio consisting of 106 constituents of the iTraxx European IG S5 CDS index.

4.1. Synthetic CDO Pricing

The key idea behind CDOs, is the tranching of the credit risk of the underlying portfolio. A given tranche n_{tr} is defined by its attachment and detachment points $u_{n_{\text{tr}}}^{\text{lower}}$ and $u_{n_{\text{tr}}}^{\text{upper}}$ respectively. The tranche notional is given by: $s_{n_{\text{tr}}} = N_{\text{prot}}(u_{n_{\text{tr}}}^{\text{upper}} - u_{n_{\text{tr}}}^{\text{lower}})$, where N_{prot} denotes the total portfolio value.

Let $L_{\text{total}}(t)$ be the percentage cumulative loss in the portfolio value at time t. The total cumulative loss at time t is then $L_{\text{total}}(t)N_{\text{prot}}$. The loss suffered by the holders of tranche n_{tr} from origination to time t is a percentage $L_{n_{\text{tr}}}(t)$ of the portfolio notional value N_{prot}:

$$L_{n_{\text{tr}}}(t) = \min\left\{ \max\left\{L_{\text{total}}(t) - L_{n_{\text{tr}}}, 0\right\}, u_{n_{\text{tr}}}^{\text{upper}} - u_{n_{\text{tr}}}^{\text{lower}}\right\}. \tag{4.1}$$

We consider a transaction initiated at time 0, with maturity T. In a CDO contract, the tranche investor adopts a position as a protection seller. Similar to the assumption under the CDS

pricing, we assume that defaults occur at the midpoints between coupon payment dates. The value of the protection leg is then given by the following:

$$\mathbb{E}\left[\mathrm{PV}_{\mathrm{prot}}^{(n_{\mathrm{tr}})}\right] = \sum_{j=1}^{J} D\left(0, \frac{t_j - t_{j-1}}{2}\right)\left(\mathbb{E}\left[L_{n_{\mathrm{tr}}}(t_j)\right] - \mathbb{E}\left[L_{n_{\mathrm{tr}}}(t_{j-1})\right]\right), \tag{4.2}$$

where $\mathbb{E}[L_{n_{\mathrm{tr}}}(\,\cdot\,)]$ is the expectation with respect to the risk neutral measure $\mathbb{Q}$ and is calculated by a simulation model. The tranche loss profiles under each scenario are calculated and stored. The expectation is found by calculating the weighted average over all scenarios. This only applies if a weighted Monte Carlo scheme is used.

The tranche investors need to be compensated for bearing the default risk in the underlying credits. The holders of tranche n receive a periodic coupon payment. Let the coupon payment dates be denoted as $0 = t_0 \leq t_1 \leq t_2 \cdots \leq t_J = T$. The predetermined frequency of the coupon payment dates is usually on a quarterly basis. The spread paid for protection on a given tranche does not vary during the life of the contract. and is usually quoted in basis points per annum. However, the tranche notional decays through the life of the contract. The outstanding tranche notional at time t is given by the following:

$$N_{\mathrm{out}}^{n_{\mathrm{tr}}}(t) = \left(u_{n_{\mathrm{tr}}}^{\mathrm{upper}} - u_{n_{\mathrm{tr}}}^{\mathrm{lower}} - \mathbb{E}\left[L_{n_{\mathrm{tr}}}(t)\right]\right)N_{\mathrm{prot}}. \tag{4.3}$$

The expected outstanding tranche notional since the last coupon date must be considered at the coupon payment dates. This amount between coupon payment dates t_j and t_{j-1} is simply the average of $N_{\mathrm{out}}^{n_{\mathrm{tr}}}(t_j)$ and $N_{\mathrm{out}}^{n_{\mathrm{tr}}}(t_{j-1})$. We assume again that defaults can only occur at the midpoint between arbitrary coupon dates. The expected outstanding tranche notional is denoted by:

$$\mathbb{E}\left[N_{\mathrm{out}}^{n_{\mathrm{tr}}}(t_j, t_{j-1})\right] = \left(u_{n_{\mathrm{tr}}}^{\mathrm{upper}} - u_{n_{\mathrm{tr}}}^{\mathrm{lower}} - \mathbb{E}\left[L_{n_{\mathrm{tr}}}(t_j)\right] + \frac{\mathbb{E}[L_{n_{\mathrm{tr}}}(t_j)] - \mathbb{E}[L_{n_{\mathrm{tr}}}(t_{j-1})]}{2}\right). \tag{4.4}$$

Using this equation we can compute the expected present value of the coupon payments:

$$\mathbb{E}\left[\mathrm{PV}_{\mathrm{prem}}^{n_{\mathrm{tr}}}\right] = \sum_{j=1}^{J} s_{n_{\mathrm{tr}}}(t_j - t_{j-1})D(0, t_j)\mathbb{E}\left[N_{\mathrm{out}}^{n_{\mathrm{tr}}}(t_j, t_{j-1})\right]. \tag{4.5}$$

This fair spread of a tranche can be computed by equating the expected present value of the protection and premium legs. The market quotation for the equity tranche is to have a fixed 500 bps spread, and the protection buyer make an upfront payment of a fixed percentage of the tranche notional. The protection seller receives the upfront fee expressed as a percentage f of the tranche notional, so that equity investors purchase the note at a discount $f(u_{\mathrm{eq}})N_{\mathrm{prot}}$. Only the premium leg is different for equity tranche investors. This is given by the following:

$$\mathbb{E}\left[\mathrm{PV}_{\mathrm{prem}}^{(\mathrm{eq})}\right] = f(u_{\mathrm{eq}})N_{\mathrm{prot}} + \sum_{j=1}^{J} s_{\mathrm{eq}}(t_j - t_{j-1})D(0, t_j)\mathbb{E}\left[N_{\mathrm{out}}^{(\mathrm{eq})}(t_j, t_{j-1})\right]. \tag{4.6}$$

4.2. Weighted Monte Carlo Techniques

Monte Carlo algorithms can be divided (somewhat arbitrarily) into two categories: uniformly weighted and nonuniformly weighted algorithms. nonuniform weights are a mechanism for improving simulation accuracy. Consider a set of M paths, generated by a simulation procedure. A nonuniformly weighted simulation is one in which the probabilities are not necessarily equal. Suppose that we assign, respectively, probabilities $p_1, p_2, \ldots, p_M$, to the different paths. The value of the security according to the nonuniform weights is,

$$\Omega_h = \sum_{m=1}^{M} p_m \Lambda_m, \tag{4.7}$$

where Λ is the payoff function.

Two features of credit risk modelling which pose a particular challenge under simulation based procedures, namely:

(1) it requires accurate estimation of low-probability events of large credit losses; and

(2) default dependence mechanisms described in the previous chapter do not immediately lend themselves to rare-event simulation techniques used in other settings.

It is for these reasons that we implement a weighted Monte Carlo simulation procedure to put emphasis on determining directly the risk neutral probabilities of the future states of the market, which will allow the accurate determination of the credit loss distribution. This is in contrast to calibration of pricing models through traditional methods.

In what follows, we introduce the general modelling framework, present the algorithm and discuss some technical implementation details. (We use a similar methodology to Rosen and Saunders [5].) This generalised framework can be applied to credit risk models that have a factor structure.

A weighted Monte Carlo method can be used to find an implied risk-neutral distribution of the systematic factors, assuming the specification of a credit risk model with a specified set of parameter values. According to Rosen and Saunders [5] the justification for this approach follows working within a conditional independence framework, where obligor defaults are independent, conditional on a given systematic risk factor.

The methodology to obtain the implied credit loss distribution is summarised by the following steps.

(i) *Latent factor scenarios*. Define M scenarios on the systematic factor V.

(ii) *Conditional portfolio loss distribution*. Calculate the conditional portfolio loss profile for scenario m.

(iii) *Conditional tranche values*: Infer the tranche loss distributions conditional on scenario m.

(iv) *Implied scenario probabilities (optimisation problem)*. Given a set of scenario probabilities p_m; tranche values are given as expectations over all scenarios of the conditional values. We solve the resulting constrained inverse problem to find a set of implied scenario probabilities p_m.

(v) *Implied credit loss distribution*-Calculate the aggregate credit loss at each time step given the implied scenario probabilities.

The first three steps above have been discussed in detail in the previous sections of this chapter. The focus now is placed on the optimisation problem for implying the scenario probabilities.

4.2.1. Mathematical Formulation

Let $h : \mathbb{R}^M \to \mathbb{R} \cup \{+\infty\}$ be a strictly convex function and let $w_{i,M}$ denote the weight of the ith path of M paths. Now consider the optimisation problem:

$$\min_{w_{1,M}, w_{2,M}, \ldots, w_{M,M}} \sum_{i=1}^{M} h(w_{i,M}) \quad \text{subject to}$$

$$\frac{1}{M} \sum_{i=1}^{M} w_{i,M} = 1, \tag{4.8}$$

$$\frac{1}{M} \sum_{i=1}^{M} w_{i,M} \mathbf{G}_i = c_{\mathbf{G}},$$

for some fixed $c_{\mathbf{G}} \in \mathbb{R}^N$. The objective function is strictly convex and the constraints are linear, so if a feasible solution exist, with a finite objective function value, then there is a unique optimal solution $w_{1,M}, w_{2,M}, \ldots, w_{M,M}$. This optimum defines the weighted Monte Carlo estimator,

$$\widehat{\Psi}_{\mathrm{LCV}} = \sum_{i=1}^{M} w_{i,M} \Psi_i. \tag{4.9}$$

The weights derived from (4.8) can be made more explicit by introducing the Lagragian. The strict convexity of the objective function is not a sufficient condition to guarantee a unique solution.

The classical approach to solving constrained optimisation problems is the method of Lagrange multipliers. This approach transforms the constrained optimisation problem into an unconstrained one, thereby allowing the use of the unconstrained optimisation techniques.

4.2.2. Objective Functions

Taking the objective to be a symmetric separable convex function gives the optimal probabilities p. This is an interpreted as the most uniform probabilities satisfying the constraints, though different objective functions imply different measures of uniformity. A common choice for the fitness measure is *entropy*. This is a particularity interesting and in some respects a convenient objective. The principle of maximum entropy gives a method of generating a probability distribution from a limited amount of information. It is a relatively well used principle for the construction of probabilistic models.

In this setting,

$$h = \sum_{w^*=1}^{M} p_{w^*} \log p_{w^*}. \tag{4.10}$$

This case is put forward in Avellaneda et al. [34], where a Bayesian interpretation is given. The usual convention of $0 \log 0$ is followed. We now provide a problem specific description of the optimisation problem using the principle of maximum entropy.

4.2.3. Constraint Description

For the bespoke CDO pricing problem, the implied scenario probabilities must satisfy the following constraints:

(i) the sum of all scenario probabilities must equal to one;

(ii) the probabilities should be positive: $p_i \geq 0$ for all $i \in \{0, 1, \ldots, M\}$; and bounded by 1;

(iii) we match the individual CDS spreads (i.e., marginal default probabilities for each name)

$$\sum_{w^*=1}^{M} p_{w^*} p_{z^*}^{t|V_{w^*}} = p_{z^*}^{t|V} \quad \forall z^* \in \{1, 2, \ldots, N\}; \tag{4.11}$$

(iv) the current market prices of standard CDO tranches are matched given by

$$\sum_{w^*=1}^{M} p_{w^*} \mathrm{PV}_{\mathrm{prot}}^{(n_{\mathrm{tr}})} = \sum_{w^*=1}^{M} p_{w^*} \mathrm{PV}_{\mathrm{prem}}^{(n_{\mathrm{tr}})} \quad \forall n_{\mathrm{tr}} \in \{1, 2, \ldots\} \tag{4.12}$$

tranches.

Because the number of controls in the problem is typically smaller than the number of replications, these constraints do not determine the probabilities [35]. We choose a particular set of probabilities by selecting the maximum entropy measure as the objective. The problem specific optimisation problem is defined by,

$$\begin{gathered}
\max_{\mathbf{p} \in \mathbb{R}^M} \sum_{w^*=1}^{M} p_{w^*} \log p_{w^*} \quad \text{subject to} \\
\sum_{w^*=1}^{M} p_{w^*} p_{z^*}^{t|V_{w^*}} = p_{z^*}^{t|V} \quad \forall z^* \in \{1, 2, \ldots, N\}, \\
p_{w^*} \geq 0 \quad \forall w^* \in \{0, 1, \ldots, M\}, \\
\sum_{w^*=1}^{M} p_{w^*} = 1.
\end{gathered} \tag{4.13}$$

The CDO tranche prices constraint will only hold if the bespoke and index collateral portfolios have the same underlying, while the constraint on default probabilities will always hold.

4.2.4. Augmented Lagrangian Methods

The augmented Lagragian method seeks the solution by replacing the original constrained problem with a sequence of unconstrained subproblems in which the objective function is formed by the original objective of the constrained optimisation plus additional "penalty" terms. These terms are made up of constraint functions multiplied by a positive coefficient.

For the problem at hand, the augmented Lagrangian function is given by,

$$L_M^*(x;s;\lambda_M;\mu_M) = \sum_{m=1}^{M} p_m \log p_m + \lambda_M^{(1)}\left(1-\sum_{i=1}^{M} p_{i,M}\right) + \lambda_M^{(2)}\sum_{k=1}^{N}\left(p_k^{t|V} - \sum_{i=1}^{M} p_{i,M} p_k^{t|V_m}\right) + \frac{1}{2\mu_M}\left(1-\sum_{i=1}^{M} p_{i,M}\right)^2 + \frac{1}{2\mu_M} + \left(\sum_{k=1}^{N}\left(p_k^{t|V} - \sum_{i=1}^{M} p_{i,M} p_k^{t|V_m}\right)\right)^2. \tag{4.14}$$

4.2.5. Implementation Issues

In practice, a significant proprotion of computation time is not spent solving the optimisation problem itself, but rather computing the coefficients in the linear constraints. The marginal default probability constraints require the evaluation of the conditional default probability given in the previous chapter for each name under each scenario.

One has the option to only match the cumulative implied default probability to the end of the lifetime of the bespoke CDO or perhaps at selected times only. The advantage of dropping constraints is twofold. Firstly it reduces the computational burden by reducing the number of coefficients that need to be computed, thus leading to a faster pricing algorithm. Secondly, it loosens the conditions required of the probabilities $\mathbf{p}$, thus resulting in factor implied distributions with superior "quality," as given by higher values for the fitness function [5]. Matlab is used for implementing the model.

4.3. Interpreting the Implied Distributions

The valuation of CDOs depends on the portfolio loss distribution. For the pricing of a CDO or CDO^2 it is sufficient to know the portfolio loss distributions over different time horizons. The implied credit loss distributions should be considered relative to the prior models, before deriving efficient frontiers for the bespoke portfolios. This is crucial as deviations from the model will result in a sub-optimal asset allocation strategy.

Figure 2 shows the implied and model distribution of default losses for the benchmark portfolio under the Gaussian copula assumption. Kernel smoothing was applied to the density results. This approach is a nonparametric way of estimating the probability density function of a random variable. This density estimation technique makes it possible to extrapolate the data to the entire population.

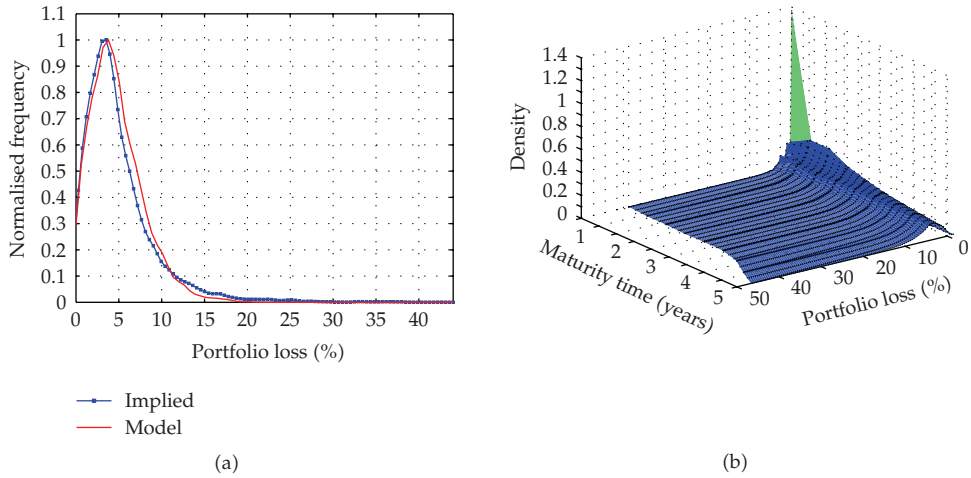


Figure 2: The comparison of the implied loss density under Gaussian copula assumption: (a) shows the deviation of the model density from the implied density at the 5 year horizon, (b) displays the implied credit loss density surface up to the 5 year horizon.

The first thing to note is the typical shape of credit loss distributions. Due to the common dependence on the factor V, defaults are correlated. This association gives rise to a portfolio credit loss density that is right-skewed and has a right-hand tail. The loss surface as a whole is widening and flattening, with an increasing expected portfolio loss.

The Clayton copula model displays similar deviations. This is shown in Figure 3. The Clayton assumption will still under-estimate small losses, but this effect is less pronounced than the Gaussian case. The maximum portfolio loss obtained over a 5 year horizon is 53.7% as compared to the 44.3% in the Gaussian case. The loss surface as a whole is widening and flattening with an increasing expected portfolio loss. The maximum loss increases from 44.2% over a 6 month horizon to 53.7% over five years. The Clayton copula model still does not perfectly capture the market dynamics, but is an improvement over the Gaussian case.

Similar sentiments on the weakness of Gaussian copula are shared by Li and Liang [36].

Figure 4 uses a logarithmic scale for the probabilities to show the tail effects more clearly.

The probabilities decrease very quickly under the Gaussian and Clayton copula assumptions. The effect of thicker tails under the Clayton copula can easily be seen to dominate the Gaussian copula. In the pricing of the super senior tranche, the Clayton model exhibits higher expected losses due to the excess mass concentrated in the tails of the distribution, the tranche spread will be higher under this model than the Gaussian case. These deviations in the implied distribution under different distributional assumptions will filter through to the resulting efficient frontiers. Due to the higher tail probabilities the Clayton ETL efficient frontiers will be significantly different from frontiers resulting from the Gaussian assumption. This feature will be shown in the subsequent sections.

The results so far also have a practical edge for credit risk management. The likelihood of extreme credit losses is increased under the Clayton copula assumption. This is due to the lower tail dependency exhibited by this copula function.

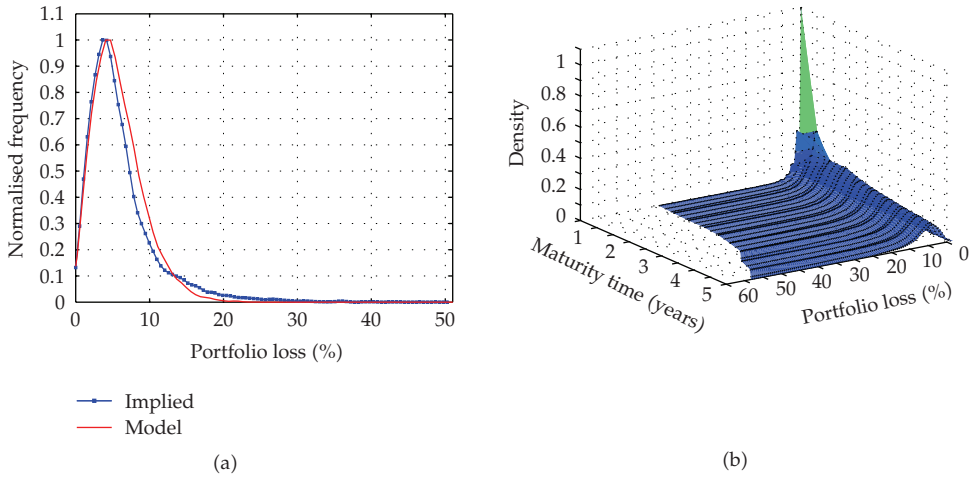


Figure 3: The comparison of the implied loss density under Clayton copula assumption: (a) shows the deviation of the model density from the implied density at the 5 year horizon, (b) displays the implied credit loss density surface up to the 5 year horizon.

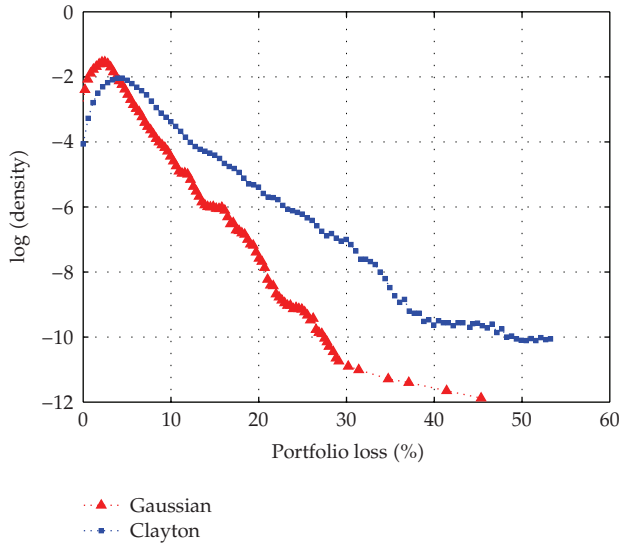


Figure 4: The comparison of implied tail probabilities under different copula assumptions.

5. Risk Characterisation of Credit portfolios

Comparison of uncertainty in outcomes is central to investor preferences. If the outcomes have a probabilistic description, a wealth of concepts and techniques from probability theory can be applied. The main objective in the following section is to present a review of the fundamental work by Artzner et al. [37] and Föllmer and Schied [38] from an optimisation point of view, since a credit risk measure will represent one of the objectives in the CDO optimisation problem.

Let R denote the set of random variables defined on the probability space $(\Omega_p, \mathcal{F}_p, \mathbb{Q}_p)$. We define Ω_p as a set of finitely many possible scenarios for the portfolio p. Financial risks are represented by a convex cone $\mathcal{M} \subseteq (\Omega_p, \mathcal{F}_p, \mathbb{Q}_p)$ of random variables. Any random variable L_c in this set will be interpreted as a possible loss of some credit portfolio over a given time horizon. The following provides a definition of a convex cone.

Definition 5.1 (convex cone). $\mathcal{M}$ is a convex cone if,

(1) $L_c^{(1)} \in \mathcal{M}$ and $L_c^{(2)} \in \mathcal{M}$ implies that $L_c^{(1)} + L_c^{(2)} \in \mathcal{M}$; and

(2) $\lambda^* L_c \in \mathcal{M}$ for every $\lambda^* \geq 0$.

Definition 5.2 (measures of risk). Given some convex cone, $\mathcal{M}$ of random variables, a measure of risk Θ with domain $\mathcal{M}$ is a mapping:

$$\Theta : \mathcal{M} \longrightarrow \mathbb{R}. \tag{5.1}$$

From an economic perspective, $\Theta(L_c)$, can be regarded as the capital buffer that should be set aside for adverse market movements. In order to measure and control the associated risk Artzner et al. [37] introduced an axiomatic framework of coherent risk measures which were recently "generalised" by Föllmer and Schied [38] to convex risk measures.

Definition 5.3 (Convex risk measures). A mapping $\Theta : \mathcal{M} \to \mathbb{R}$ is called a *convex risk measure*, if and only if it is

(1) *convex* for every $L_c^{(1)}$ and $L_c^{(2)} \in \mathcal{M}$, one has $\Theta(\Lambda L_c^{(1)} + (1-\Lambda)L_c^{(2)}) \leq \Lambda\Theta(L_c^{(1)}) + (1-\Lambda)\Theta(L_c^{(2)})$ for some $\Lambda \in \mathbb{R}$;

(2) *monotone* for every $L_c^{(1)}$ and $L_c^{(2)} \in \mathcal{M}$ with $L_c^{(1)} \leq L_c^{(2)}$, one has $\Theta(L_c^{(1)}) \leq \Theta(L_c^{(2)})$; and

(3) *translation invariant* if a_* is a constant then $\Theta(L_c + a_*\mathbf{1}) = -a_* + \Theta(L_c)$, where $\mathbf{1}$ denotes the unit vector.

By adding positive homogeneity to these properties, one obtains the following definition.

Definition 5.4 (coherent risk measures). A convex risk measure Θ is called *coherent*, if in addition it is,

Positively homogeneous if $\Lambda \geq 0$ then $\Theta(\Lambda L_c) = \Lambda\Theta(L_c)$ holds.

Denote the credit loss distribution of L_c by $F_{L_c}(l_c) = \mathbb{Q}(L_c \leq l_c)$. In the analysis we are concerned solely with two risk measures which are based on the loss distribution F_{L_c}, namely VaR and ETL. We now recall the definition of these risk measures.

Definition 5.5 (value-at-risk (VaR)). Given some confidence level $\beta_* \in (0,1)$, the *Value-at-Risk* (VaR) of the credit portfolio at the confidence level β_* is given by the smallest number l_c such that the probability that the loss L_c exceeds l_c is no larger than $(1-\beta_*)$. Formally,

$$\text{VaR} = \inf\left(l_c \in \mathbb{R} : \mathbb{Q}\left(L_c > l\right) \leq 1 - \beta_*\right). \tag{5.2}$$

This definition of VaR coincides with the definition of an β_*-quantile of the distribution of L_c in terms of a generalised inverse of the distribution function F_{L_c}. We observe this coincidence by noting,

$$\begin{aligned} \text{VaR} &= \inf\left(l_c \in \mathbb{R} : 1 - F_{L_c}(l_c) \leq 1 - \beta_*\right) \\ &= \inf\left(l_c \in \mathbb{R} : F_{L_c}(l_c) \geq \beta_*\right). \end{aligned} \tag{5.3}$$

For a random variable Y we will denote the β_*-quantile of the distribution by $q_{\beta_*}(F_Y)$, and write $\text{VaR}_{\beta_*}(Y)$ when we wish to stress that the quantile should be interpreted as a VaR number. A simple definition of ETL which suffices for continuous loss distributions is as follows.

Definition 5.6 (expected-tail-loss (ETL)). Consider a loss L with continuous $F_L df$ satisfying

$$\int_{\mathbb{R}} |l| dF_{L_c}(l_c) < \infty. \tag{5.4}$$

Then the *Expected-Tail-Loss* at confidence level $\alpha \in (0,1)$, is defined to be,

$$\text{ETL}_{\beta_*} = \mathbb{E}\left[L_c \mid L_c \geq \text{VaR}_{\beta_*}\right] = \frac{\mathbb{E}[L_c; L_c \geq \text{VaR}_{\beta_*}(L_c)]}{\mathbb{Q}[L_c \geq \text{VaR}_{\beta_*}(L_c)]}. \tag{5.5}$$

6. Optimal Bespoke CDO Design

Credit portfolio optimisation plays a critical role in determining bespoke CDO strategies for investors. The most crucial tasks in putting together bespoke CDOs is choosing the underlying credits that will be included in the portfolio. Usually investors often express preferences on individual names to which they willing to have the exposure, while there are likely to be credit rating constraints and industry/geographical concentration limits imposed by rating agencies and/or investors.

Given these various requirements, it is up to the credit structurer to optimise the portfolio and achieve the best possible tranche spreads for investors. In the following analysis, we focus on the asset allocation rather than credit selection strategy, which remains a primary modelling challenge for credit structurers.

Davidson [2] provides an interesting analogy between bespoke CDO optimisation and Darwin's evolutionary cycles. In the natural world, life adapts to suit the particulars of its environment. The adaptation to a specific environment is possible due to the application of powerful set evolutionary techniques—reproduction, mutation and survival of the fittest. Nature then explores the full range of possible structures to hone in on those that are most perfectly suited to the surroundings.

Creating a portfolio for a CDO can broadly be seen in similar ways. Given a certain set of investor defined, constraintsstructurers need to be able to construct a credit portfolio that is best suited to the market environment. Added to these constraints are market constraints such as trade lot restrictions and liquidity and availability of underlying credits. If the portfolio does not suit these conditions, it evolves so that only those with the best fit (highest tranche spreads) will survive. Many of the same techniques used in the natural world can be applied to this structuring process.

Evolutionary computation methods are exploited to allow for a generalisation of the underlying problem structure and to solve the resulting optimisation problems, numerically in a systematic way. The next section will briefly discuss some of the basic concepts of multiobjective optimisation and outline the NSGA-II algorithm used for solving the challenging CDO optimisation problem. The CDO optimisation model is then outlined before conducting a prototype experiment on the test portfolio constructed from the constituents of the iTraxx Europe IG S5 index.

6.1. Multiobjective Optimisation

Many real-world problems involve simultaneous optimisation of several incommensurable and often competing objectives. In single-objective optimisation the solution is usually clearly defined, this does not hold for multiobjective optimisation problems. We define formally, the multiobjective optimisation problem to define other improtant concepts used in this chapter. All problems are assumed to be minimisation problems unless otherwise specified To avoid inserting the same reference every few lines, note that the terms and definitions are taken from Zitzler [39], and adapted to CDO optimisation problem.

Definition 6.1 (multiobjective optimisation problem). A general *multiobjective optimisation Problem* (MOP) includes a set of n_* parameters (decision variables), a set of k_1 objective functions, and a set of k_2 constraints. Objective functions and constraints are functions of the decision variables. The optimisation goal is to obtain,

$$\begin{aligned} \min \mathbf{y} = \mathbf{f}(\mathbf{x}) &= \left(f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_{k_1}(\mathbf{x})\right) \quad \text{subject to,} \\ e(\mathbf{x}) &= \left(e_1(\mathbf{x}), e_2(\mathbf{x}), \ldots, e_{k_2}(\mathbf{x})\right) \leq 0, \\ \mathbf{x} &= \left(x_1, x_2, \ldots, x_n\right) \in \mathbf{X}, \\ \mathbf{y} &= \left(y_1, y_2, \ldots, y_{k_1}\right) \in \mathbf{Y}, \end{aligned} \tag{6.1}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the decision and objective vectors, respectively, whilst $\mathbf{X}$ and $\mathbf{Y}$ denote the decision and objective spaces, respectively. The constraints $e(\mathbf{x}) \leq \mathbf{0}$ determine the set of feasible solutions.

Without loss of generality, a minimisation problem is assumed here. For maximisation or mixed maximisation/minimisation problems the definitions are similar.

Definition 6.2 (vector relationship). For any two vectors $\mathbf{r}$ and $\mathbf{r}^*$

$$\begin{aligned} \mathbf{r} = \mathbf{r}^* \quad &\text{iff} \quad r_i = r_i^* \ \forall i \in \{1, 2, \ldots, k_1\}, \\ \mathbf{r} \lneq \mathbf{r}^* \quad &\text{iff} \quad r_i \leq r_i^* \ \forall i \in \{1, 2, \ldots, k_1\}, \\ \mathbf{r} < \mathbf{r}^* \quad &\text{iff} \quad \mathbf{r} \leq \mathbf{r}^* \wedge \mathbf{r} \leq \mathbf{r}^*. \end{aligned} \tag{6.2}$$

The relations $\geq$ and $>$ are similarly defined.

Although the concepts and terminology of Pareto optimality are frequently invoked, most often they are erroneously used in literature. We now define this set of concepts to ensure understanding and consistency.

Definition 6.3 (Pareto dominance). For any two vectors $\mathbf{r}$ and $\mathbf{r}*$

$$\begin{aligned}
&\mathbf{r} \prec \mathbf{r}^* \quad \text{iff} \quad \mathbf{f}(\mathbf{r}) \leq \mathbf{f}(\mathbf{r}^*) \text{ then } \mathbf{r} \text{ dominates } r^*,\\
&\mathbf{r} \not\preceq r^* \quad \text{iff} \quad \mathbf{f}(\mathbf{r}) \leq \mathbf{f}(\mathbf{r}^*) \text{ then } r \text{ weakly dominates } r^*,\\
&\mathbf{r} \sim \mathbf{r}^* \quad \text{iff} \quad \mathbf{f}(\mathbf{r}) \not\leq \mathbf{f}(\mathbf{r}^*) \wedge \mathbf{f}(\mathbf{r}^*) \not\leq \mathbf{f}(\mathbf{r}) \text{ then } \mathbf{r} \text{ is indifferent to } \mathbf{r}^*.
\end{aligned} \tag{6.3}$$

The definitions for a maximisation problems ($\succ, \not\succeq, \sim$) are analogous.

Definition 6.4 (Pareto optimality). A decision vector $\mathbf{x} \in \mathbf{X}_f$ is said to be nondominated regarding a set $\mathbf{K} \subset \mathbf{X}_f$ iff

$$\nexists\, \mathbf{k} \in \mathbf{K} : \mathbf{k} \prec \mathbf{K}. \tag{6.4}$$

If it is clear within the context which set $\mathbf{K}$ is meant, it is simply left out. Moreover, $\mathbf{x}$ is said to be Pareto optimal iff $\mathbf{x}$ is nondominated regarding $\mathbf{X}_f$.

The entirety of all Pareto-optimal solutions is termed the *Pareto-optimal set*; the corresponding objective vectors form the *Pareto-optimal frontier* or surface.

Definition 6.5 (nondominated sets and frontiers). Let $\mathbf{K} \subset \mathbf{X}_f$. The function $\mathcal{N}_d(\mathbf{K})$ gives the set of nondominated decision vectors in $\mathbf{K}$:

$$\mathcal{N}_d(\mathbf{K}) = \{\mathbf{k} \in \mathbf{K} \mid \mathbf{k} \text{ is nondominating regarding } \mathbf{K}\}. \tag{6.5}$$

The set $\mathcal{N}_d(\mathbf{K})$ is the nondominated set regarding $\mathbf{K}$, the corresponding set of objective vectors $\mathbf{f}(\mathcal{N}_d(\mathbf{K}))$ is the nondominated front regarding $\mathbf{K}$. Furthermore, the set $\mathbf{X}_{\mathcal{N}_d} = \mathcal{N}_d(\mathbf{X}_f)$ is called the *Pareto-optimal set* and the set $\mathbf{Y}_{\mathcal{N}_d} = \mathbf{f}(\mathbf{X}_f)$, is denoted as the *Pareto-optimal frontier*.

6.2. Nondominating Sort Genetic Algorithm (NSGA-II)

The second generation NSGA-II is a fast and elitist multiobjective evolutionary algorithm. The main features are (http://www.kxcad.net/ESTECO/modeFRONTIER320/html/userman/ch07s01s10.html).

(i) *Implementation of a fast nondominated sorting procedure*: Sorting of the individuals of a given population is according to the level of nondomination. Generally, nondominated sorting algorithms are computationally expensive for large population sizes however, the adopted solution performs a clever sorting strategy.

(ii) *Implementation of elitism for multiobjective search*: Using an elitism-preserving approach introduces storing all nondominated solutions discovered so far, beginning from the initial population. Elitism enhances the convergence properties towards the true Pareto-optimal set.

(iii) *Adopting a parameter-less diversity preservation mechanism*: Diversity and spread of solutions is guaranteed without use of sharing parameters, since NSGA-II adopts a suitable parameter-less niching approach. This niche is accomplished by

the crowding distance measure, which estimates the density of solutions in the objective space, and the crowded comparison operator, which guides the selection process towards a uniformly spread Pareto frontier.

(iv) *Constraint handling method does not make use of penalty parameters*: The algorithm implements a modified definition of dominance in order to solve constrained multiobjective problems efficiently.

(v) *Real-coded and binary-coded design variables*: A new feature is the application of the genetic algorithms in the field of continuous variables.

6.2.1. Representation of Individuals

The first stage of building the EA is to link the "real world" to the "EA world." This linking involves setting up a bridge between the original problem context and the problem-solving solving space, where evolution takes place. The objects forming possible solutions within the original problem context are referred to as a *phenotype*, while their encoding are called *genotypes* or more commonly *chromosomes*.

The mapping from phenotype space to the genotype space is termed *encoding*. The inverse mapping is termed *decoding*.

Choosing an appropriate representation for the problem being solved is improtant in the design of a successful EA. Often it comes down to good knowledge of the application domain [40].

Many different encoding methods have been proposed and used in EA development. Few frequently applied representations are: binary, integer and real valued representation. Real-valued or Floating-point representation is often the most sensible way to represent a candidate solution to of a problem. This approach is appropriate when the values that we want to represent as genes, originate from a continuous distribution.

The solutions to the proposed CDO optimisation models are real-valued. This study opted to use the real-valued encoding, for the sake of operational simplicity. The genes of a chromosome are real numbers between 0 and 1, which represents the weights invested in the different CDS contracts. However, the summation of these weights might not be 1 in the initialisation stage or after genetic operations. To overcome this problem, the weights are normalised as follows:

$$x_i' = \frac{x_i}{\sum_{i=1}^{N} x_i}. \tag{6.6}$$

6.2.2. Evaluation Function

The role of the *evaluation function* is to represent the requirements to which the population should adapt. This role forms the basis of selection. Technically, the function assigns a quality measure to the genotypes. Typically the function is composed from a quality measure in the phenotype space and the inverse representation [40]. In the evolutionary context the function is usually referred to as the fitness function.

In the bespoke CDO optimisation problem, we introduce two objective to the problem. These objectives are namely the CDO tranche return, and the portfolio tail risk measured by ETL. Various constraints are then introduced to study the dynamics of the Pareto frontier under various conditions.

6.2.3. Population Models and Diversity Preservation

Unlike operators, which operate on individuals, selection operators like parent and survivor selection work at population level. In most EA applications, the population size is constant and does not change during the evolutionary search. The *diversity* of the population is a measure of the number of different solutions present.

In NSGA-I, the well-known sharing function approach was used, which was found to maintain sustainable diversity in a population [4]. There are however two difficulties with this sharing function. According to Deb et al. [4], the performance of the sharing function method in maintaining a spread of solutions depends largely on the value of σ_s, and since each solution must be compared with all other solutions in the population, the overall complexity of the sharing function approach is $\mathbf{O}(N_{\text{pop}}^2)$.

Deb et al. [4] replace the sharing function approach with a crowded-comparison approach in NSGA-II. This eliminates both the above difficulties to some extent. This approach does not require any user-defined parameter for maintaining diversity among population members, and also has a better computational complexity.

To describe this approach, we first define a metric for density estimation and then present the description of the crowded-comparison operator.

6.2.4. Crowded Comparison Operator

The density of solutions surrounding a particular solution in the population must firstly be estimated by calculating the average distance of two points on either side of this point, along each of the objectives.

Once the nondominated sort is complete, the crowding distance is assigned. Individuals are selected based on rank and crowding distance. The crowding-distance computation requires sorting the population according to each objective function value in ascending order of magnitude. The boundary solutions (solutions with smallest and largest function values) for each objective function are assigned an infinite distance value. All other intermediate solutions are assigned a distance value equal to the absolute normalised difference in the function values of two adjacent solutions. This calculation is continued with other objective functions. The overall crowding-distance value is calculated as the sum of individual distance values corresponding to each objective. Each objective function is normalised before calculating the crowding distance. (See Deb et al. [4] for more details regarding the algorithm.)

The complexity of this procedure is governed by the sorting algorithm and has computational complexity of $\mathbf{O}(M_* N_{\text{pop}} \log N_{\text{pop}})$, where there are M_* independent sorting of at most N_{pop} individuals, when all population members are in one front $\mathcal{O}$.

After all population members in the set $\mathcal{O}$ are assigned a distance metric, we can compare two solutions for their extent of proximity with other solutions. A solution with a smaller value of this distance measure is, in some sense, more crowded by other solutions.

6.2.5. Nondomination Sorting Approach

In order to identify solutions in the NSGA of the first nondominated front, each solution is compared with every other solution in the population to find if it is dominated (We summarise the algorithm discussed in Deb et al. [4].). This procedure requires $\mathbf{O}(k_1 N_{\text{pop}})$

comparisons for each solution, where k_1 is the number of objectives. This process is continued to find all members of the first nondominated level in the population, the total complexity is $\mathbf{O}(k_1 N_{\text{pop}}^2)$ computations. At this stage, all individuals in the first nondominated front are found. In order to find the individuals in the next nondominated front, the solutions of the first front are discounted temporarily and the above procedure is repeated. In the worst case, the task of finding the second front also requires $\mathbf{O}(k_1 N_{\text{pop}}^2)$ computations, particularly when number of solutions belong to the second and higher nondominated levels. This argument is true for finding third and higher levels of nondomination.

The worst case is when there are N_{pop} fronts and there exists only one solution in each front. This requires an overall $\mathbf{O}(k_1 N_{\text{pop}}^3)$ computations. For each solution we calculate two entities:

(1) the domination count n_q, the number of solutions which dominate the solution; and

(2) the set of solutions $\mathbf{H}_q$ that the solution dominates. This requires $\mathbf{O}(k_1 N_{\text{pop}}^2)$ comparisons.

Solutions in the first nondominated front will have their domination count as zero. For each solution with $n_q = 0$, each member $q \in \mathbf{H}_q$ is visited and reduce its domination count by one. By doing so, these members belong to the second nondominated front. This process continues until all fronts are identified. For each solution in the second or higher level of nondomination, the domination count can be at most $N_{\text{pop}} - 1$. Since there are at most such $N_{\text{pop}} - 1$ solutions, the total complexity of the procedure is $\mathbf{O}(N_{\text{pop}}^2)$. Thus, the overall complexity of the procedure is $\mathbf{O}(k_1 N_{\text{pop}}^3)$.

In NSGA-II the complexity reduction is due to the realisation that the body of the first inner loop (for each $p_i^* \in \mathcal{J}_i$) is executed exactly N_{pop} times as each individual can be the member of at most one front and the second inner loop (for each $q \in \mathbf{H}_p$) can be executed at maximum $N_{\text{pop}} - 1$ times for results in the overall computations. It is improtant to note that although the time complexity has reduced to $\mathbf{O}(k_1 N_{\text{pop}}^2)$, the storage requirement has increased to $\mathbf{O}(N_{\text{pop}}^2)$.

6.2.6. Parent Selection Mechanism

The selection operator determines, which individuals are chosen for mating and how many offspring each selected individual produces. Once the individuals are sorted based on nondomination with crowding distance assigned, the selection is carried out using a crowded comparison operator described above. The comparison is carried out as below based on

(1) nondomination rank p_{rank}^*, that is, individuals in FRONT_i will have their rank $p_{\text{rank}}^* = i$; and

(2) crowding distance $\text{FRONT}_i(d_j)$.

The selection of individuals is through a binary tournament selection with the crowed-comparison operator. In *tournament selection* the individuals are chosen randomly from the population and the best individual from this group is selected as parent. This process is repeated as often as individuals must be chosen. These selected parents produce uniform at random offspring. The parameter for tournament selection is the tournament size. This takes values ranging from 2 to the number of individuals in population [41].

6.2.7. Mutation Operation

Mutation causes individuals to be randomly altered. These variations are mostly small. They will be applied to the variables of the individuals with a low probability. Offspring are mutated after being created by recombination. Mutation of real variables means, that randomly created values are added to the variables with a low mutation probability. The probability of mutating a variable is inversely proprotional to the number of variables (dimensions). The more dimensions one individual has, the smaller is the mutation probability. Different papers reproted results for the optimal mutation rate [41].

The mutation step size is usually difficult to choose. The optimal step-size depends on the problem considered and may even vary during the optimisation process. Small mutation steps are often successful, especially when the individual is already well adapted. However, large mutation steps can produce good results with a faster convergence rate. According Pohleim [41] good mutation operators should often produce small step-sizes with a high probability and large step-sizes with a low probability.

In the NSGA-II a polynomial mutation operator is used. This operator is defined by the following:

$$\Psi_i^{\text{offspring}} = \Psi_i^{\text{parent}} + \left(\Psi_i^{\text{up}} + \Psi_i^{\text{low}}\right)\delta_i, \tag{6.7}$$

where $\Psi_i^{\text{offspring}}$ is the child, and Ψ_i^{parent} is the parent, with Ψ_i^{up} and Ψ_i^{low} being the upper and lower bounds of the parent component. δ_i is a small variation which is calculated from a polynomial distribution given below.

$$\delta_i = \begin{cases} \left(2r_i\right)^{1/(\eta_m+1)} - 1 & r_i < 0.5 \\ 1 - \left(2\left(1 - r_i\right)\right)^{1/(\eta_m+1)} - 1 & r_i \geq 0.5, \end{cases} \tag{6.8}$$

where $r_i \sim \mathcal{U}(0,1)$ and η_m is mutation distribution index.

6.2.8. Recombination Operation

The recombination operator produces new individuals in combining the information contained in two or more parents in the mating population. This mating is done by combining the variable values of the parents. Depending on the representation of the variables different methods must be used.

Using real-value representation, the *Simulated Binary Crossover* operator is used for recombination and *polynomial mutation* is used for mutating the offspring population [42].

Simulated binary crossover simulates the binary crossover observed in nature and is give by the following.

$$\begin{aligned} \Psi_{1,i}^{\text{offspring}} &= \frac{1}{2}\left(\left(1 - \beta_k\right)\Psi_{1,k}^{\text{Parent}} + \left(1 + \beta_k\right)\Psi_{2,k}^{\text{Parent}}\right), \\ \Psi_{2,i}^{\text{offspring}} &= \frac{1}{2}\left(\left(1 + \beta_k\right)\Psi_{1,k}^{\text{Parent}} + \left(1 - \beta_k\right)\Psi_{2,k}^{\text{Parent}}\right), \end{aligned} \tag{6.9}$$

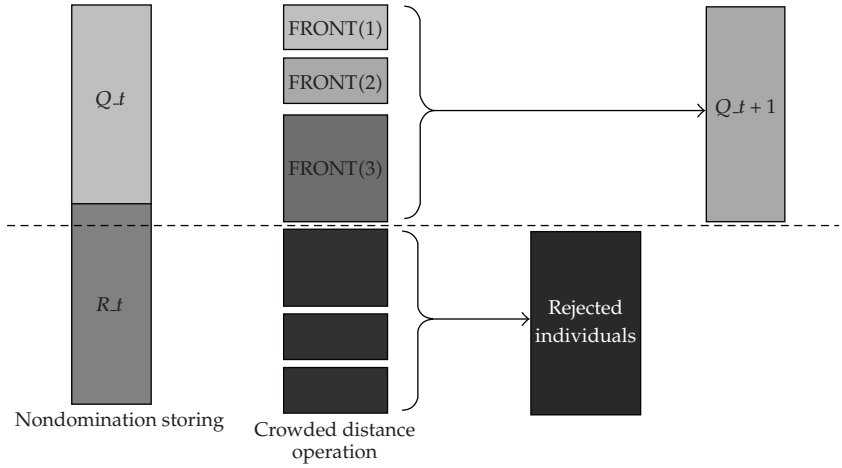


Figure 5: An outline of the NSGA-II procedure.

where $\beta_k (\geq 0)$ is a sample from a random number generated having the density

$$f_c(\beta) = \begin{cases} \dfrac{1}{2}(\eta_c + 1)\beta^{\eta_c} & 0 \leq \beta \leq 1, \\ \dfrac{1}{2}(\eta_c + 1)\dfrac{1}{\beta^{\eta_{c+2}}} & \beta \geq 1. \end{cases} \tag{6.10}$$

This distribution can be obtained from a $u \sim \mathcal{U}(0,1)$ source. η_c is the distribution index for the recombination operator.

6.2.9. Main Loop

First, a random parent population is created. The population is then sorted based on the nondomination outlined above. Each individual is assigned a fitness value (or rank) equal to its nondomination level, with 1 representing the best level. Binary tournament selection, recombination, and mutation operators are then applied to create an offspring population of size N_{pop}.

Due to the introduction of elitism, the current population is compared to previously found best nondominated solutions, the procedure is different after the initial generation. Figure 5 outlines the tth generation of the NSGA-II procedure.

The first step is to combined parent population Q_t with the offspring population R_t. The population is now of size $2N_{\text{pop}}$. The resulting population is then sorted according to nondomination. Since all previous and current population members are included in $N\_t = Q\_t \cup R\_t$, elitism is ensured.

Solutions belonging to the best nondominated set FRONT(1), are of best solutions in the combined population and must be emphasised more than any other solution in the combined population. If FRONT(1) is smaller than N_{pop}, then all individuals in this set are chosen for the next generation, else a truncation operation will have to be applied. If the first

case holds, then the remaining members of the next generation are chosen from subsequent nondominated fronts in the order of their ranking.

In general, the count of solutions in all sets from FRONT(1) to FRONT(l) would be larger than the population size. To choose exactly N_{pop} population members, the solutions of the last frontier are sorted and then the crowded-comparison operator is applied in descending order. The best solutions are chosen to fill all population slots. This procedure is valid for minimisation problems only.

The requisite diversity among nondominated solutions is introduced by using the crowded comparison procedure, which is used in the tournament selection and during the population truncation phases [4].

6.2.10. Initialisation and Termination Conditions

Initialisation is kept simple in most EA applications, the first population is seeded by randomly generated individuals. In principle, problem-specific heuristics can be used in this step, to create an initial population with higher fitness values [40].

We can distinguish two cases for an appropriate termination condition. If the problem has a known optimum fitness level, then an acceptable error bound would be a suitable condition. However, in many multiobjective problems optimum values are not known. In this case one needs to extend this condition with one that certainly stops the algorithm. As noted by Eiben and Smith [40], a common choice is to impose a limit on the total number of fitness evaluations.

6.3. Optimisation Problem

The optimal bespoke structure can be derived by solving the following problem:

$$\begin{aligned} &\min_{\mathbf{x}\in[0,1]^N} \Theta(\mathbf{x}), \\ &\max_{\mathbf{x}\in[0,1]^N} s_{n_t r}(\mathbf{x}) \\ \text{subject to} \quad & \\ &\sum_{i=1}^{N} x_i = 1, \end{aligned} \tag{6.11}$$

where in the above problem, we minimise the portfolio ETL, and simultaneously maximise the $n_{\text{tr}}^{\text{th}}$ tranche spread denoted by $s_{n_{\text{tr}}}$ subject to the portfolio weights summing up to unity. This minimality may typically be the case if only long positions in the underlying credits are allowed.

For practical purposes, it may be crucial to include other real-world constraints. Let $\mathcal{G}^*$ denote the set of organisational, regulatory and physical constraints. The list of constraints, which are commonly included into $\mathcal{G}^*$ are summarised below.

(1) *Upper and lower limits on issuer weights*: especially due to regulatory issues, the portfolio fractional amount invested into each asset may be limited, that is, there

may be individual lower and upper bounds on the portfolio weights.

$$l_{\text{upper}}^{(i)} \leq x_i \leq l_{\text{upper}}^{(i)} \quad \forall i \in \{1, 2, 3, \ldots, N\}. \tag{6.12}$$

When short selling is disallowed, it can be explicitly modelled by setting $l_i = 0$ in this formulation.

(2) *Regional/industry concentration constraints*: due to rating agency and regulatory restrictions, the portfolio may have industry and regional concentration limits, then

$$\sum_{i=1}^{k} x_i \leq U^{(I_{\text{ind}}, I_{\text{geo}})}, \quad \text{where } 0 \leq U^{(I_{\text{ind}}, I_{\text{geo}})} \leq 1, \tag{6.13}$$

where the coordinate $(I_{\text{ind}}, I_{\text{geo}})$ refers to a particular industry I_{ind} in region I_{geo}.

(3) *Cardinality constraints*: these constraints limit the number of assets in the portfolio. They can either be strict

$$\mathcal{N}(x_i > 0) = K^*, \quad 0 < K^* \leq a,\ K^* \in \mathbb{N}. \tag{6.14}$$

or lower and upper bounds on the cardinality can be defined,

$$K_l^* \leq \mathcal{N}(x_i > 0) \leq K_u^*, \quad 0 < K_l^* < K_u^* \leq a,\ K_l^*, K_u^* \in \mathbb{N}. \tag{6.15}$$

(4) *Weighted average rating factor (WARF) constraints*: due to regulatory or organisational restrictions, the portfolio WARF must be above some minimum value,

$$R_{\text{prot}} \geq R_{\text{target}}, \tag{6.16}$$

where in the case of S and P ratings $R_{\text{prot}} \in \{AAA, AA+, AA, AA-, \ldots\}$. (Due to the nature of CDOs, ratings higher than the portfolio rating can be achieved. In this case the constraint may be defined as $R_{\text{nth tranche}} \geq R_{\text{target}}$.)

Rating agencies presuppose that there are two broad credit considerations that determine the risk of a CDO portfolio: collateral diversity by industry and by reference entity, and the credit quality of each asset in the portfolio. With respect to the latter, rating agencies use the concept of the WARF. Each security in the CDO portfolio has a rating (actual or implied) and each rating has a corresponding rating factor. The lower a security's rating, the higher its corresponding rating factor. In order to calculate the weighted average debt rating for a pool of assets, take the par amount of each performing asset and multiply it by its respective rating factor. Then, sum the resulting amounts for all assets in the pool and divide this number by the sum of the par values of the performing assets [1].

Krahnen and Wilde [43] show that the common presumption about macro factor tail risk (extreme systematic risk) being largely held by senior tranches is false. While senior

tranches are in fact primarily exposed to tail risk, the reverse is not true, as the share of tail risk borne by senior tranches is quite limited. They conclude that tail risk in standard CDO transactions is held by all tranches. This argument motivates the use of a tail-risk measure like ETL in the tranche return-risk optimisation to derive optimal structures. The multiobjective optimisation problem including the above constraints is now defined in detail.

$$\begin{aligned}
&\min_{x\in\mathcal{X}} \mathrm{ETL}_{\beta_*}\left(L(t_j)\right),\\
&\max_{x\in\mathcal{X}} s_n = \frac{\sum_{j=1}^{J} D(0,(t_j - t_{j-1})/2)\left(\mathbb{E}[L_{n_{\mathrm{tr}}}(t_j)] - \mathbb{E}[L_{n_{\mathrm{tr}}}(t_{j-1})]\right)}{\sum_{j=1}^{J}(t_j - t_{j-1})D(0,t_j)\mathbb{E}[N_{\mathrm{out}}^{(n_{\mathrm{tr}})}(t_j,t_{j-1})]}\\
&\text{subject to}\\
&l_{\mathrm{upper}}^{(i)} \le x_i \le l_{\mathrm{upper}}^{(i)} \quad \forall i \in \{1,2,3,\ldots,N\},\\
&R_{\mathrm{prot}} \ge R_{\min},\\
&\mathcal{N}(x_i > 0) = 106.
\end{aligned} \tag{6.17}$$

We do not include the industry/region concentration constraints since the numerical analysis is conducted on the benchmark portfolio consisting of 106 constituents of the iTraxx Europe IG S5 index. The regional constraint will fall away, however, the industry constraint will still hold. The study will analyse the industry distributions of the resulting collateral portfolio's, for optimal [6,9]% mezzanine tranches that define the pareto frontier. This is an improtant visualisation tool for bespoke CDO investors, which depicts the trade-off between tranche return and risk, to allow for picking a structure given the respective risk preference. A flat Libor curve of 4.6% is applied. The CDO structure is priced as at 18-April-2008.

The two cases investigated are as follows.

(1) First an examination of the pareto frontiers for bespoke CDO structures under both Gaussian and Clayton copula assumptions is conducted. Long-only positions in the underlying CDS are allowed. The upper trading limit in any particular reference entity is set to 2.5% of the portfolio, whilst the lower limit is set to 0.5% to satisfy the cardinality constraint of having 106 credits in the collateral portfolio.

(2) Then an investigation into the behaviour of the pareto frontiers under increasing upper trading limit is conducted. This consequently allows an investigation of concentration risk and its effects on the producing optimal CDO structures. The upper trading limit is increased to 5% in this case.

Table 1 reprots the NSGA-II parameter setting used to solve the optimisation problem (6.17).

The parameters settings were found to be most appropriate after several runs of the algorithm. The computational time played a crucial role in the assignment of parameter values.

6.4. Prototype Experiment: Pareto Frontier Analysis

6.4.1. Optimal Structures with Long Only Credit portfolio

Bespoke CDOs are commonly preferred among investors because they can be used to execute a variety of customised investment objectives and strategies. Flexibility in choosing the

Table 1: NSGA-II parameter settings.

Evolutionary parameter	Value
Population size	100
Number of generations	500
Crossover probability	90%
Distribution index for crossover	0.01
Distribution index for mutation	0.01
Pool size	50
Tour size	2

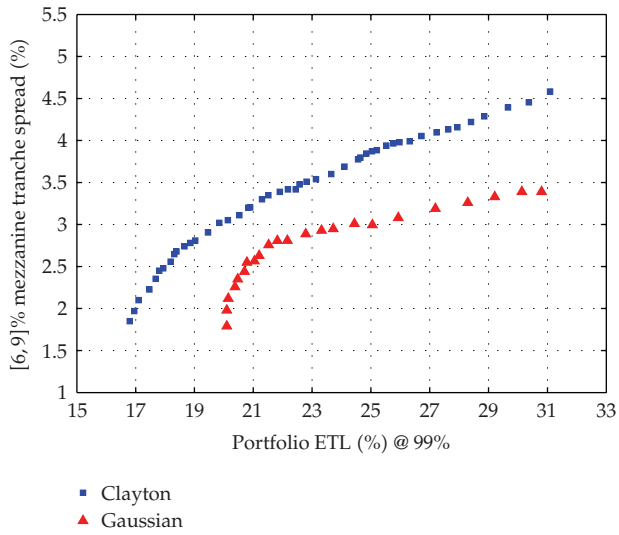


Figure 6: The comparison of bespoke CDO pareto frontiers under different default dependence assumptions.

reference portfolio allows investors to structure the portfolio investments such that it can be properly matched with their investment portfolios, as a separate investment strategy or as a hedging position. Most often investors choose credit portfolios such that diversification is only as a result of buying protection on reference entities in the underlying portfolio.

We derive the pareto frontiers solving problem (6.17) under both the Gaussian and Clayton copula assumptions when investors are restricted to the long-only condition.

Figure 6 displays the difference in pareto frontiers for the two default dependence assumptions. The first observation is that the Gaussian allocation is sub-optimal. For all levels of portfolio risk, measured by ETL, the Clayton allocation will result in higher $[6, 9]\%$ tranche spreads. This deviation is an increasing function of portfolio tail risk and widens to a maximum of 120 bps for an ETL of 30% of the portfolio losses.

The minimum tail risk portfolio can provide a $[6, 9]\%$ mezzanine tranche that pays 180 bps under the Clayton allocation, whilst the Gaussian allocation provides spread of 198 bps. However, there exist a 3.3% difference in ETL for these minimum risk portfolios. A similar spread under the Clayton allocation can be achieved with a ETL of just 16.95%.

We concluded that the asset allocation strategies based on the Gaussian copula will results in sub-optimal bespoke CDO structures.

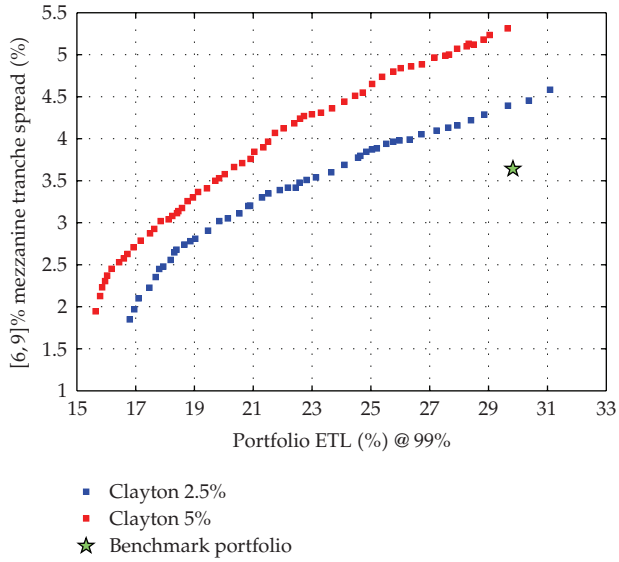


Figure 7: The portfolio notional distribution for the concentrated portfolio case, by issuer credit rating.

6.4.2. Issuer Concentrated Credit Structures

We now investigate the effects of issuer concentration on the pareto frontier. As we concluded in the previous chapter, that investors demand a higher premium for taking on concentration risk, combining this with the leverage effects of tranche technology, will result in higher tranche spreads than the well diversified case. The resulting frontiers are shown in Figure 7 for the $[6, 9]\%$ mezzanine tranche.

Figure 7 compares the resulting pareto-frontiers for the well-diversified and issuer concentrated cases. The CDO spread of the concentrated case over the well-diversified case is an increasing function of portfolio risk. For a 29.64% level of risk, a spread of 92 bps over the well-diversified case can be achieved.

7. Conclusions

The main objective of this research has been to develop and illustrate a methodology to derive optimal bespoke CDO structures using the NSGA-II technology. Until recently, the derivation of such structures was a near to an impossible task. However, with the advent of advance pricing and optimisation techniques applied in this research, credit structurers can use these tools to provide investors with optimal structures. Investors can use this type of methodology to compare similar rated credit structures and make an informed investment decision, especially under the current credit market conditions.

The most improtant finding is that the Gaussian copula allocation produces suboptimal CDO tranche investments. Better tranche returns at all levels of risk are obtained under the Clayton assumption. In the constrained long-only case, for all levels of portfolio risk, measured by ETL, the Clayton allocation will result in higher $[6, 9]\%$ tranche spread. This deviation was shown to be an increasing function of portfolio tail risk and widen to a maximum of 120 bps for an ETL of 30%.

It was also demonstrated that significant improvement of returns over standardised CDOs tranches of similar rating can be achieved using the methodology presented in this paper. The findings also concluded that the leverage effects should became more pronounced for tranches lower down the capital structure. More concentrated bespoke CDO portfolios result in higher tranche spreads. Investor demand higher premiums for taking on the concentration risk. For a 29.64% level of portfolio tail risk, an excess return of 92 bps over the well diversified case can be achieved.

Since the current analysis is restricted to single-factor copula models, one obvious extension for future studies should be on the added benefits multifactor copula models may provide in producing optimal CDO structures. On the use of evolutionary technology, a future study should compare NSGA-II results to the rival, the second generation of the strength pareto evolutionary algorithm (SPEA-2) proposed by the Zitzler et al. [44]. Stability of and convergence to the pareto frontier for optimal CDO structures should also be investigated.

References

[1] A. Rajan, G. McDermott, and R. Roy, *The Structured Credit Handbook*, Wiley Finance Series, John Wiley & Sons, New York, NY, USA, 2006.

[2] C. Davidson, "The evolution will be analysed," *Risk*, vol. 20, no. 2, 2007.

[3] D. Jewan, R. Guo, and G. Witten, "Copula marginal expected tail loss efficient frontiers for CDOs of bespoke portfolios," in *Proceedings of the World Congress in Financial Engineering Conference (WCE '08)*, vol. 2, pp. 1124–1129, London, UK, July 2008.

[4] K. Deb, A. Pratep, S. Agrawal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithmml: NGSA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[5] D. Rosen and D. Saunders, "Valuing CDOs of bespoke portfolios with implied multi-factor models," The Field Institute for Research in Mathematical Sciences, 2007, Working paper, http://www.defaultrisk.com/pp_cdo_23.htm.

[6] M. Whetten, "The Bespoke—a guide to single-tranche synthetic CDOs," Nomura, Fixed Income Research, 2004.

[7] G. Tejwani, A. Yoganandan, and A. Jha, "How do dealers risk manage synthetic cdos," Lehman Brothers, Structured Credit Research, 2007.

[8] M. Sklar, "Fonctions de répartition à n dimensions et leurs marges," *Publications de l'Institut de Statistique de l'Université de Paris*, vol. 8, pp. 229–231, 1959.

[9] P. Embrechths, F. Lindskog, and D. Straumann, "Correlation: pitfalls and alternatives," *Risk*, vol. 12, no. 1, pp. 69–71, 1999.

[10] R. Frey and A. McNeil, "Dependent defaults in models of portfolio credit risk," *Journal of Risk*, vol. 6, no. 1, pp. 59–92, 2003.

[11] M. Gordy, "A risk-factor model foundation for ratings-based bank capital rules," *Journal of Financial Intermediation*, vol. 12, no. 3, pp. 199–232, 2003.

[12] S. Merino and M. Nyfeler, "Calculating portfolio loss," *Risk*, vol. 15, no. 3, pp. 82–86, 2002.

[13] M. Pykthin and A. Dev, "Analytical approach to credit risk modelling," *Risk*, vol. 15, no. 3, pp. 26–32, 2002.

[14] O. Vasicek, "Probability of loss on loan portfolio," KMV Corporation, 1987, http://www.moodyskmv.com/research/files/wp/Probability_of_Loss_on_Loan_Portfolio.pdf.

[15] P. J. Schönbucher and D. Schubert, "Copula-dependent default risk in intensity models," no. TIK-Report 103, Bonn University, 2001, Working paper, http://www.defaultrisk.com/pp_corr_22.htm.

[16] D. G. Clayton, "A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence," *Biometrika*, vol. 65, no. 1, pp. 141–151, 1978.

[17] D. Oakes, "A model for association in bivariate survival data," *Journal of the Royal Statistical Society. Series B*, vol. 44, no. 3, pp. 414–422, 1982.

[18] A. Friend and E. Rogge, "Correlation at first sight," *Economic Notes*, vol. 34, no. 2, pp. 155–183, 2004.

[19] J.-P. Laurent and J. Gregory, "Basket default swaps, CDO's and factor copulas," *Journal of Risk*, vol. 7, no. 4, pp. 103–122, 2005.

[20] L. Schloegl and D. O'Kane, "A note on the large homogeneous portfolio approximation with the Student-*t* copula," *Finance and Stochastics*, vol. 9, no. 4, pp. 577–584, 2005.

[21] P. J. Schönbucher, "Taken to the limit: simple and not-so-simple loan loss distributions," Bonn University, 2002, Working paper, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=378640.

[22] A. W. Marshall and I. Olkin, "Families of multivariate distributions," *Journal of the American Statistical Association*, vol. 83, no. 403, pp. 834–841, 1988.

[23] D. X. Li, "On default correlation: a copula function approach," *Journal of Financial Economics*, vol. 9, no. 4, pp. 43–54, 2000.

[24] H. Brunlid, *Comparative analysis of Hyperbolic copulas induced by a one factor Lévy model*, M.S. thesis, Department of Economics, Lund University, Lund, Sweden, 2007, Working paper, http://www.defaultrisk.com/pp_cdo_17.htm.

[25] X. Burtschell, J. Gregory, and J.-P. Laurent, "Beyond the Gaussian copula: stochastic and local correlation," *Journal of Credit Risk*, vol. 3, no. 1, pp. 31–62, 2007.

[26] M. van der Voort, "An implied loss model," Abn Amro Bank & Erasmus University, Rotterdam, 2006, Working paper, http://www.defaultrisk.com/pp_crdrv115.htm.

[27] Y. Elouerkhaouia, "The Marshall-Olkin copula: a new approach to correlation skew modelling," in *Proceedings of the 7th Annual Risk Management Conference*, Paris, France, August 2005.

[28] J. Hull and A. White, "Valuing credit derivatives using an implied copula approach," *Journal of Derivatives*, vol. 14, no. 4, pp. 8–28, 2006.

[29] M. S. Joshi and A. Stacy, "Intensity gamma: a new approach to pricing portfolio credit derivatives," *Risk*, vol. 20, no. 7, pp. 1–12, 2006.

[30] M. B. Walker, "An incomplete market model for collateralized debt obligations," Department of Physics, University of Toronto, 2005, Working paper, http://www.physics.utoronto.ca/~qocmp/incompleteCDO.2005.10.05.pdf.

[31] M. B. Walker, "A term structure model for collateralized debt obligations," Department of Physics, University of Toronto, July 2005, Working paper, http://www.physics.utoronto.ca/~qocmp/walkerRisk.2005.07.30.pdf.

[32] S. Willemann, "Fitting the CDO correlation skew: a tractable structural jump-diffusion model," *Journal of Credit Risk*, vol. 3, no. 1, pp. 63–90, 2007.

[33] J. Turc and P. Very, "Pricing and hedging correlation products," Society General, Quantitative Strategy, 2004.

[34] M. Avellaneda, R. Buff, C. Friedman, N. Grandechamp, L. Kruk, and J. Newman, "Weighted Monte Carlo: a new technique for calibrating asset-pricing models," *International Journal of Theoretical and Applied Finance*, vol. 4, no. 1, pp. 91–119, 2001.

[35] P. Glasserman and B. Yu, "Large sample properties of weighted monte carlo estimators," *Operations Research*, vol. 53, no. 2, pp. 298–312, 2005.

[36] D. X. Li and M. Liang, "CDO^2 squared pricing using Gaussian mixture model with transformation of loss distribution," Barclays Capital, Quantitative Analytics Global Credit Derivatives, 2005.

[37] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent measures of risk," *Mathematical Finance*, vol. 9, no. 3, pp. 203–228, 1999.

[38] H. Föllmer and A. Schied, "Robust preferences and convex measures of risk," in *Advances in Finance and Stochastics*, S. Klaus, P. J. Schonbucher, and D. Sondermann, Eds., pp. 39–56, Springer, Berlin, Germany, 2002.

[39] E. Zitzler, *Evolutionary algorithms for multi-objective optimization: methods and applications*, Ph.D. thesis, Swiss Federal Institute of Technology, Zurich, Switzerland, 1999, ftp://ftp.tik.ee.ethz.ch/pub/people/thiele/paper/zitzler-diss.ps.gz.

[40] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing*, Natural Computing Series, Springer, Berlin, Germany, 2007.

[41] H. Pohleim, "Evolutionary algorithms: overview, methods and operators," GEATbx: Gentic & Evolutionary AlgorithmToolbox for Matlab, 2005, http://www.geatbx.com/download.html.

[42] K. Deb and R. B. Agrawal, "Simulated binary crossover for continuous search space," *Complex Systems*, vol. 9, no. 2, pp. 115–148, 1995.

[43] J. P. Krahnen and C. Wilde, "Risk transfer in CDOs," Goethe University Frankfurt, Finance Department, 2008, Working paper, http://www.defaultrisk.com/pp_cdo_56.htm.

[44] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: improving the strength pareto evolutionary algorithm," TIK-Report 103, Swiss Federal Institute of Technology, Zurich, Switzerland, 2001, http://www.tik.ee.ethz.ch/sop/publicationListFiles/zlt2001a.pdf.

Hindawi Publishing Corporation
Journal of Applied Mathematics and Decision Sciences
Volume 2009, Article ID 125308, 22 pages
doi:10.1155/2009/125308

Research Article

Modified Neural Network Algorithms for Predicting Trading Signals of Stock Market Indices

C. D. Tilakaratne,[1] M. A. Mammadov,[2] and S. A. Morris[2]

[1] *Department of Statistics, University of Colombo, P.O. Box 1490, Colombo 3, Sri Lanka*
[2] *Graduate School of Information Technology and Mathematical Sciences, University of Ballarat, P.O. Box 663, Ballarat, Victoria 3353, Australia*

Correspondence should be addressed to C. D. Tilakaratne, cdt@stat.cmb.ac.lk

Received 29 November 2008; Revised 17 February 2009; Accepted 8 April 2009

Recommended by Lean Yu

The aim of this paper is to present modified neural network algorithms to predict whether it is best to buy, hold, or sell shares (trading signals) of stock market indices. Most commonly used classification techniques are not successful in predicting trading signals when the distribution of the actual trading signals, among these three classes, is imbalanced. The modified network algorithms are based on the structure of feedforward neural networks and a modified Ordinary Least Squares (OLSs) error function. An adjustment relating to the contribution from the historical data used for training the networks and penalisation of incorrectly classified trading signals were accounted for, when modifying the OLS function. A global optimization algorithm was employed to train these networks. These algorithms were employed to predict the trading signals of the Australian All Ordinary Index. The algorithms with the modified error functions introduced by this study produced better predictions.

1. Introduction

A number of previous studies have attempted to predict the price levels of stock market indices [1–4]. However, in the last few decades, there have been a growing number of studies attempting to predict the direction or the trend movements of financial market indices [5–11]. Some studies have suggested that trading strategies guided by forecasts on the direction of price change may be more effective and may lead to higher profits [10]. Leung et al. [12] also found that the classification models based on the direction of stock return outperform those based on the level of stock return in terms of both predictability and profitability.

The most commonly used techniques to predict the trading signals of stock market indices are feedforward neural networks (FNNs) [9, 11, 13], probabilistic neural networks (PNNs) [7, 12], and support vector machines (SVMs) [5, 6]. FNN outputs the value of the

stock market index (or a derivative), and subsequently this value is classified into classes (or direction). Unlike FNN, PNN and SVM directly output the corresponding class.

Almost all of the above mentioned studies considered only two classes: the upward and the downward trends of the stock market movement, which were considered as buy and sell signals [5–7, 9, 11]. It was noticed that the time series data used for these studies are approximately equally distributied among these two classes.

In practice, the traders do not participate in trading (either buy or sell shares) if there is no substantial change in the price level. Instead of buying/selling, they will hold the money/shares in hand. In such a case it is important to consider the additional class which represents a hold signal. For instance, the following criterion can be applied to define three trading signals, buy, hold, and sell.

Criterion A.

$$\begin{aligned} &\textit{buy} \quad \textit{if} \quad \Upsilon(t+1) \geq l_u, \\ &\textit{hold} \quad \textit{if} \quad l_l < \Upsilon(t+1) < l_u, \\ &\textit{sell} \quad \textit{if} \quad \Upsilon(t+1) \leq l_l, \end{aligned} \tag{1.1}$$

where $\Upsilon(t+1)$ *is the relative return of the Close price of day* $(t+1)$ *of the stock market index of interest, while* l_l *and* l_u *are thresholds.*

The values of l_l and l_u depend on the traders' choice. There is no standard criterion found in the literature how to decide the values of l_l and l_u , and these values may vary from one stock index to another. A trader may decide the values for these thresholds according to his/her knowledge and experience.

The proper selection of the values for l_l and l_u could be done by performing a sensitivity analysis. The Australian All Ordinary Index (AORD) was selected as the target stock market index for this study. We experimented different pairs of values for l_l and l_u [14]. For different windows, different pairs gave better predictions. These values also varied according to the prediction algorithm used. However, for the definition of trading signals, these values needed to be fixed.

By examining the data distribution (during the study period, the minimum, maximum, and average for the relative returns of the Close price of the AORD are −0.0687, 0.0573, and 0.0003, resp.), we chose $l_u = -l_l = 0.005$ for this study, assuming that 0.5% increase (or decrease) in Close price of day t+1 compared to that of day t is reasonable enough to consider the corresponding movement as a buy (or sell) signal. It is unlikely that a change in the values of l_l and l_u would make a qualitative change in the prediction results obtained.

According to Criterion A with $l_u = -l_l = 0.005$, one cannot expect a balanced distribution of data among the three classes (trading signals) because more data falls into the hold class while less data falls into the other two classes.

Due to the imbalance of data, the most classification techniques such as SVM and PNN produce less precise results [15–17]. FNN can be identified as a suitable alternative technique for classification when the data to be studied has an imbalanced distribution. However, a standard FNN itself shows some disadvantages: (a) use of local optimization methods which do not guarantee a deep local optimal solution; (b) because of (a), FNN needs to be trained many times with different initial weights and biases (multiple training results in more than one solution and having many solutions for network parameters prevent getting a clear

picture about the influence of input variables); (c) use of the ordinary least squares (OLS; see (2.1)) as an error function to be minimised may not be suitable for classification problems.

To overcome the problem of being stuck in a local minimum, finding a global solution to the error minimisation function is required. Several past studies attempted to find global solutions for the parameters of the FNNs, by developing new algorithms (e.g., [18–21]). Minghu et al. [19] proposed a hybrid algorithm of global optimization of dynamic learning rate for FNNs, and this algorithm shown to have global convergence for error backpropagation multilayer FNNs (MLFNNs). The study done by Ye and Lin [21] presented a new approach to supervised training of weights in MLFNNs. Their algorithm is based on a "subenergy tunneling function" to reject searching in unpromising regions and a "ripple-like" global search to avoid local minima. Jordanov [18] proposed an algorithm which makes use of a stochastic optimization technique based on the so-called low-discrepancy sequences to trained FNNs. Toh et al. [20] also proposed an iterative algorithm for global FNN learning.

This study aims at modifying neural network algorithms to predict whether it is best buy, hold, or sell the shares (trading signals) of a given stock market index. This trading system is designed for short-term traders to trade under normal conditions. It assumes stock market behaviour is normal and does not take unexceptional conditions such as bottlenecks into consideration.

When modifying algorithms, two matters were taken into account: (1) using a global optimization algorithm for network training and (2) modifying the ordinary least squares error function. By using a global optimization algorithm for network training, this study expected to find deep solutions to the error function. Also this study attempted to modify the OLS error function in a way suitable for the classification problem of interest.

Many previous studies [5–7, 9, 11] have used technical indicators of the local markets or economical variables to predict the stock market time series. The other novel idea of this study is the incorporation of the intermarket influence [22, 23] to predict the trading signals.

The organisation of the paper is as follows. Section 2 explains the modification of neural network algorithms. Section 3 describes the network training, quantification of intermarket influence, and the measures of evaluating the performance of the algorithms. Section 4 presents the results obtained from the proposed algorithms together with their interpretations. This section also compares the performance of the modified neural network algorithms with that of the standard FNN algorithm. The last section is the conclusion of the study.

2. Modified Neural Network Algorithms

In this paper, we used modified neural network algorithms for forecasting the trading signals of stock market indices. We used the standard FNN algorithm as the basis of these modified algorithms.

A standard FNN is a fully connected network with every node in the lower layer linked to every node in the next higher layer. These linkages are attached with some weights, $w = (w_1, \ldots, w_M)$, where M is the number of all possible linkages. Given weight, w, the network produces an output for each input vector. The output corresponding to the ith input vector will be denoted by $o_i \equiv o_i(w)$.

FNNs adopt the backpropagation learning that finds optimal weights w by minimising an error between the network outputs and given targets [24]. The most commonly used error

function is the Ordinary Least Squares function (OLS):

$$E_{\mathrm{OLS}} = \frac{1}{N}\sum_{i=1}^{N}(a_i - o_i)^2, \tag{2.1}$$

where N is the total number of observations in the training set, while a_i and o_i are the target and the output corresponding to the ith observation in the training set.

2.1. Alternative Error Functions

As described in the Introduction (see Section 1), in financial applications, it is more important to predict the direction of a time series rather than its value. Therefore, the minimisation of the absolute errors between the target and the output may not produce the desired accuracy of predictions [24, 25]. Having this idea in mind, some past studies aimed to modify the error function associated with the FNNs (e.g., [24–27]). These studies incorporated factors which represent the direction of the prediction (e.g., [24–26]) and the contribution from the historical data that used as inputs (e.g., [24, 25, 27]).

The functions proposed in [24–26] penalised the incorrectly predicted directions more heavily, than the correct predictions. In other words, higher penalty was applied if the predicted value, o_i, is negative when the target, a_i, is positive or viceversa.

Caldwell [26] proposed the Weighted Directional Symmetry (WDS) function which is given as follows:

$$f_{\mathrm{WDS}}(i) = \frac{100}{N}\sum_{i=1}^{N} w_{ds}(i)|a_i - o_i|, \tag{2.2}$$

where

$$w_{ds}(i) = \begin{cases} 1.5 & \text{if } (a_i - a_{i-1})(o_i - o_{i-1}) \leq 0, \\ 0.5, & \text{otherwise}, \end{cases} \tag{2.3}$$

and N is the total number of observations.

Yao and Tan [24, 25] argued that the weight associated with f_{WDS} (i.e., $w_{ds}(i)$) should be heavily adjusted if a wrong direction is predicted for a larger change, while it should be slightly adjusted if a wrong direction is predicted for a smaller change and so on. Based on this argument, they proposed the Directional Profit adjustment factor:

$$f_{\mathrm{DP}}(i) = \begin{cases} c_1 & \text{if } (\Delta a_i \times \Delta o_i) > 0\,,\ \Delta a_i \leq \sigma, \\ c_2 & \text{if } (\Delta a_i \times \Delta o_i) > 0,\ \Delta a_i > \sigma, \\ c_3 & \text{if } (\Delta a_i \times \Delta o_i) < 0,\ \Delta a_i \leq \sigma, \\ c_4 & \text{if } (\Delta a_i \times \Delta o_i) < 0,\ \Delta a_i > \sigma, \end{cases} \tag{2.4}$$

where $\Delta a_i = a_i - a_{i-1}$, $\Delta o_i = o_i - o_{i-1}$, and σ is the standard deviation of the training data (including validation set). For the experiments authors used $c_1 = 0.5$, $c_2 = 0.8$, $c_3 = 1.2$, and $c_4 = 1.5$ [24, 25]. By giving these weights, they tried to impose a higher penalty the predictions whose direction is wrong and the magnitude of the error is lager, than the other predictions.

Based on this Directional Profit adjustment factor (2.4), Yao and Tan [24, 25] proposed Directional Profit (DP) model [24, 25]:

$$E_{\mathrm{DP}} = \frac{1}{N}\sum_{i=1}^{N} f_{\mathrm{DP}}(i)(a_i - o_i)^2. \tag{2.5}$$

Refenes et al. [27] proposed Discounted Least Squares (LDSs) function by taking the contribution from the historical data into accounts as follows:

$$E_{\mathrm{DLS}} = \frac{1}{N}\sum_{i=1}^{N} w_b(i)(a_i - o_i)^2, \tag{2.6}$$

where $w_b(i)$ is an adjustment relating to the contribution of the ith observation and is described by the following equation:

$$w_b(i) = \frac{1}{1 + \exp(b - 2bi/N)}. \tag{2.7}$$

Discount rate b denotes the contribution from the historical data. Refenes et al. [27] suggested $b = 6$.

Yao and Tan [24, 25] proposed another error function, Time Dependent directional Profit (TDP) model, by incorporating the approach suggested by Refenes et al. [27] to their Directional Profit Model (2.5):

$$E_{\mathrm{TDP}} = \frac{1}{N}\sum_{i=1}^{N} f_{\mathrm{TDP}}(i)(a_i - o_i)^2, \tag{2.8}$$

where $f_{\mathrm{TDP}}(i) = f_{\mathrm{DP}}(i) \times w_b(i)$. $f_{\mathrm{DP}}(i)$ and $w_b(i)$ are described by (2.4) and (2.7), respectively.

Note. Refenes et al. [27] and Yao and Tan [24, 25] used $1/2N$ instead of $1/N$ in the formulas given by (2.5), (2.6), and (2.8).

2.2. Modified Error Functions

We are interested in classifying trading signals into three classes: buy, hold, and sell. The hold class includes both positive and negative values (see Criterion A in Section 1). Therefore, the least squares functions, in which the cases with incorrectly predicted directions (positive or negative) are penalised (e.g., the error functions given by (2.5) and (2.8)), will not give the desired prediction accuracy. For example, suppose that $a_i = 0.0045$ and $o_i = -0.0049$. In this case the predicted signal is correct, according to Criterion A. However, the algorithms used in [24, 25] try to minimise error function as $\Delta a_i \times \Delta o_i < 0$ (refer (2.8)). In fact such a minimisation

is not necessary, as the predicted signal is correct. Therefore, instead of the weighing schemes suggested by previous studies, we proposed a different scheme of weighing.

Unlike the weighing schemes suggested in [24, 25], which impose a higher penalty on the predictions whose sign (i.e., negative or positive) is incorrect, this novel scheme is based on the correctness of the classification of trading signals. If the predicted trading signal is correct, we assign a very small (close to zero) weight and, otherwise, assign a weight equal to 1. Therefore, the proposed weighing scheme is

$$w_d(i) = \begin{cases} \delta & \text{if the predicted trading signal is correct,} \\ 1, & \text{otherwise,} \end{cases} \tag{2.9}$$

where δ is a very small value. The value of δ needs to be decided according to the distribution of data.

2.2.1. Proposed Error Function 1

The weighing scheme, $f_{\mathrm{DP}}(i)$, incorporated in the Directional Profit (DP) error function (2.5) considers only two classes, upward and downward trends (direction) which are corresponding to buy and sell signals. In order to deal with three classes, buy, hold, and sell, we modified this error function by replacing $f_{\mathrm{DP}}(i)$ with the new weighing scheme $w_d(i)$ (see (2.9)). Hence, the new error function (E_{CC}) is defined as

$$E_{CC} = \frac{1}{N}\sum_{i=1}^{N} w_d(i)(a_i - o_i)^2. \tag{2.10}$$

When training backpropagation neural networks using (2.10) as the error minimisation function, the error is forced to take a smaller value, if the predicted trading signal is correct. On the other hand, the actual size of the error is considered in the cases of misclassifications.

2.2.2. Proposed Error Function 2

The contribution from the historical data also plays an important role in the prediction accuracy of financial time series. Therefore, Yao and Tan [24, 25] went further by combining DP error function (see (2.5)) with DLS error function (see (2.6)) and proposed Time Dependent Directional Profit (TDP) error function (see (2.8)).

Following Yao and Tan [23, 24], this study also proposed a similar error function, E_{TCC}, by combining first new error function (E_{CC}) described by (2.10) with the DLS error function (E_{DLS}). Hence the second proposed error function is

$$E_{TCC} = \frac{1}{N}\sum_{i=1}^{N} w_b(i) \times w_d(i)(a_i - o_i)^2, \tag{2.11}$$

where $w_b(i)$ and $w_d(i)$ are defined by (2.7) and (2.9), respectively.

The difference between the TDP error function (see (2.8)) and this second new error function (2.11) is that $f_{\mathrm{DP}}(i)$ is replaced by $w_d(i)$ in order to deal with three classes: buy, hold, and sell.

2.3. Modified Neural Network Algorithms

Modifications to neural network algorithms were done by (i) using the OLS error function as well as the modified least squares error functions; (ii) employing a global optimization algorithm to train the networks.

The importance of using global optimization algorithms for the FNN training was discussed in Section 1. In this paper, we applied the global optimization algorithm, AGOP (introduced in [28, 29]), for training the proposed network algorithms.

As the error function to be minimised, we considered E_{OLS} (see (2.1)) and E_{DLS} (see (2.6)) together with the two modified error functions E_{CC} (see (2.10)) and E_{TCC} (see (2.11)). Based on these four error functions, we proposed the following algorithms:

(i) NN_{OLS}—neural network algorithm based on the Ordinary Least Squares error function, E_{OLS} (see (2.1));

(ii) NN_{DLS}—neural network algorithm based on the Discounted Least Squares error function, E_{DLS} (see (2.6));

(iii) NN_{CC}—neural network algorithm based on the newly proposed error function 1, E_{CC} (see (2.10));

(iv) NN_{TCC}—neural network algorithm based on the newly proposed error function 2, E_{TCC} (see (2.11)).

The layers are connected in the same structure as the FNN (Section 2). A tan-sigmoid function was used as the transfer function between the input layer and the hidden layer, while the linear transformation function was employed between the hidden and the output layers.

Algorithm NN_{OLS} differs from the standard FNN algorithm since it employs a new global optimization algorithm for training. Similarly, NN_{DLS} also differs from the respective algorithm used in [24, 25] due to the same reason. In addition to the use of new training algorithm, NN_{CC} and NN_{TCC} are based on two different modified error functions. The only way to examine whether these new modified neural network algorithms perform better than the existing ones (in the literature) is to conduct numerical experiments.

3. Network Training and Evaluation

The Australian All Ordinary Index (AORD) was selected as the stock market index whose trading signals are to be predicted. The previous studies done by the authors [22] suggested that the lagged Close prices of the US S\&P 500 Index (GSPC), the UK FTSE 100 Index (FTSE), French CAC 40 Index (FCHI), and German DAX Index (GDAXI) as well as that of the AORD itself showed an impact on the direction of the Close price of day t of the AORD. Also it was found that only the Close prices at lag 1 of these markets influence the Close price of the AORD [22, 23]. Therefore, this study considered the relative return of the Close prices at lag 1 of two combinations of stock market indices when forming input sets: (i) a combination which includes the GSPC, FTSE, FCHI, and the GDAXI; (ii) a combination which includes the AORD in addition to the markets included in (i).

The input sets were formed with and without incorporating the quantified intermarket influence [22, 23, 30] (see Section 3.1). By quantifying intermarket influence, this study tries to identify the influential patterns between the potential influential markets and the AORD. Training the network algorithms with preidentified patterns may enhance their learning. Therefore, it can be expected that the using quantified intermarket influence for training algorithms produces more accurate output.

The quantification of intermarket influence is described in Section 3.1, while Section 3.2 presents the input sets used for network training.

Daily relative returns of the Close prices of the selected stock market indices from 2nd July 1997 to 30th December 2005 were used for this study. If no trading took place on a particular day, the rate of change of price should be zero. Therefore, before calculating the relative returns, the missing values of the Close price were replaced by the corresponding Close price of the last trading day.

The minimum and the maximum values of the data (relative returns) used for network training are −0.137 and 0.057, respectively. Therefore, we selected the value of δ (see Section 2.2) as 0.01. If the trading signals are correctly predicted, 0.01 is small enough to set the value of the proposed error functions (see (2.10) and (2.11)) to approximately zero.

Since, influential patterns between markets are likely to vary with time [30], the whole study period was divided into a number of moving windows of a fixed length. Overlapping windows of length three trading years were considered (1 trading year ≡ 256 trading days) . A period of three trading years consists of enough data (768 daily relative returns) for neural network experiments. Also the chance that outdated data (which is not relevant for studying current behaviour of the market) being included in the training set is very low.

The most recent 10% of data (the last 76 trading days) in each window were accounted for out of sample predictions, while the remaining 90% of data were allocated for network training. We called the part of the window which allocated for training the *training window*. Different number of neurons for the hidden layer was tested when training the networks with each input set.

As described in Section 2.1, the error function, E_{DLS} (see (2.6)), consists of a parameter b (discount rate) which decides the contribution from the historical data of the observations in the time series. Refenes et al. [27] fixed $b = 6$ for their experiments. However, the discount rate may vary from one stock market index to another. Therefore, this study tested different values for b when training network NN_{DLS}. Observing the results, the best value for b was selected, and this best value was used as b when training network NN_{TCC}.

3.1. Quantification of Intermarket Influences

Past studies [31–33] confirmed that the most of the world's major stock markets are integrated. Hence, one integrated stock market can be considered as a part of a single global system. The influence from one integrated stock market on a dependent market includes the influence from one or more stock markets on the former.

If there is a set of influential markets to a given dependent market, it is not straightforward to separate influence from individual influential markets. Instead of measuring the individual influence from one influential market to a dependent market, the relative strength of the influence from this influential market to the dependent market can be measured compared to the influence from the other influential markets. This study used the approach proposed in [22, 23] to quantify intermarket influences. This approach estimates

the combined influence of a set of influential markets and also the contribution from each influential market to the combined influence.

Quantification of intermarket influences on the AORD was carried out by finding the coefficients, ξ_i, $i = 1, 2, \ldots$ (see Section 3.1.1), which maximise the median rank correlation between the relative return of the Close of day $(t + 1)$ of the AORD market and the sum of ξ_i multiplied by the relative returns of the Close prices of day t of a combination of influential markets over a number of small nonoverlapping windows of a fixed size. The two combinations of markets, which are previously mentioned this section, were considered. ξ_i measures the contribution from the ith influential market to the combined influence which is estimated by the optimal correlation.

There is a possibility that the maximum value leads to a conclusion about a relationship which does not exist in reality. In contrast, the median is more conservative in this respect. Therefore, instead of selecting the maximum of the optimal rank correlation, the median was considered.

Spearman's rank correlation coefficient was used as the rank correlation measure. For two variables X and Y, Spearman's rank correlation coefficient, r_s, can be defined as

$$r_s = \frac{n(n^2-1) - 6\sum {d_i}^2 - (T_x - T_y)/2}{\sqrt{(n(n^2-1) - T_x)(n(n^2-1) - T_Y)}}, \tag{3.1}$$

where n is the total number of bivariate observations of x and y, d_i is the difference between the rank of x and the rank of y in the ith observation, and T_x and T_y are the number of tied observations of X and Y, respectively.

The same six training windows employed for the network training were considered for the quantification of intermarket influence on the AORD. The correlation structure between stock markets also changes with time [31]. Therefore, each moving window was further divided into a number of small windows of length 22 days. 22 days of a stock market time series represent a trading month. Spearman's rank correlation coefficients (see (3.1)) were calculated for these smaller windows within each moving window.

The absolute value of the correlation coefficient was considered when finding the median optimal correlation. This is appropriate as the main concern is the strength rather than the direction of the correlation (i.e., either positively or negatively correlated).

The objective function to be maximised (see Section 3.1.1 given below) is defined by Spearman's correlation coefficient, which uses ranks of data. Therefore, the objective function is discontinuous. Solving such a global optimization problem is extremely difficult because of the unavailability of gradients. We used the same global optimization algorithm, AGOP, which was used for training the proposed algorithms (see Section 2.3) to solve this optimization problem.

3.1.1. Optimization Problem

Let $Y(t + 1)$ be the relative return of the Close price of a selected dependent market at time $t+1$, and let $X_j(t)$ be the relative return of the Close price of the jth influential market at time t. Define $X_\xi(t)$ as

$$X_\xi(t) = \sum_j \xi_j X_j(t), \tag{3.2}$$

where the coefficient $\xi_j \geq 0$, $j = 1, 2, \ldots, m$ measures the strength of influence from each influential market X_j, while m is the total number of influential markets.

The aim is to find the optimal values of the coefficients, $\xi = (\xi_1, \ldots, \xi_m)$, which maximise the rank correlation between $Y(t+1)$ and $X_\xi(t)$ for a given window.

The correlation can be calculated for a window of a given size. This window can be defined as

$$T\left(t^0, l\right) = \left\{t^0,\ t^0 + 1, \ldots,\ t^0 + (l-1)\right\}, \tag{3.3}$$

where t^0 is the starting date of the window, and l is its size (in days). This study sets $l = 22$ days.

Spearman's correlation (see (3.1)) between the variables $Y(t+1)$, $X_\xi(t)$, $t \in T(t^0, l)$, defined on the window $T(t^0, l)$, will be denoted as

$$C(\xi) = \mathrm{Corr}\left(Y(t+1),\ X_\xi(t) \| \ T\left(t^0, l\right)\right). \tag{3.4}$$

To define optimal values of the coefficients for a long time period, the following method is applied. Let $[1, T] = \{1, 2, \ldots, T\}$ be a given period (e.g., a large window). This period is divided into n windows of size l (we assume that $T = l \times n$, $n > 1$ is an integer) as follows:

$$T(t_k, l), \quad k = 1, 2, 3, \ldots, n, \tag{3.5}$$

so that,

$$\begin{gathered} T(t_k, l) \cap T(t_{k'}, l) = \phi \quad \text{for} \quad \forall k \neq k', \\ \bigcup_{k=1}^{n} T(t_k, l) = [1, T]. \end{gathered} \tag{3.6}$$

The correlation coefficient between $Y(t+1)$ and $X_\xi(t)$ defined on the window $T(t_k, l)$ is denoted as

$$C_k(\xi) = \mathrm{Corr}\left(Y(t+1),\ X_\xi(t) \| \ T(t_k, l)\right), \quad k = 1, \ldots, n. \tag{3.7}$$

To define an objective function over the period $[1, T]$, the median of the vector, $(C_1(\xi), \ldots, C_n(\xi))$, is used. Therefore, the optimization problem can be defined as

$$\begin{gathered} \text{Maximise} \quad f(\xi) = \mathrm{Median}(C_1(\xi), \ldots, C_n(\xi)), \\ \text{s. t.} \quad \sum_j \xi_j = 1, \quad \xi_j \geq 0,\ j = 1, 2, \ldots, m. \end{gathered} \tag{3.8}$$

The solution to (3.8) is a vector, $\xi = (\xi_1, \ldots, \xi_m)$, where ξ_j, $j = 1, 2, \ldots, m$ denotes the strength of the influence from the jth influential market.

In this paper, the quantity, $\xi_j X_j$, is called the quantified relative return corresponding to the jth influential market.

3.2. Input Sets

The following six sets of inputs were used to train the modified network algorithms introduced in Section 2.3.

(1) Four input features of the relative returns of the Close prices of day t of the market combination (i) (i.e., GSPC(t), FTSE(t), FCHI(t), and GDAXI(t))—denoted by **GFFG**.

(2) Four input features of the quantified relative returns of the Close prices of day t of the market combination (i) (i.e., ξ_1 GSPC(t), ξ_2 FTSE(t), ξ_3 FCHI(t), and ξ_4 GDAXI(t))—denoted by **GFFG-q**.

(3) Single input feature consists of the sum of the quantified relative returns of the Close prices of day t of the market combination (i) (i.e., ξ_1 GSPC(t) + ξ_2 FTSE(t) + ξ_3 FCHI(t) + ξ_4 GDAXI(t))—denoted by **GFFG-sq**.

(4) Five input features of the relative returns of the Close prices of day t of the market combination (ii) (i.e., GSPC(t), FTSE(t), FCHI(t), GDAXI(t), and AORD(t))—denoted by **GFFGA**.

(5) Five input features of the quantified relative returns of the Close prices of day t of the market combination (ii) (i.e., ξ_1^A GSPC(t), ξ_2^A FTSE(t), ξ_3^A FCHI(t), ξ_4^A GDAXI(t), and ξ_5^A AORD(t))—denoted by **GFFGA-q**.

(6) Single input feature consists of the sum of the quantified relative returns of the Close prices of day t of the market combination (ii) (i.e., ξ_1^A GSPC(t) + ξ_2^A FTSE(t) + ξ_3^A FCHI(t) + ξ_4^A GDAXI(t) + ξ_5^A AORD(t))—denoted by **GFFGA-sq**.

(ξ_1, ξ_2, ξ_3, ξ_4) and (ξ_1^A, ξ_2^A, ξ_3^A, ξ_4^A) are solutions to (3.8) corresponding to the market combinations (i) and (ii), previously mentioned in Section 3. These solutions relating to the market combinations (i) and (ii) are shown in the Tables 1 and 2, respectively. We note that ξ_i and ξ_i^A, $i = 1, 2, 3, 4$ are not necessarily be equal.

3.3. Evaluation Measures

The networks proposed in Section 2.3 output the $(t + 1)$th day relative returns of the Close price of the AORD. Subsequently, the output was classified into trading signals according to Criterion A (see Section 1).

The performance of the networks was evaluated by the overall classification rate (r_{CA}) as well as by the overall misclassification rates (r_{E1} and r_{E2}) which are defined as follows:

$$r_{CA} = \frac{N_0}{N_T} \times 100, \tag{3.9}$$

where N_0 and N_T are the number of test cases with correct predictions and the total number of cases in the test sample, respectively, as follows:

$$\begin{aligned} r_{E1} &= \frac{N_1}{N_T} \times 100, \\ r_{E2} &= \frac{N_2}{N_T} \times 100, \end{aligned} \tag{3.10}$$

Table 1: Optimal values of quantification coefficients (ξ) and the median optimal Spearman's correlations corresponding to market combination (i) for different training windows.

Training window no.	Optimal values of ξ				Optimal median Spearman's correlation
	GSPC	FTSE	FCHI	GDAXI	
1	0.57	0.30	0.11	0.02	0.5782*
2	0.61	0.18	0.08	0.13	0.5478*
3	0.77	0.09	0.13	0.01	0.5680*
4	0.79	0.06	0.15	0.00	0.5790*
5	0.56	0.17	0.03	0.24	0.5904*
6	0.66	0.06	0.08	0.20	0.5359*

* Significant at 5% level

Table 2: Optimal values of quantification coefficients (ξ) and the median optimal Spearman's correlations corresponding to market combination (ii) for different training windows.

Training window no.	Optimal values of ξ					Optimal median Spearman's correlation
	GSPC	FTSE	FCHI	GDAXI	AORD	
1	0.56	0.29	0.10	0.03	0.02	0.5805*
2	0.58	0.11	0.12	0.17	0.02	0.5500*
3	0.74	0.00	0.17	0.02	0.07	0.5697*
4	0.79	0.07	0.14	0.00	0.00	0.5799*
5	0.56	0.17	0.04	0.23	0.00	0.5904*
6	0.66	0.04	0.09	0.20	0.01	0.5368*

* Significant at 5% level

where N_1 is the number of test cases where a buy/sell signal is misclassified as a hold signals or vice versa. N_2 is the test cases where a sell signal is classified as a buy signal and vice versa.

From a trader's point of view, the misclassification of a hold signal as a buy or sell signal is a more serious mistake than misclassifying a buy signal or a sell signal as a hold signal. The reason is in the former case a trader will loses the money by taking part in an unwise investment while in the later case he/she only lose the opportunity of making a profit, but no monetary loss. The most serious monetary loss occurs when a buy signal is misclassified as a sell signal and viceversa. Because of the seriousness of the mistake, r_{E2} plays a more important role in performance evaluation than r_{E1}.

4. Results Obtained from Network Training

As mentioned in Section 3, different values for the discount rate, b, were tested. $b = 1, 2, \ldots, 12$ was considered when training NN_{DLS}. The prediction results improved with the value of b up to 5. For $b > 5$ the prediction results remained unchanged. Therefore, the value of b was fixed at 5. As previously mentioned (see Section 3), $b = 5$ was used as the discount rate also in NN_{TCC} algorithm.

We trained the four neural network algorithms by varying the structure of the network; that is by changing the number of hidden layers as well as the number of neurons per hidden layer. The best four prediction results corresponding to the four networks were obtained when the number of hidden layers equal to one is and, the number of neurons per hidden layer is equal to two (results are shown in Tables 12, 13, 14, 15). Therefore, only the

Table 3: Results obtained from training neural network, NN_{OLS}. The best prediction results are shown in bold colour.

Input set	Average r_{CA}	Average r_{E2}	Average r_{E1}
GFFG	64.25	0.00	35.75
GFFGA	64.25	0.00	35.75
GFFG-q	**64.69**	**0.00**	**35.31**
GFFGA-q	64.04	0.00	35.96
GFFG-sq	63.82	0.00	36.18
GFFGA-sq	63.60	0.00	36.40

Table 4: Results obtained from training neural network, NN_{DLS}. The best prediction results are shown in bold colour.

Input set	Average r_{CA}	Average r_{E2}	Average r_{E1}
GFFG	64.25	0.44	35.31
GFFGA	64.04	0.44	35.53
GFFG-q	64.47	0.22	35.31
GFFGA-q	64.25	0.22	35.53
GFFG-sq	63.82	0.00	36.18
GFFGA-sq	**64.04**	**0.00**	**35.96**

Table 5: Results obtained from training neural network, NN_{CC}. The best prediction results are shown in bold colour.

Input set	Average r_{CA}	Average r_{E2}	Average r_{E1}
GFFG	**65.35**	**0.00**	**34.65**
GFFGA	64.04	0.22	35.75
GFFG-q	63.82	0.00	36.18
GFFGA-q	64.04	0.00	35.96
GFFG-sq	64.25	0.00	35.75
GFFGA-sq	63.82	0.00	36.18

Table 6: Results obtained from training neural network, NN_{TCC}. The best prediction results are shown in bold colour.

Input set	Average r_{CA}	Average r_{E2}	Average r_{E1}
GFFG	66.67	0.44	32.89
GFFGA	64.91	0.22	34.87
GFFG-q	**66.23**	**0.00**	**33.37**
GFFGA-q	63.82	0.22	35.96
GFFG-sq	64.25	0.44	35.31
GFFGA-sq	64.69	0.22	35.09

results relevant to networks with two hidden neurons are presented in this section. Table 3 to Table 6 present the results relating to neural networks, NN_{OLS}, NN_{DLS}, NN_{CC}, and NN_{TCC}, respectively.

The best prediction results from NN_{OLS} were obtained when the input set GFFG-q (see Section 3.2) was used as the input features (see Table 3). This input set consists of four inputs of the quantified relative returns of the Close price of day t of the GSPC and the three European stock indices.

Table 7: Results obtained from training standard FNN algorithms. The best prediction results are shown in bold colour.

Input set	Average r_{CA}	Average r_{E2}	Average r_{E1}
GFFG	62.06	0.22	37.72
GFFGA	62.06	0.22	37.72
GFFG-q	**62.72**	**0.00**	**37.28**
GFFGA-q	**62.72**	**0.00**	**37.28**
GFFG-sq	62.28	0.00	37.72
GFFGA-sq	62.50	0.00	37.50

Table 8: Average (over six windows) classification and misclassification rates of the best prediction results corresponding to NN_{OLS} (trained with input set GFFG-q; refer Table 3).

Actual class	Average classification (misclassification) rates		
	Predicted class		
	Buy	Hold	Sell
Buy	23.46%	(76.54%)	(0.00%)
Hold	(5.00%)	88.74%	(6.27%)
Sell	(0.00%)	(79.79%)	20.21%

Table 9: Average (over six windows) classification and misclassification rates of the best prediction results corresponding to NN_{DLS} (trained with input set GFFGA-sq; refer Table 4).

Actual class	Average classification (misclassification) rates		
	Predicted class		
	Buy	Hold	Sell
Buy	22.10%	(77.90%)	(0.00%)
Hold	(4.97%)	89.20%	(5.83%)
Sell	(0.00%)	(83.06%)	16.94%

NN_{DLS} yielded nonzero values for the more serious classification error, r_{E2}, when the multiple inputs (either quantified or not) were used as the input features (see Table 4). The best results were obtained when the networks were trained with the single input representing the sum of the quantified relative returns of the Close prices of day t of the GSPC, the European market indices, and the AORD (input set GFFGA-sq; see Section 3.2). When the networks were trained with the single inputs (input sets GFFG-sq and GFFGA-sq; see Section 3.2) the serious misclassifications were prevented.

The overall prediction results obtained from the NN_{OLS} seem to be better than those relating to NN_{DLS}, (see Tables 3 and 4).

Compared to the predictions obtained from NN_{DLS}, those relating to NN_{CC} are better (see Tables 4 and 5). In this case the best prediction results were obtained when the relative returns of day t of the GSPC and the three European stock market indices (input set GFFG) were used as the input features (see Table 5). The classification rate was increased by 1.02% compared to that of the best prediction results produced by NN_{OLS} (see Tables 3 and 5).

Table 6 shows that NN_{TCC} also produced serious misclassifications. However, these networks produced high overall classification accuracy and also prevented serious misclassifications when the quantified relative returns of the Close prices of day t of the GSPC and the European stock market indices (input set GFFG-q) were used as the input features.

Table 10: Average (over six windows) classification and misclassification rates of the best prediction results corresponding to NN_{CC} (trained with input set GFFG; refer Table 5).

Actual class	Average classification (misclassification) rates Predicted class		
	Buy	Hold	Sell
Buy	23.94%	(76.06%)	(0.00%)
Hold	(5.00%)	89.59%	(6.66%)
Sell	(0.00%)	(77.71%)	22.29%

Table 11: Average (over six windows) classification and misclassification rates of the best prediction results corresponding to NN_{TCC} (trained with input set GFFG-q; refer Table 6).

Actual class	Average classification (misclassification) rates Predicted class		
	Buy	Hold	Sell
Buy	27.00%	(73.00%)	(0.00%)
Hold	(4.56%)	89.22%	(6.22%)
Sell	(0.00%)	(75.49%)	24.51%

The accuracy was the best among all four types of neural network algorithms considered in this study.

NN_{TCC} provided 1.34% increase in the overall classification rate compared to NN_{CC}. When compared with the NN_{OLS}, NN_{TCC} showed a 2.37% increase in the overall classification rate, and this can be considered as a good improvement in predicting trading signals.

4.1. Comparison of the Performance of Modified Algorithms with that of the Standard FNN Algorithm

Table 7 presents the average (over six windows) classification rates, and misclassification rates related to prediction results obtained by training the standard FNN algorithm which consists of one hidden layer with two neurons. In order to compare the prediction results with those of the modified neural network algorithms, the number of hidden layers was fixed as one, while the number of hidden neurons were fixed as two. These FNNs was trained for the same six windows (see Section 3) with the same six input sets (see Section 3.2). The transfer functions employed are same as those of the modified neural network algorithms (see Section 2.3).

When the overall classification and overall misclassification rates given in Table 7 are compared with the respective rates (see Tables 3 to 6) corresponding to the modified neural network algorithms, it is clear that the standard FNN algorithm shows poorer performance than those of all four modified neural network algorithms. Therefore, it can be suggested that all modified neural network algorithms perform better when predicting the trading signals of the AORD.

4.2. Comparison of the Performance of the Modified Algorithms

The best predictions obtained by each algorithm were compared by using classification and misclassification rates. The classification rate indicates the proportion of correctly classified signals to a particular class out of the total number of actual signals in that class whereas,

Table 12: Results obtained from training neural network, NN_{OLS} with different number of hidden neurons.

Input set	No. of hidden neurons	Average r_{CA}	Average r_{E2}	Average r_{E2}
GFFG	1	64.25	0.00	35.75
	2	64.25	0.00	35.75
	3	64.25	0.00	35.75
	4	64.25	0.22	35.53
	5	64.25	0.00	35.75
	6	64.25	0.00	35.75
GFFGA	1	64.25	0.00	35.75
	2	64.25	0.00	35.75
	3	64.04	0.00	35.96
	4	64.25	0.00	35.75
	5	64.25	0.00	35.75
	6	64.25	0.00	35.75
GFFG-q	1	64.47	0.00	35.53
	2	**64.69**	**0.00**	**35.31**
	3	64.47	0.00	35.53
	4	64.04	0.00	35.96
	5	**64.69**	**0.00**	**35.31**
	6	64.25	0.00	35.75
GFFGA-q	1	64.25	0.00	35.75
	2	64.04	0.00	35.96
	3	63.60	0.22	36.18
	4	64.04	0.00	35.96
	5	64.25	0.00	35.75
	6	63.82	0.00	36.18
GFFG-sq	1	63.82	0.00	36.18
	2	63.82	0.00	36.18
	3	63.82	0.00	36.18
	4	63.82	0.00	36.18
	5	63.82	0.00	36.18
	6	63.82	0.00	36.18
GFFGA-sq	1	63.60	0.00	36.40
	2	63.60	0.00	36.40
	3	63.60	0.00	36.40
	4	63.60	0.00	36.40
	5	63.60	0.00	36.40
	6	63.60	0.00	36.40

the misclassification rate indicates the proportion of incorrectly classified signals from a particular class to another class out of the total number of actual signals in the former class.

4.2.1. Prediction Accuracy

The average (over six windows) classification and misclassification rates related to the best prediction results obtained from NN_{OLS}, NN_{DLS}, NN_{CC}, and NN_{TCC} are shown in Tables 8 to 11, respectively.

Table 13: Results obtained from training neural network, NN_{DLS} with different number of hidden neurons.

Input set	No. of hidden neurons	Average r_{CA}	Average r_{E2}	Average r_{E1}
GFFG	1	64.47	0.44	35.09
	2	64.25	0.44	35.71
	3	64.03	0.44	35.53
	4	64.25	0.44	35.31
	5	64.25	0.44	35.31
	6	64.25	0.44	35.31
GFFGA	1	64.03	0.44	35.53
	2	64.03	0.44	35.53
	3	64.03	0.44	35.53
	4	64.03	0.44	35.53
	5	64.03	0.44	35.53
	6	64.03	0.44	35.53
GFFG-q	1	64.47	0.22	35.31
	2	64.47	0.22	35.31
	3	64.69	0.22	35.09
	4	64.47	0.22	35.31
	5	64.25	0.22	35.53
	6	64.47	0.22	35.31
GFFGA-q	1	64.69	0.22	35.09
	2	64.25	0.22	35.53
	3	63.82	0.22	35.96
	4	64.25	0.44	35.31
	5	64.47	0.44	35.09
	6	64.25	0.22	35.53
GFFG-sq	1	63.82	0.00	36.18
	2	63.82	0.00	36.18
	3	63.82	0.00	36.18
	4	63.82	0.00	36.18
	5	63.82	0.00	36.18
	6	63.82	0.00	36.18
GFFGA-sq	**1**	**64.04**	**0.00**	**35.96**
	2	**64.04**	**0.00**	**35.96**
	3	**64.04**	**0.00**	**35.96**
	4	**64.04**	**0.00**	**35.96**
	5	**64.04**	**0.00**	**35.96**
	6	**64.04**	**0.00**	**35.96**

Among the best networks corresponding to the four algorithms considered, the best network of the algorithm based on the proposed error function 2 (see (2.11)) showed the best classification accuracies relating to buy and sell signals (27% and 25%, resp.; see Tables 8 to 11). Also this network classified more than 89% of the hold signals accurately and it is the second best rate for the hold signal. The rate of misclassification from hold signals to buy is the lowest when this network was used for prediction. The rate of misclassification from hold class to sell class is also comparatively low (6.22%, which is the second lowest among the four best predictions).

Table 14: Results obtained from training neural network, NN_{CC} with different number of hidden neurons.

Input set	No. of hidden neurons	Average r_{CA}	Average r_{E2}	Average r_{E1}
GFFG	1	62.72	0.66	36.62
	2	**65.35**	**0.00**	**34.65**
	3	63.60	0.00	36.40
	4	63.38	0.22	36.40
	5	64.25	0.00	35.75
	6	64.69	0.00	35.31
GFFGA	1	64.04	0.00	35.96
	2	64.03	0.22	35.75
	3	63.16	0.00	36.84
	4	64.04	0.00	35.96
	5	64.03	0.44	35.53
	6	64.04	0.00	35.96
GFFG-q	1	63.38	0.00	36.62
	2	63.82	0.00	36.18
	3	63.60	0.00	36.40
	4	64.91	0.22	34.87
	5	64.03	0.22	35.75
	6	64.69	0.00	35.31
GFFGA-q	1	65.35	0.22	34.43
	2	64.04	0.00	35.96
	3	64.04	0.00	35.96
	4	63.38	0.00	36.62
	5	65.13	0.00	34.87
	6	63.82	0.00	36.18
GFFG-sq	1	64.25	0.00	35.75
	2	64.25	0.00	35.75
	3	64.04	0.00	35.96
	4	64.04	0.00	35.96
	5	64.25	0.00	35.75
	6	64.04	0.00	35.96
GFFGA-sq	1	63.82	0.00	36.18
	2	63.82	0.00	36.18
	3	63.82	0.00	36.18
	4	63.82	0.00	36.18
	5	63.82	0.00	36.18
	6	63.82	0.00	36.18

The network corresponding to the algorithm based on the proposed error function 1 (see (2.10)) produced the second best prediction results. This network accounted for the second best prediction accuracies relating to buy and sell signals while it produced the best predictions relating to hold signals (Table 10).

4.3. Comparisons of Results with Other Similar Studies

Most of the studies [8, 9, 11, 13, 22], which used FNN algorithms for predictions, are aimed at predicting the direction (up or down) of a stock market index. Only a few studies [14, 17],

Table 15: Results obtained from training neural network, NN_{TCC} with different number of hidden neurons.

Input set	No. of hidden neurons	Average r_{CA}	Average r_{E2}	Average r_{E1}
GFFG	1	65.57	0.44	33.99
	2	66.67	0.44	32.89
	3	64.47	0.44	35.09
	4	65.57	0.22	34.21
	5	65.13	0.22	34.65
	6	64.91	0.22	34.87
GFFGA	1	64.69	0.22	35.09
	2	64.91	0.22	34.87
	3	65.13	0.00	34.87
	4	65.13	0.22	34.35
	5	64.13	0.22	34.65
	6	65.57	0.22	34.21
GFFG-q	1	64.91	0.22	34.87
	2	**66.23**	**0.00**	**33.77**
	3	65.57	0.00	34.43
	4	65.79	0.22	33.99
	5	65.13	0.22	34.65
	6	66.23	0.22	33.55
GFFGA-q	1	65.57	0.22	34.21
	2	63.82	0.22	35.96
	3	64.91	0.00	35.09
	4	63.82	0.22	35.96
	5	64.69	0.22	35.09
	6	64.47	0.00	35.53
GFFG-sq	1	65.13	0.44	34.43
	2	64.25	0.44	35.31
	3	64.91	0.44	34.65
	4	64.47	0.44	35.09
	5	64.69	0.44	34.87
	6	64.69	0.44	34.87
GFFGA-sq	1	64.69	0.22	35.09
	2	64.69	0.22	35.09
	3	64.69	0.22	35.09
	4	64.91	0.22	34.87
	5	64.91	0.22	34.87
	6	64.69	0.22	35.09

which used the AORD as the target market index, predicted whether to buy, hold or sell stocks. These studies employed the standard FNN algorithm (that is with OLS error function) for prediction. However, the comparison of results obtained from this study with the above mentioned two studies is impossible as they are not in the same form.

5. Conclusions

The results obtained from the experiments show that the modified neural network algorithms introduced by this study perform better than the standard FNN algorithm in predicting the

trading signals of the AORD. Furthermore, the neural network algorithms, based on the modified OLS error functions introduced by this study (see (2.10) and (2.11)), produced better predictions of trading signals of the AORD. Of these two algorithms, the one-based on (2.11) showed the better performance. This algorithm produced the best predictions when the network consisted of one hidden layer with two neurons. The quantified relative returns of the Close prices of the GSPC and the three European stock market indices were used as the input features. This network prevented serious misclassifications such as misclassification of buy signals to sell signals and viceversa and also predicted trading signals with a higher degree of accuracy.

Also it can be suggested that the quantified intermarket influence on the AORD can be effectively used to predict its trading signals.

The algorithms proposed in this paper can also be used to predict whether it is best to buy, hold, or sell shares of any company listed under a given sector of the Australian Stock Exchange. For this case, the potential influential variables will be the share price indices of the companies listed under the stock of interest.

Furthermore, the approach proposed by this study can be applied to predict trading signals of any other global stock market index. Such a research direction would be very interesting especially in a period of economic recession, as the stock indices of the world's major economies are strongly correlated during such periods.

Another useful research direction can be found in the area of marketing research. That is the modification of the proposed prediction approach to predict whether market share of a certain product goes up or not. In this case market shares of the competitive brands could be considered as the influential variables.

References

[1] B. Egeli, M. Ozturan, and B. Badur, "Stock market prediction using artificial neural networks," in *Proceedings of the 3rd Hawaii International Conference on Business*, pp. 1–8, Honolulu, Hawaii, USA, June 2003.

[2] R. Gençay and T. Stengos, "Moving average rules, volume and the predictability of security returns with feedforward networks," *Journal of Forecasting*, vol. 17, no. 5-6, pp. 401–414, 1998.

[3] M. Qi, "Nonlinear predictability of stock returns using financial and economic variables," *Journal of Business & Economic Statistics*, vol. 17, no. 4, pp. 419–429, 1999.

[4] M. Safer, "A comparison of two data mining techniques to predict abnormal stock market returns," *Intelligent Data Analysis*, vol. 7, no. 1, pp. 3–13, 2003.

[5] L. Cao and F. E. H. Tay, "Financial forecasting using support vector machines," *Neural Computing & Applications*, vol. 10, no. 2, pp. 184–192, 2001.

[6] W. Huang, Y. Nakamori, and S.-Y. Wang, "Forecasting stock market movement direction with support vector machine," *Computers and Operations Research*, vol. 32, no. 10, pp. 2513–2522, 2005.

[7] S. H. Kim and S. H. Chun, "Graded forecasting using an array of bipolar predictions: application of probabilistic neural networks to a stock market index," *International Journal of Forecasting*, vol. 14, no. 3, pp. 323–337, 1998.

[8] H. Pan, C. Tilakaratne, and J. Yearwood, "Predicting Australian stock market index using neural networks exploiting dynamical swings and intermarket influences," *Journal of Research and Practice in Information Technology*, vol. 37, no. 1, pp. 43–54, 2005.

[9] M. Qi and G. S. Maddala, "Economic factors and the stock market: a new perspective," *Journal of Forecasting*, vol. 18, no. 3, pp. 151–166, 1999.

[10] Y. Wu and H. Zhang, "Forward premiums as unbiased predictors of future currency depreciation: a non-parametric analysis," *Journal of International Money and Finance*, vol. 16, no. 4, pp. 609–623, 1997.

[11] J. Yao, C. L. Tan, and H. L. Poh, "Neural networks for technical analysis: a study on KLCI," *International Journal of Theoretical and Applied Finance*, vol. 2, no. 2, pp. 221–241, 1999.

[12] M. T. Leung, H. Daouk, and A.-S. Chen, "Forecasting stock indices: a comparison of classification and level estimation models," *International Journal of Forecasting*, vol. 16, no. 2, pp. 173–190, 2000.

[13] K. Kohara, Y. Fukuhara, and Y. Nakamura, "Selective presentation learning for neural network forecasting of stock markets," *Neural Computing & Applications*, vol. 4, no. 3, pp. 143–148, 1996.

[14] C. D. Tilakaratne, M. A. Mammadov, and S. A. Morris, "Effectiveness of using quantified intermarket influence for predicting trading signals of stock markets," in *Proceedings of the 6th Australasian Data Mining Conference (AusDM '07)*, vol. 70 of *Conferences in Research and Practice in Information Technology*, pp. 167–175, Gold Coast, Australia, December 2007.

[15] R. Akbani, S. Kwek, and N. Japkowicz, "Applying support vector machines to imbalanced datasets," in *Proceedings of the 15th European Conference on Machine Learning (ECML '04)*, pp. 39–50, Springer, Pisa, Italy, September 2004.

[16] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority oversampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[17] C. D. Tilakaratne, S. A. Morris, M. A. Mammadov, and C. P. Hurst, "Predicting stock market index trading signals using neural networks," in *Proceedings of the 14th Annual Global Finance Conference (GFC '07)*, pp. 171–179, Melbourne, Australia, September 2007.

[18] I. Jordanov, "Neural network training and stochastic global optimization," in *Proceedings of the 9th International Conference on Neural Information Processing (ICONIP '02)*, vol. 1, pp. 488–492, Singapore, November 2002.

[19] J. Minghu, Z. Xiaoyan, Y. Baozong, et al., "A fast hybrid algorithm of global optimization for feedforward neural networks," in *Proceedings of the 5th International Conference on Signal Processing (WCCC-ICSP '00)*, vol. 3, pp. 1609–1612, Beijing, China, August 2000.

[20] K. A. Toh, J. Lu, and W. Y. Yau, "Global feedforward neural network learning for classification and regression," in *Proceedings of the 3rd International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR '01)*, pp. 407–422, Shophia Antipolis, France, September 2001.

[21] H. Ye and Z. Lin, "Global optimization of neural network weights using subenergy tunneling function and ripple search," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '03)*, vol. 5, pp. 725–728, Bangkok, Thailand, May 2003.

[22] C. D. Tilakaratne, M. A. Mammadov, and C. P. Hurst, "Quantification of intermarket influence based on the global optimization and its application for stock market prediction," in *Proceedings of the 1st International Workshop on Integrating AI and Data Mining (AIDM '06)*, pp. 42–49, Horbart, Australia, December 2006.

[23] C. D. Tilakaratne, S. A. Morris, M. A. Mammadov, and C. P. Hurst, "Quantification of intermarket influence on the Australian all ordinary index based on optimization techniques," *The ANZIAM Journal*, vol. 48, pp. C104–C118, 2007.

[24] J. Yao and C. L. Tan, "A study on training criteria for financial time series forecasting," in *Proceedings of the International Conference on Neural Information Processing (ICONIP '01)*, pp. 1–5, Shanghai, China, November 2001.

[25] J. Yao and C. L. Tan, "Time dependent directional profit model for financial time series forecasting," in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN '00)*, vol. 5, pp. 291–296, Como, Italy, July 2000.

[26] R. B. Caldwell, "Performances metrics for neural network-based trading system development," *NeuroVe$t Journal*, vol. 3, no. 2, pp. 22–26, 1995.

[27] A. N. Refenes, Y. Bentz, D. W. Bunn, A. N. Burgess, and A. D. Zapranis, "Financial time series modelling with discounted least squares backpropagation," *Neurocomputing*, vol. 14, no. 2, pp. 123–138, 1997.

[28] M. A. Mammadov, "A new global optimization algorithm based on dynamical systems approach," in *Proceedings of the 6th International Conference on Optimization: Techniques and Applications (ICOTA '04)*, A. Rubinov and M. Sniedovich, Eds., Ballarat, Australia, December 2004.

[29] M. Mammadov, A. Rubinov, and J. Yearwood, "Dynamical systems described by relational elasticities with applications," in *Continuous Optimization: Current Trends and Applications*, V. Jeyakumar and A. Rubinov, Eds., vol. 99 of *Applied Optimization*, pp. 365–385, Springer, New York, NY, USA, 2005.

[30] C. D. Tilakaratne, "A study of intermarket influence on the Australian all ordinary index at different time periods," in *Proceedings of the 2nd International Conference for the Australian Business and Behavioural Sciences Association (ABBSA '06)*, Adelaide, Australia, September 2006.

[31] C. Wu and Y.-C. Su, "Dynamic relations among international stock markets," *International Review of Economics & Finance*, vol. 7, no. 1, pp. 63–84, 1998.

[32] J. Yang, M. M. Khan, and L. Pointer, "Increasing integration between the United States and other international stock markets? A recursive cointegration analysis," *Emerging Markets Finance and Trade*, vol. 39, no. 6, pp. 39–53, 2003.

[33] M. Bhattacharyya and A. Banerjee, "Integration of global capital markets: an empirical exploration," *International Journal of Theoretical and Applied Finance*, vol. 7, no. 4, pp. 385–405, 2004.

Hindawi Publishing Corporation
Journal of Applied Mathematics and Decision Sciences
Volume 2009, Article ID 179230, 19 pages
doi:10.1155/2009/179230

Research Article

Selecting the Best Forecasting-Implied Volatility Model Using Genetic Programming

Wafa Abdelmalek,[1] Sana Ben Hamida,[2] and Fathi Abid[1]

[1] *RU: MODESFI, Faculty of Economics and Business, Road of the Airport Km 4, 3018 Sfax, Tunisia*
[2] *Laboratory of Intelligent IT Engineering, Higher School of Technology and Computer Science, 2035 Charguia, Tunisia*

Correspondence should be addressed to Wafa Abdelmalek, wafa.abdelmalek@fsegs.rnu.tn

Received 29 November 2008; Revised 15 April 2009; Accepted 10 June 2009

Recommended by Lean Yu

The volatility is a crucial variable in option pricing and hedging strategies. The aim of this paper is to provide some initial evidence of the empirical relevance of genetic programming to volatility's forecasting. By using real data from S&P500 index options, the genetic programming's ability to forecast Black and Scholes-implied volatility is compared between time series samples and moneyness-time to maturity classes. Total and out-of-sample mean squared errors are used as forecasting's performance measures. Comparisons reveal that the time series model seems to be more accurate in forecasting-implied volatility than moneyness time to maturity models. Overall, results are strongly encouraging and suggest that the genetic programming approach works well in solving financial problems.

1. Introduction

Although the Black-Scholes (BS) formula [1] is commonly used to price derivative securities, it has some well-known deficiencies. When the BS formula is inverted to imply volatilities from reported option prices, the volatility estimates exhibit a dependency on both strike price and maturity, giving rise to what options professionals call "volatility smile" and "volatility term structure." In many markets, prior to the October 1987 stock market crash, there appeared to be symmetry around the zero moneyness, where in-the-money (ITM) or out-of-the-money (OTM) options have higher implied volatilities than at-the-money (ATM) options. This dependency of implied volatility on the strike, for a given maturity, became known as the smile effect, although the exact structure of volatility varied across markets and even within a particular market from day to day. However, after the crash, the smile had a changed shape in many markets, particularly for traded options on stock market indexes

(such as S&P500 index options), where the symmetric smile pattern (U-shape) has changed to more of a sneer. This is often referred to in the markets as the "volatility smirk," where call option-implied volatilities are observed to decrease monotonically with strike price. This can arise when the market places a relatively greater probability on a downward price movement than an upward movement, resulting in a negatively skewed implied terminal asset distribution.

The idea of the volatility smile had its genesis in the early papers documenting the systematic pricing biases of the BS option pricing model. Early tests by Black [2] found that the BS model underprices OTM stock options and overprices ITM stock options. In contrast to Black [2], Macbeth and Merville [3], studying call options, listed on the Chicago Board Options Exchange (CBOE) from December 1975 to December 1976, found evidence that the BS model underprices ITM options and overprices OTM options. Later, Rubinstein [4], studying options price data for the thirty most actively traded option classes on the CBOE between August 1976 and October 1978, found some confusing patterns. Rubinstein [4] reported a systematic mispricing pattern similar to that reported by Macbeth and Merville [3], where the BS model overprices OTM options and underprices ITM options, for a time period between August 1976 and October 1977. However, for a time period between October 1977 and October 1978, Rubinstein [4] reported a systematic mispricing pattern similar to that reported by Black [2], where the BS model underprices OTM options and overprices ITM options. Clewlow and Xu [5], studying options on stock index futures listed on the Chicago Mercantile Exchange (CME), found asymmetric patterns for the smile. Xu and Taylor [6], studying currency options traded in Philadelphia between 1984 and 1992, found empirically that the option bias was twice the size they were expecting, increasing its magnitude as maturity approaches. Duque and Paxson [7] also found the smile effect for options traded on the London International Financial Futures and Options Exchange (LIFFE) during March 1991, and conjectured that there is a possible empirical relation between time to maturity and the U-shape format of the smile. Heynen [8] also found empirical evidence for the exercise price bias when observing stock index options during 1989 on the European Option Exchange (EOE). Gemmill [9] found the same bias for options on the FTSE100 during 5-year period, although the smile showed different patterns for different days extracted from the sample. Dumas et al. [10] also found empirical smile patterns for options on the S&P500 stock index, but its shape seemed to be asymmetric and changing along time to maturity.

Volatility smiles are generally thought to result from the parsimonious assumptions used to derive the BS model. In particular, the BS model assumes that security log prices follow a constant variance diffusion process. The existence of systematic bias in the model reflects a departure from the BS assumptions, with the result that the probability distribution is not lognormal. Black [2] suggested that the nonstationary behavior of volatility would lead the BS model to overprice or underprice options. Since then, a considerable body of research has attempted to explain this bias. Some authors attributed the smile to transaction costs [11]. Some other authors imputed the smile to the behavior of traders [12], others tried to explain the smile by the introduction of conditional heteroskedasticity [13], stochastic interest rate [14], jump processes [15], and stochastic volatility [16]. Literature is not unanimous in finding causes for the smile effect and the models developed in order to cover this bias have only partially solved the problem. Unfortunately, these increasingly complex models have some caveats in common: they require strong assumptions, their parameters are difficult to estimate, and they are particularly prone to in-sample overfitting. Furthermore, their stochastic volatility process itself is unobservable and must be proxied for.

In contrast to parametric forecasting models, nonparametric approach has the distinct advantage of not relying on specific assumptions about the underlying asset price dynamics and is therefore robust to specification errors that might affect adversely parametric models. As mentioned in Ma et al. [17], traditional financial engineering methods based on parametric models such as the GARCH model family seem to have difficulty to improve the accuracy in volatility forecasting due to their rigid as well as linear structure. The complex and nonlinear nature of processes means that more advanced techniques are often needed. Genetic programming (GP) could effectively deal with nonlinearity and flexibility, which opens up an alternative path besides other data-based approaches.

This paper intends to establish a systematic approach and a GP software tool for analysts to improve the accuracy of the forecast. This volatility's forecasting approach should be free of strong assumptions and more flexible than parametric method. Reliable volatility forecasts can greatly benefit professional option traders, market makers who need to price derivatives, and all investors with risk management concerns. Implied volatilities, generated from option markets, can be particularly useful in such contents as they are forward-looking measures of the market's expected volatility during the remaining life of an option. We attempt then to forecast the implied volatility of BS by formulating a nonlinear and nonparametric approach based on GP algorithm.

It is important to notice here that we do not claim to have found the perfect implied volatility's forecasting model. Ultimately, all implied volatility's forecasting models, no matter how complex, are misspecified and the search over the universe of volatility forecasting should concentrate on finding the least misspecified model.

The remainder of the paper is divided into three sections. Section 2 describes the design of the GP and the methodology applied in the current research. A brief introduction to the GP is given in Section 2.1, while the data to conduct the analysis and the GP implementation are described in Section 2.2. Section 3 reports the experimental results of the GP search for volatility forecast. Selection of the best generated GP-implied volatility forecasting models is presented in Section 3.1. Implied volatility patterns are described in Section 3.2. Finally, conclusion and future directions of the research are provided in Section 4.

2. Research Design and Methodology

2.1. Overview of Genetic Programming

GP [18] is an evolutionary algorithm which extends the basic genetic algorithms [19] to process nonlinear problem structure. This optimization technique is based on the principles of natural evolution. The algorithm attempts to evolve solutions by using the principle of survival and reproduction of the fittest and genetic operators analogous to those occurring in biological species. In GP, solutions are represented as tree structures that can vary in size and shape, rather than fixed length character strings as in genetic algorithms. This means that GP can be used to perform optimization at a structural level. The advantage of the GP approach is that it allows one to be agnostic about the general form of optimal forecasting model, and to search efficiently in a nondifferentiable space of models. It also has the attractive feature that one can build into the search procedure the relevant performance criterion directly in the form of the measure of fitness. It differs from most other numerical search techniques in that it simultaneously involves a parallel search involving many points in the search space.

Initialize population
while *(termination condition not satisfied)* **do**
begin
Evaluate the performance of each function according to the fitness criterion
Select the best functions in the population using the selection algorithm
Generate new individuals by crossover and mutation
endwhile
Report the best solution found
end

Algorithm 1: A flow chart summarizing the GP algorithm structure.

Given the advantages of GP, it has been widely used to solve complex high-dimensional and nonconvex optimization problems (Kallel et al. [20]). In fact, the GP has been successfully applied to forecasting financial time series, including stock prices (Tsang et al. [21]), stock returns (Kaboudan [22]), and international currency exchange rates (Kaboudan [23]). In particular, the GP has proved successful at forecasting time series volatility in different markets. Zumbach et al. [24] and Neely and Weller [25] have applied GP to forecast foreign exchange rate volatility. Neely and Weller [25] have tested the forecasting performance of GP for USD-DEM and USD-YEN daily exchange rates against that of GARCH $(1,1)$ model and a related RiskMetrics volatility forecast over different time horizons, using various accuracy criteria. While the GP rules didn't usually match the GARCH $(1,1)$ or RiskMetrics models' MSE or R^2, its performance on those measures was generally close; but the GP did consistently outperform the GARCH model on mean absolute error (MAE) and model error bias at all horizons. Overall, on some dimensions, the GP has produced significantly superior results. Using high-frequency foreign exchange USD-CHF and USD-JPY time series, Zumbach et al. [24] have compared the GP forecasting accuracy to that of historical volatilities, the GARCH $(1,1)$, FIGARCH, and HARCH models. According to the root-mean-squared errors, the generated GP volatility models did consistently outperform the benchmarks.

The GPs forecasting accuracy has been shown also in index market. Using historical returns of Nikkei 225 and S&P500 indices, Chen and Yeh [26] have applied a recursive genetic programming (RGP) approach to estimate volatility by simultaneously detecting and adapting to structural changes. Results have shown that the RGP is a promising tool for the study of structural changes. When RGP discovers structural changes, it will quickly suggest a new class of models so that overestimation of volatility due to ignorance of structural changes can be avoided. Applying a combination of theory and techniques such as wavelet transform, time series data mining, Markov chain-based discrete stochastic optimization, and evolutionary algorithms GA and GP, Ma et al. [27, 28] have proposed a systematic approach to address specifically nonlinearity problems in the forecast of financial indices using intraday data of S&P100 and S&P500 indices. As a result, accuracy of forecasting had reached an average of over 75% surpassing other publicly available results on the forecast of any financial index.

These encouraging findings suggest that a GP may be a powerful tool for forecasting historical volatility based on previous returns. The present paper extends the studies mentioned earlier by forecasting the implied volatility instead of historical volatility using

GP. In fact, implied volatility has been found to be a more accurate and efficient forecast of future volatility than historical volatility [29–32]. Furthermore, the implied volatility is estimated from prices of index options in a wide range of strike prices, not just ATM strikes.

Our GP software is referred to as symbolic regression written in C++ language. Symbolic regression was one of the earliest applications of GP (Koza [18]), and continues to be widely studied (Cai et al. [33]; Gustafson et al. [34]). It is designed to perform genetic regression which is a function that relates a set of inputs to an output. The set data on which the program operates to determine the relationship between input parameters and output is called training set. The set of data on which the resulting formula is tested is called test set. The GP's algorithm includes the following steps (Algorithm 1).

2.1.1. Terminal and Function Sets

Two of the most important aspects of the GP algorithm are the function and terminal sets that building blocks use to construct expression trees which are solutions to the problem. Together, they define the ingredients that are available to GP to create computer programs.

The terminal set corresponds to the inputs of the program. It is determined according to the domain of problems and the inputs can be constants or variables. The function set may contain standard arithmetic operations, standard programming operations, standard mathematical functions, logical functions, or domain-specific functions. The function set combined with the terminal set enables the algorithm to generate a tree-structured solution to the problem, where nodes define function set and where leaf nodes define terminal set.

2.1.2. Initialization

GP starts by randomly creating an initial population of trees, which are generated by randomly picking nodes from a given terminal set and function set. A restriction on the maximum allowed depth is imposed to avoid that the generated trees to be too complex.

2.1.3. Fitness Evaluation

The evolutionary process is driven by a fitness function that evaluates the performance of each individual (tree) in the population. This fitness function assigns numerical values to each individual in the population. These values are then used to select individuals for reproduction.

2.1.4. Selection

After the fitness of each individual in the population is measured, GP then probabilistically selects the fitter individuals from the population to act as the parents of the next generation. Selection determines which individuals of the population will have all or some of their genetic material passed on the next generation. In general, parents displaying a higher level of performance are more likely to be selected with the hope that they can produce better offsprings with larger chance.

2.1.5. Genetic Operators

Crossover and mutation are the two basic operators which are applied to the selected individuals in order to generate the next population. The mutation operator affects random changes in a tree by randomly altering certain nodes or subtrees, whereas the crossover operator performs swaps of randomly subtrees between two selected parents. The parameter choices for crossover and mutation are clearly critical in ensuring a successful GP application. They impact on populational diversity and the ability of GP to escape from local optima [35].

After applying the genetic operators, the new individuals are evaluated through the fitness function in order to replace those of the current population.

2.1.6. Termination Criterion

Once the new population has been created, the current population is replaced by the new one, and the process of fitness evaluation, selection, and the application of genetic operators are repeated until some termination criterion is satisfied. This may be the maximum number of generations or convergence level of evolution generations.

2.2. Data and Implementation

Data used to perform our empirical analysis are daily prices of European S&P500 index calls options, from the CBOE for the sample period January 02, 2003 to August 29, 2003. S&P500 index options are among the most actively traded financial derivatives in the world. The minimum tick for series trading below 3 is 1/16 and for all other series 1/8. Strike price intervals are 5 points, and 25 for far months. The expiration months are three near term months followed by three additional months from the March quarterly cycle (March, June, September, and December). Option prices are set to be the average of the bid and ask prices. The risk-free interest rate is approximated by using 3-month US Treasury bill rates.

The beginning sample contains 42504 daily observations of call option prices and their determinants. Four exclusion filters are applied, which yields a final sample of 6670 daily observations and this is the starting point for our empirical analysis. First, as call options with time to maturity less than 10 days may induce liquidity biases, they are excluded from the sample. Second, call options with low quotes are eliminated to mitigate the impact of price discreteness on option valuation. Third, deep-ITM and deep-OTM option prices are also excluded due to the lack of trading volume. Finally, option prices not satisfying the arbitrage restriction [14], $C \geq S - Ke^{-r\tau}$, are not included.

The full sample is sorted by time series and moneyness-time to maturity. For time series, data are divided into 10 successive samples each containing 667 daily observations. For moneyness-time to maturity, data are divided into 9 classes, classes are, respectively, OTMST (C1), OTMMT (C2), OTMLT (C3), ATMST (C4), ATMMT (C5), ATMLT (C6), ITMST (C7), ITMMT (C8), and ITMLT (C9). According to moneyness criterion: a call option is said OTM if $S/K < 0.98$; ATM if $S/K \in [0.98, 1.03[$; and ITM if $S/K \geq 1.03$. According to time to maturity criterion: A call option is short-term (ST) if $\tau < 60$ days; medium term (MT) if $\tau \in [60, 180]$ days; long-term (LT) if $\tau > 180$ days.

The first preparatory step to run a GP is to specify the ingredients the evolutionary process can use to construct potential solutions. The terminal and function sets are described in Table 1.

Table 1: Terminal set and function set.

Expression	Definition
C/K	Call price/Strike price
S/K	Index price/Strike price
τ	Time to maturity
$+$	Addition
$-$	Subtraction
$*$	Multiplication
%	Protected division: $x\%y = 1$ if $y = 0$; $x\%y = x\%y$ otherwise
ln	Protected natural log: $\ln(x) = \ln(\lvert x\rvert)$
Exp	Exponential function: $\exp(x) = e^x$
Sqrt	Protected square root: $\sqrt{x} = \sqrt{\lvert x\rvert}$
Ncdf	Normal cumulative distribution function Φ

The search space for GP is the space of all possible parse trees that can be recursively created from the sets of terminal and function. The terminal set includes the inputs variables, notably, the call option price divided by strike price C/K, the index price divided by strike price S/K and time to maturity τ. The function set includes unary and binary nodes. Unary nodes consist of mathematical functions, notably, cosinus function (cos), sinus function (sin), log function (ln), exponential function (exp), square root function ($\sqrt{\ }$), and the normal cumulative distribution function (Φ). Binary nodes consist of the four basic mathematical operators, notably, addition ($+$), subtraction ($-$), multiplication ($\times$), and division (%). The basic division operation is protected against division by zero and the log and square root functions are protected against negative arguments. The basic measure of model accuracy used is the mean squared error (MSE) between the target (y_t) and forecasted ($\widehat{y}_t$) output volatility, computed as follows:

$$\text{MSE} = \frac{1}{N}\sum_{t=1}^{N}(y_t - \widehat{y}_t)^2, \tag{2.1}$$

where y_t is the implied volatility computed using the BS formula ($C = SN(d_1) - Ke^{-r\tau}N(d_2)$, $d_1 = (\ln(S/K) + (r + 0.5\sigma^2)\tau)/\sigma\sqrt{\tau}$, $d_2 = d_1 - \sigma\sqrt{\tau}$.), $\widehat{y}_t$ is the generated GP volatility, and N is the number of data sample.

The implementation of genetic programming involves different parameters that determine its efficiency. A common approach in tuning a GP is to undertake a series of trial and error experiments before making parameter choices for the final GP runs. A two-step approach is used. First, the optimal set of genetic parameters is determined based on time series data. By varying parameters for the genetic programs when training them on the first sample data, their performance is tested on a test data set from a later date. Various genetic programs are tested. Each program is run ten times with ten different random seeds. The choice of the best genetic program is made according to the mean and median of mean squared errors (MSEs) for training and testing sets. Second, the optimal set of parameters, relative to the best genetic program selected in the first step, are used for training the genetic program separately on each of the first 9 samples using ten different seeds and testing the program's performance only on the test data from the immediately following date. This limits

Table 2: Summary of GP settings and parameters.

Method of generation:	Ramped half and half
Population size:	100
Offspring size:	200
Generations number:	400
Maximum depth of new individual:	6
Maximum depth of the tree:	17
Selection algorithm:	Tournament selection (4)
Crossover probability:	60%
Mutation probability:	40%
Branch mutation:	20%
Point Mutation:	10%
Expansion mutation:	10%
Replacement strategy:	Comma replacement
Stopping criterion:	Maximum number of generations

the effects of nonstationarities and avoid data snooping. The same genetic parameters and random seeds used for time series data are applied to moneyness-time to maturity data in an attempt to analyze the effect of data' group on the GP performance; each moneyness-time to maturity class is subdivided into training set and test set. Parameters used in this paper are listed in Table 2.

An initial population of 100 individuals is randomly generated using random number generator algorithm with random seeds. The generative method used for initialization is the ramped half and half as detailed in Koza [18]. This method involves generating an equal number of trees using a maximum initial depth that ranges from 2 to 6. For each level of depth, 50% of the initial trees are generated via the full method and the other 50% are generated via the grow method. A maximum size of tree measured by depth is 17. This is a popular number used to limit the size of tree (Koza [18]). It is large enough to accommodate complicated formulas and works in practice. Based on the fitness criterion, the selection of the individuals for reproduction is done with the tournament selection algorithm. A group of individuals is selected from the population with a uniform random probability distribution. The fitness values of each member of this group are compared and the actual best is selected. The size of the group is given by the tournament size which is equal here to 4. The crossover operator is used to generate about 60% of the individuals in the population, while the mutation operator is used to generate about 40% of the population. Different mutation operators are used. Point mutation operator consists of replacing a single node in a tree with another randomly generated node of the same arity. Branch mutation operator randomly selects an internal node in the tree, and then it replaces the subtree rooted at that node with a new randomly generated subtree. Expansion mutation operator randomly selects a terminal node in the tree, and then replaces it with a new randomly-generated subtree. Branch mutation is applied with a rate of 20%; point and expansion mutations are applied with a rate of 10% each. The method of replacing parents for the next generation is comma replacement strategy, which selects the best offspring to replace the parents. Of course, it assumes that offspring size is higher than parents size. The stopping criterion is the maximum number of generations 400.

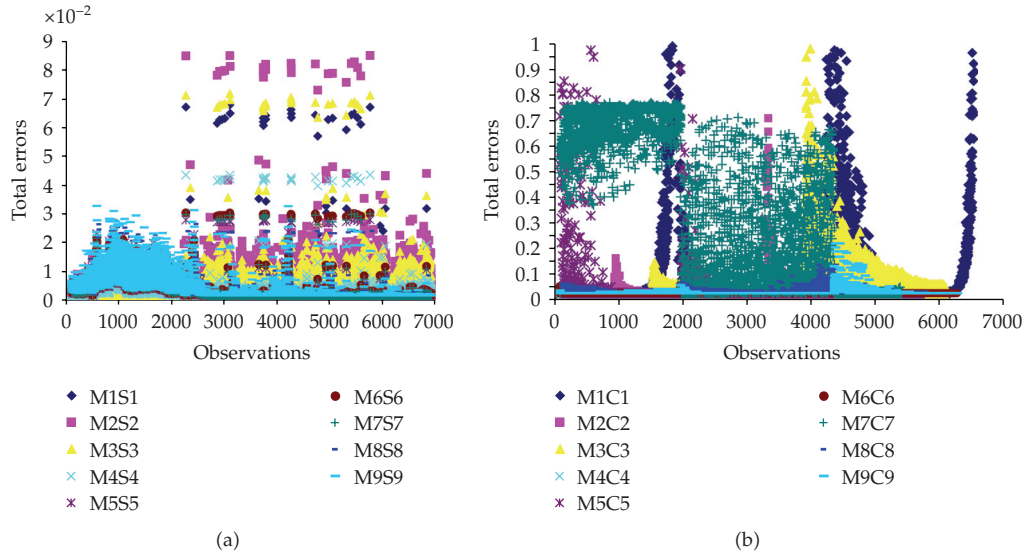


Figure 1: Evolution of the squared errors for total sample of the best generated GP volatility models relative to time series (a) and moneyness-time to maturity (b).

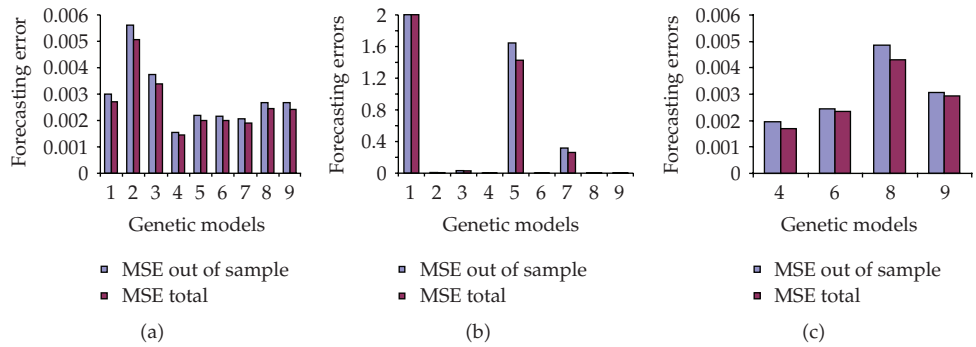


Figure 2: Performance of the generated GP volatility models according to MSE total and out-of-sample for the time series samples (a) and the moneyness-time to maturity classes (b) and (c).

3. Findings and Results Analysis

3.1. *Selection of the Best Genetic Programming-Implied Volatility Forecasting Models*

Selection of the best generated GP volatility model, relative to each training set, for time series and moneyness-time to maturity data, is made according to the training and test MSE. In definitive, nine generated GP volatility models are selected for time series (M1S1$\cdots$M9S9) and similarly nine generated GP volatility models are selected for moneyness-time to maturity classification (M1C1$\cdots$M9C9). To assess the internal and external accuracy of these generated GP volatility models, we use the MSE total computed for the enlarged sample and the MSE out-of-sample computed for external samples to the training one, as performance's

measures. Figure 2 describes the evolution's pattern of the squared errors for these volatility models.

It appears throughout Figure 2 that for the enlarged sample, the nine generated GP volatility models relative to each data group display different patterns of forecasting errors. It is important to note that each model fits well to the data on which it is trained. Furthermore, the performance of the models is not uniform. Total errors are higher for the moneyness-time to maturity classes than for the time series samples. Time series models are adaptive not only to training samples but also to the enlarged sample. In contrast, the moneyness-time to maturity models are adaptive to training classes, but not all to the enlarged sample. The same patterns of forecasting errors are found using out-of-sample errors as performance criterion. Although we have used the same data, the same genetic parameters, and the same random seeds, the GP has generated new models adaptive to the moneyness-time to maturity classes, which are different from the time series models. This implies that the data group has an effect on the GP results. In other words, the GP fits well to the data on which is trained.

Figure 3 shows that, for the time series samples, the generated GP volatility model M4S4 has the smallest MSE in enlarged sample and out-of-sample. For the moneyness-time to maturity classes, the generated GP volatility models M4C4 and M6C6 seem to be more accurate in forecasting implied volatility than the other models. They present near MSE on the enlarged sample and out-of-sample. It is interesting to notice that these models belong, respectively, to ATMST and ATMLT classes. This can be explained by the fact that ATM options are generally more actively traded than other options. Theoretically, Corrado and Miller Jr. [36] have shown that implied volatility from ATM options is the most efficient and also unbiased estimator of the exante underlying volatility when return volatility is stochastic.

Based on the MSE total and the MSE out-of-sample as performance criteria, the generated GP volatility models M4S4, M4C4, and M6C6 are selected.

Table 3 reports the best generated GP volatility models relative to time series samples and moneyness-time to maturity classes and shows that the time series model M4S4 seems to be more performing than moneyness-time to maturity models M4C4 and M6C6 for the enlarged and out-of-sample samples. The decoding of these models yields the following generated GP volatility forecasting formulas:

$$\begin{aligned}
\text{M4S4:}\quad & \sigma_{\text{GP}} = \exp\left[\left(\ln\left(\Phi\left(\frac{C}{K}\right)\right) * \sqrt{\tau - 2 * \frac{C}{K} + \frac{S}{K}}\right) - \cos\left(\frac{C}{K}\right)\right], \\
\text{M4C4}: \quad & \sigma_{\text{GP}} = \sqrt{\left|\frac{C}{K} * \ln\left(\sqrt{\left|\ln\left(\ln\left(\frac{S}{K} - \frac{C}{K}\right) + \left(\frac{S}{K} + \tau\right)\right) + \frac{S}{K}\right)}\right|}, \\
\text{M6C6}: \quad & \sigma_{\text{GP}} = \frac{((C/K)/(S/K)) - \ln(S/K) + (\sin(C/K)/(S/K))}{\Phi\left(\tau * \sqrt{\tau}\right)}.
\end{aligned} \tag{3.1}$$

A detailed examination of the trees shows that the volatility generated by GP is function of all the inputs used, namely, the call option price divided by strike price C/K, the index price divided by strike price S/K and time to maturity τ. The four basic mathematical operators $(+, -, *, /)$, the square root function, and the log function appear in all genetic trees. The normal cumulative distribution function is used only in trees relative to M4S4 and M6C6

Table 3: Selection of the best generated GP volatility models relative to time series and moneyness-time to maturity in terms of the MSE total and the MSE out-of-sample (The numbers in parentheses are standard errors of the corresponding measures of the generated GP volatility models).

Models	MSE total	MSE out-of-sample
M4S4	**0,001444 (0,002727)**	**0,001539 (0,002788)**
M4C4	0,001710 (0,0046241)	0,001944 (0,005108)
M6C6	0,002357 (0,0040963)	0,002458 (0,00415354)

models. The exponential function, which represents the root node, and the cosinus function appear only in the tree relative to M4S4. The sinus function is used only in the tree relative to M6C6.

The implied volatility is defined as the standard deviation which equates the BS model price to the observed market option price. Since there is no explicit formula available to compute directly the implied volatility, the latter can be obtained by inverting the BS model. On the contrary, the GP offers explicit formulas which can compute directly the implied volatility expressed as a function of time to maturity, index, strike, and option prices. Therefore, the generated GP volatilities take account of moneyness and time to maturity. This contradicts the BS assumption of constant volatility. Furthermore, the comparison of the GP forecasting models reveals that the volatility generated by the M6C6 model can be negative, whereas this phenomenon did not exist in the M4S4 and M4C4 models since the implied volatility is computed using the square root and the exponential functions as the root nodes. It appears that even if the variables are nulls, the implied volatility of M4S4 model is always different from zero. This means that a call option can never get a zero value.

3.2. Implied Volatility Patterns: Volatility's Smile, Term Structure, and Surface

Implied volatility patterns of the generated GP volatility models are compared across both moneyness and maturity, as was done in Rubinstein [4]. The call option-implied volatilities illustrated in Figure 4 conform to the general shape hypothesized by Dumas et al. [10] for S&P500 options since the 1987 crash. ITM call options are generally trading at higher implied volatilities than OTM call options. In other words, the implied volatility, at a fixed maturity, is an increasing (decreasing) nonlinear function of moneyness (strike price). The volatility smile is consistent with the market view that the market falls more quickly than it rises. Options writers are concerned with the direction and nature of price movements. Call option writers prefer prices that creep upward and gap downward, while put option writers like the reverse. Option traders aware of a changing volatility may be able to take advantage of the knowledge. For example, if there are indications that the skew may be flattening, one strategy could involve selling at the higher volatility and reversing the position at the lower volatility. Brown [37] has examined the implied volatility smile for call and put options trading on the share price index (SPI) futures contract on the Sydney Futures Exchange (SFE), and offered a possible explanation for its shape. Implied volatilities are viewed as prices reflective of the willingness of market participants to take on and lay off the risks involved in trading volatility, and other risks not priced by the model. Figure 4 shows that the M4S4 model's implied volatility pattern smiles the least for short-term, medium-term, and long-term options. However, the M4C4, M6C6, and BS models still exhibit a significant smile for all

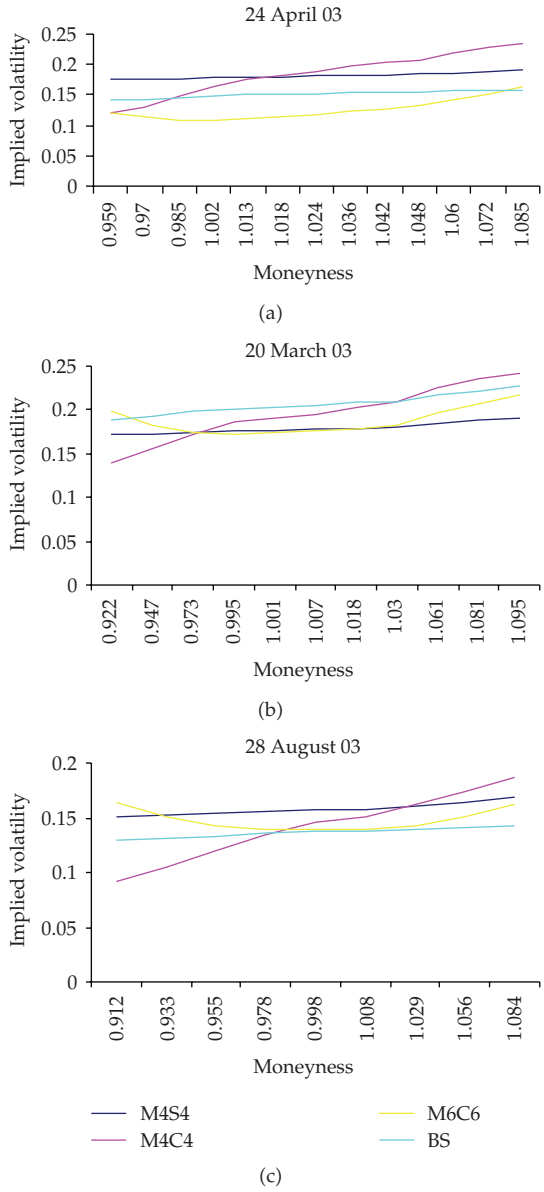


Figure 3: Volatility smile of the generated GP volatility models against BS model for short-term call options (a), medium term call options (b) and long-term call options (c) at different dates.

maturities. The volatility smiles are the strongest for short-term options, indicating that short-term options are the most severely mispriced by these models. The implied volatility models yield, for each given maturity category, the same implied volatility values for some options. For instance, for long-term options, the implied volatility curves of M4C4, M6C6, and BS models intersect all at approximately the same level of moneyness. Options with moneyness of 0.98 (ATM options) are priced at about the same volatility for long-term maturity.

Figure 5 shows that the smile effect is most clear for short-term options. For long-term options, the smile effect is almost vanished. In fact, it is clear that the degree of curvature

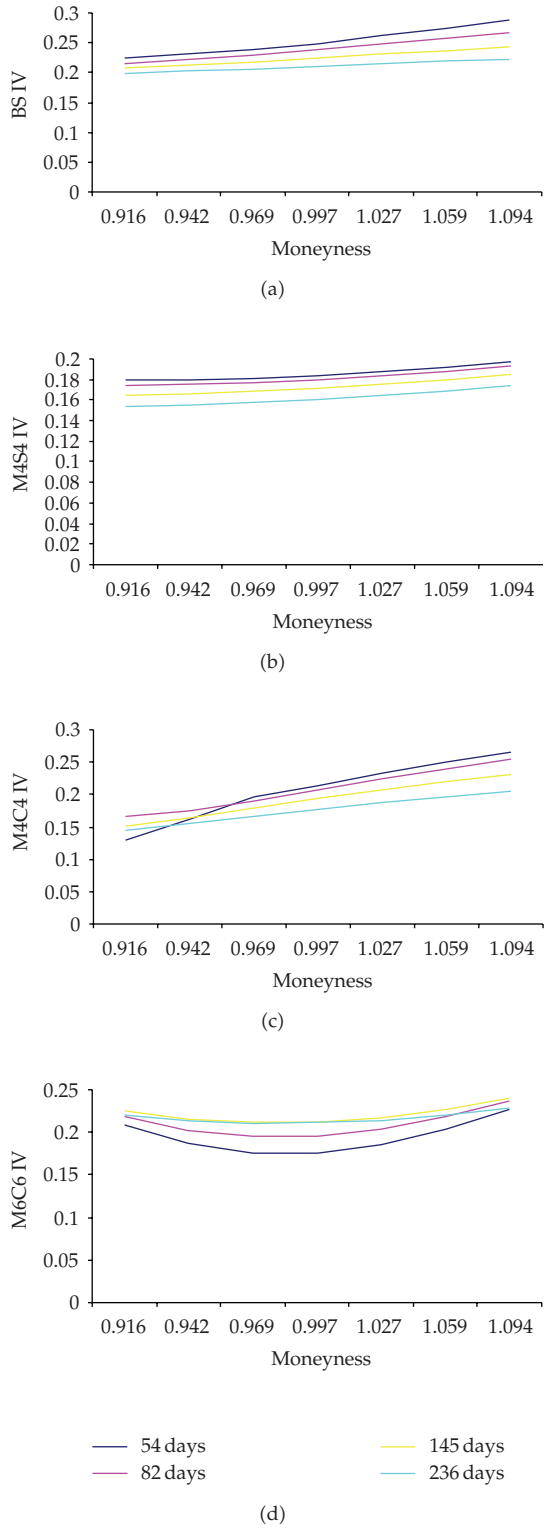


Figure 4: Implied volatility of the generated GP volatility and BS models against degree of moneyness for different times to maturity at the date 27/01/2003.

of M4S4, M4C4, M6C6, and BS models is most extreme when the options are closest to expiration and becomes more linear with longer expirations. This effect was found in [38] for DAX index options from January 1999 to February 2003. It corresponds to the empirical fact that near the expiry, the smile effect becomes stronger. The fact that the implicit volatility smile is more pronounced for short-term options contrary to long-term options is called the vanishing smile. This effect is caused by the smaller time value of short-term options which renders them more sensitive to forecasting errors. As opposed to longer term options which are strongly affected by the value of volatility, short-term options are not much affected by the value of volatility and their price cannot give much information on short-term volatility, which therefore leads to a large uncertainty on short-term volatility. A further result is the changing shape of the smile as the time to maturity increases, except the genetic model M4S4. The smiles of the M4S4 model are the flattest for all maturities, whereas the smiles of the M6C6 model are the steepest for all maturities. The implied volatilities of M4S4, M4C4, and BS models are higher for short-term maturity than for long-term maturity. The M6C6 model has an implied volatility that is higher for long-term maturity than for short-term maturity.

In contrast to the numerous studies relating models to volatility smiles, there has been relatively little research reported that relates model choice to the term structure of volatility (TSV). One reason for this is that the TSV is a far more complex phenomenon. While the smile has the same general shape for all maturities, the TSV can vary widely with the specific choice of strike price. Figure 6 illustrates the volatility term structure of the generated GP volatility models against BS model for all moneyness categories. As Figure 6 shows, the implied volatility of M4S4, M4C4, and BS models is a decreasing function of time to maturity for all moneyness classes except the genetic model M6C6. Rubinstein [4] has shown that for out-of-the money or deep out-of-the money call, the shorter the time to expiration, the higher its implied volatility. The same conclusion was found for at the money call in the second period from 24 October 1977 to 31 October 1978. For ATM call options, all the curves intersect at the same time to maturity, and in particular the curves of M4C4 and BS are very close together. The differences in ATM-implied volatilities across maturities can be attributed to the variation of implied volatility over the full range of traded strikes.

The dynamic properties of implied volatility time series have been studied in many markets by various authors. Most of these studies either focus on the term structure of ATM-implied volatilities or study separately implied volatility smiles for different maturities [39–46]. Other studies focus on the joint dynamics of all implied volatilities quoted on the market, looking simultaneously at all available maturity and moneyness values [38, 47]. In order to investigate the generated GP volatility functions, a three-dimensional graph of implied volatility against moneyness and maturity of the option, called the volatility surface, is plotted in Figure 6. This volatility surface plays an important role in the pricing of options in practice. The implied volatility surface gives a snapshot of market prices of vanilla options: specifying the implied volatility surface is equivalent to specifying prices of all vanilla options quoted on the strike price and expiration date at a given date. While the BS model predicts a flat profile for the implied volatility surface, it is a well documented empirical fact that it exhibits both a nonflat strike and term structure [8, 10, 39, 40, 48, 49].

It is, therefore, interesting to see if the genetic models reproduce realistic instantaneous profiles for the surface. Figure 6 regroups all these candidate solutions on the same graph at the same date. As seen in Figure 6, the best performing volatility functions obtained have quite different surfaces. Volatility varies across the maturity and the moneyness. The implied volatility surfaces appear to display some of the features reported by Derman and Kani [50], and Corrado and Su [51], and there is some degree of negative skewness. Graphs in Figure 6

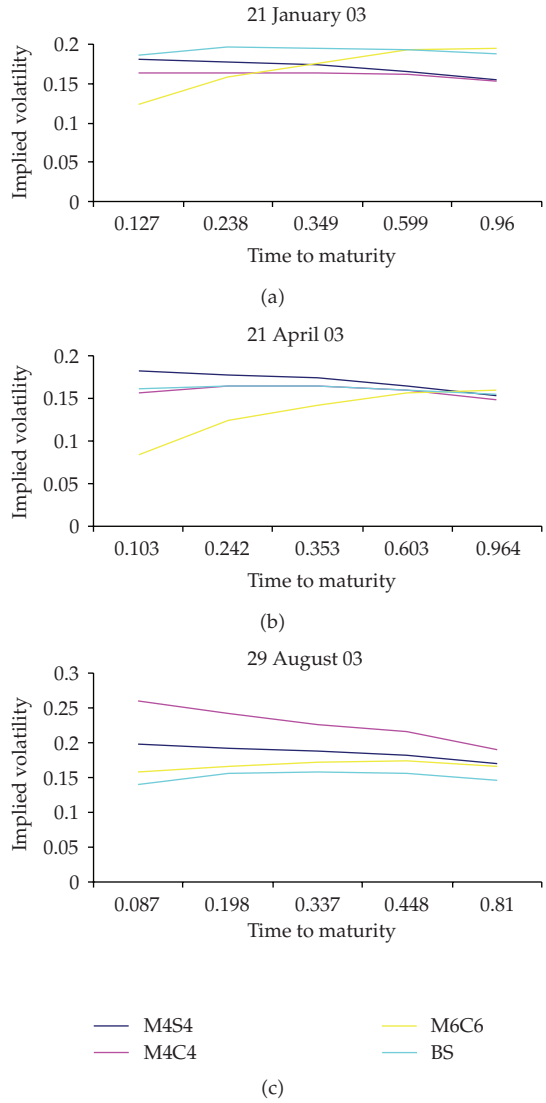


Figure 5: Volatility term structure of the generated GP volatility models against BS model for OTM call options (a), ATM call options (b), and ITM call options (c).

indicate that the volatility smiles are more asymmetric for the M4C4, M6C6, and BS than the M4S4 model which produces quasi-symmetric smile such as stochastic volatility models, the class of parametric models most commonly used to generate smile effects. Options are in fact capturing skewness of expectations especially for the M4C4, M6C6, and BS models. For the S&P500 index, negative skewness in the distribution appearing since the 1987 stock market crash reflects either a higher probability attached to a future crash, or higher valuation of returns in the event of a crash. In a stochastic volatility model, this skewness results from a negative correlation between shocks to price and shocks to volatility, as discussed in Hull and White [52] and Heston [16]. Jackwerth and Rubinstein [53] showed that this negative skewness has been roughly constant in S&P500 options since about one year after 1987 crash.

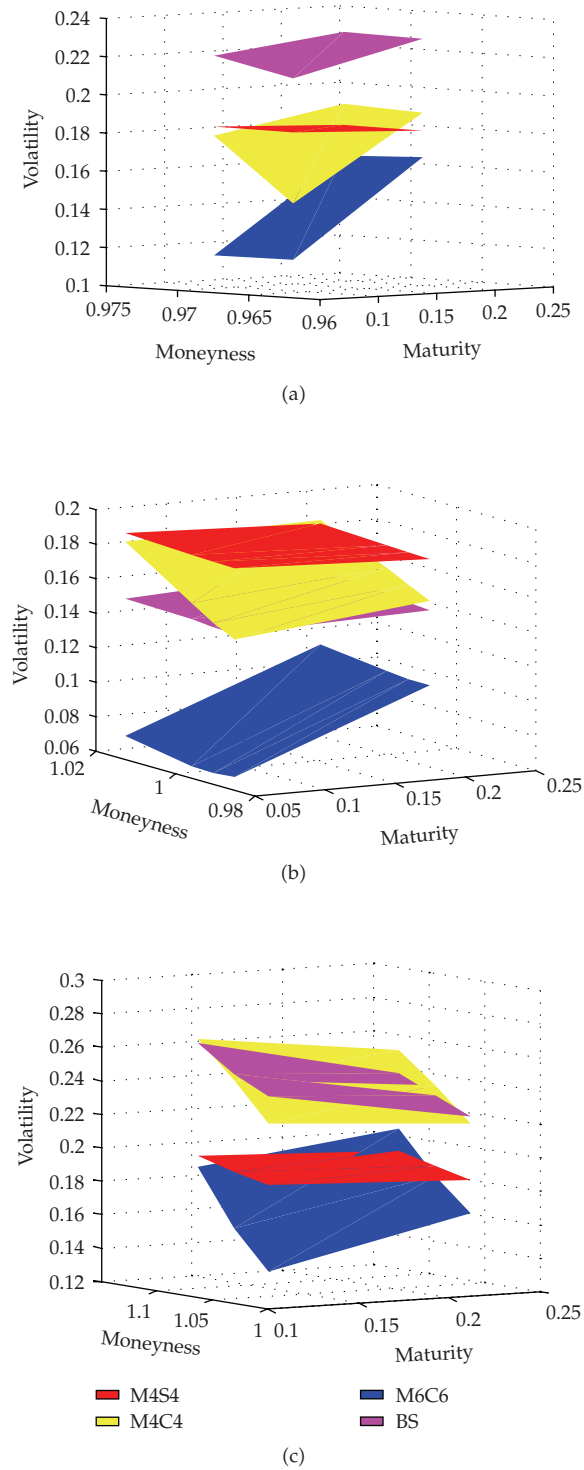


Figure 6: Volatility surfaces of the generated GP volatility models against BS model for short-term OTM call options at 30/01/2003 (a), short-term ATM call options at 30/04/2003 (b), and short-term ITM call options at 20/03/2003 (c).

4. Conclusion

In this study, we have used GP to forecast implied volatility from the S&P500 index options. Based on the MSE total and out-of-sample as performance criteria, the generated GP volatility models M4S4, M4C4, and M6C6 can be considered to the best selected models. According to total and out-of-sample forecasting errors, the time series model M4S4 seems to be more accurate in forecasting-implied volatility than moneyness-time to maturity models M4C4 and M6C6. In terms of volatility patterns, the M4S4 model smiles the least for all maturity options and generates a quasiflat and symmetric implied volatility surface. Our results suggest some interesting issues for further investigation. First, the GP can be used to forecast-implied volatility of other models than BS model, notably stochastic volatility models and models with jump. Second, this work can be reexamined using data from individual stock options, American style index options, options on futures, currency, and commodity options. Third, the performance of the generated GP volatility models can be measured in terms of hedging. Finally, the GP can be extended to allow for dynamic selection techniques applied to the enlarged sample. We believe these extensions are of interest for application and will be object of our future works.

Acknowledgments

The authors are grateful to Giovanni Barone-Adesi for research assistance.

References

[1] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *Journal of Political Economy*, vol. 81, no. 3, pp. 637–654, 1973.
[2] F. Black, "Fact and fantasy in the use of options," *Financial Analysts Journal*, vol. 31, pp. 36–41, 1975.
[3] J. Macbeth and L. Merville, "An empirical examination of the Black and Scholes call option pricing model," *Journal of Finance*, vol. 34, pp. 1173–1186, 1979.
[4] M. Rubinstein, "Non-parametric test of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE option classes from august 23, 1976 through august 31, 1978," *Journal of Finance*, vol. 40, pp. 455–480, 1985.
[5] L. Clewlow and X. Xu, "The dynamics of stochastic volatility," Working Paper, Financial Options Research Centre, Warwick University, 1993.
[6] X. Xu and S. J. Taylor, "The term structure of volatility implied by foreign exchange options," *Journal of Financial and Quantitative Analysis*, vol. 29, no. 1, pp. 57–74, 1994.
[7] J. Duque and D. Paxson, "Implied volatility and dynamic hedging," *The Review of Futures Markets*, vol. 13, no. 2, pp. 381–421, 1994.
[8] R. Heynen, "An empirical investigation of observed smile patterns," *The Review of Futures Markets*, vol. 13, no. 2, pp. 317–353, 1994.
[9] G. Gemmill, "Did option traders anticipate the crash? Evidence from volatility smiles in the U.K. with U.S. comparisons," *The Journal of Futures Markets*, vol. 16, no. 8, pp. 881–897, 1996.
[10] B. Dumas, J. Fleming, and R. E. Whaley, "Implied volatility functions: empirical tests," *Journal of Finance*, vol. 53, no. 6, pp. 2059–2106, 1998.
[11] G. M. Constantinides, "Transaction costs and the volatility implied by option prices," Tech. Rep., University of Chicago, Chicago, Ill, USA, 1998.
[12] R. Bookstaber, *Option Pricing and Investment Strategies*, Probus, Chicago, Ill, USA, 3rd edition, 1991.
[13] R. F. Engle, "Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.
[14] R. C. Merton, "Theory of rational option pricing," *Bell Journal of Economics and Management Science*, vol. 4, no. 1, pp. 141–183, 1973.

[15] R. C. Merton, "Option pricing when underlying stock returns are discontinuous," *Journal of Financial Economics*, vol. 3, no. 1-2, pp. 125–144, 1976.

[16] S. L. Heston, "A closed-form solution for options with stochastic volatility with applications to bond and currency options," *Review of Financial Studies*, vol. 6, pp. 327–343, 1993.

[17] I. Ma, T. Wong, T. Sankar, and R. Siu, "Volatility forecasts of the S&P100 by evolutionary programming in a modified time series data mining framework," in *Proceedings of the World Automation Congress (WAC '04)*, Seville, Spain, June 2004.

[18] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, The MIT Press, Cambridge, Mass, USA, 1992.

[19] J. H. Holland, *Adaptation in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbour, Mich, USA, 1975.

[20] L. Kallel, B. Naudts, and A. Rogers, "Theoretical aspects of evolutionary computing," in *Proceedings of the 2nd EvoNet Summer School Held at the University of Antwerp*, Springer, Berlin, Germany, September 1999.

[21] E. Tsang, P. Yung, and J. Li, "EDDIE-automation, a decision support tool for financial forecasting," *Decision Support Systems*, vol. 37, no. 4, pp. 559–565, 2004.

[22] M. A. Kaboudan, "A measure of time series' predictability using genetic programming applied to stock returns," *Journal of Forecasting*, vol. 18, no. 5, pp. 345–357, 1999.

[23] M. Kaboudan, "Extended daily exchange rates forecasts using wavelet temporal resolutions," *New Mathematics and Natural Computing*, vol. 1, pp. 79–107, 2005.

[24] G. Zumbach, O. V. Pictet, and O. Masutti, "Genetic programming with syntactic restrictions applied to financial volatility forecasting," Working Paper, Olsen & Associates Research Institute, 2001.

[25] C. J. Neely and P. A. Weller, "Predicting exchange rate volatility: genetic programming versus GARCH and riskmetrics," Working Paper, Federal Reserve Bank of St. Louis, May-June 2002.

[26] S. H. Chen and C. H. Yeh, "Using genetic programming to model volatility in financial time series," in *Proceedings of the 2nd Annual Conference on Genetic Programming*, pp. 288–306, Morgan Kaufmann, 1997.

[27] I. Ma, T. Wong, and T. Sanker, "An engineering approach to forecast volatility of financial indices," *International Journal of Computational Intelligence*, vol. 3, no. 1, pp. 23–35, 2006.

[28] I. Ma, T. Wong, and T. Sankar, "Volatility forecasting using time series data mining and evolutionary computation techniques," in *Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation (GECCO '07)*, p. 2262, ACM Press, London, UK, July 2007.

[29] B. J. Christensen and N. R. Prabhala, "The relation between implied and realized volatility," *Journal of Financial Economics*, vol. 50, no. 2, pp. 125–150, 1998.

[30] L. H. Ederington and W. Guan, "Is implied volatility an informationally efficient and effective predictor of future volatility?" NBER Working Paper, University of Oklahoma and Delaware State University, 2000.

[31] B. J. Blair, S.-H. Poon, and S. J. Taylor, "Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns," *Journal of Econometrics*, vol. 105, no. 1, pp. 5–26, 2001.

[32] T. Busch, B. J. Christensen, and M. Ø. Nielsen, "The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets," CREATES Research Paper 2007-9, Aarhus School of Business, University of Copenhagen, 2007.

[33] W. Cai, A. Pacheco-Vega, M. Sen, and K. T. Yang, "Heat transfer correlations by symbolic regression," *International Journal of Heat and Mass Transfer*, vol. 49, no. 23-24, pp. 4352–4359, 2006.

[34] S. Gustafson, E. K. Burke, and N. Krasnogor, "On improving genetic programming for symbolic regression," in *Proceedings of IEEE Congress on Evolutionary Computation (CEC '05)*, D. Corne, et al., Ed., vol. 1, pp. 912–919, IEEE Press, Edinburgh, UK, 2005.

[35] Z. Yin, A. Brabazon, C. O'Sullivan, and M. O'Neil, "Genetic programming for dynamic environments," in *Proceedings of the International Multiconference on Computer Science and Information Technology*, pp. 437–446, 2007.

[36] C. J. Corrado and T. W. Miller Jr., "Efficient option-implied volatility estimators," *The Journal of Futures Markets*, vol. 16, no. 3, pp. 247–272, 1996.

[37] C. A. Brown, "The volatility structure implied by options on the SPI futures contract," *Australian Journal of Management*, vol. 24, no. 2, pp. 115–130, 1999.

[38] S. Borak, M. Fengler, and W. Härdle, "DSFM fitting of implied volatility surfaces. SFB 649," Discussion Paper 2005-022, Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Berlin, Germany, 2005.
[39] R. Heynen, A. Kemma, and T. Vorst, "Analysis of the term structure of implied volatilities," *Journal of Financial and Quantitative Analysis*, vol. 29, no. 1, pp. 31–56, 1994.
[40] S. R. Das and R. K. Sundaram, "Of smiles and smirks: a term structure perspective," *Journal of Financial and Quantitative Analysis*, vol. 34, no. 2, pp. 211–239, 1999.
[41] G. Skiadopoulos, S. Hodges, and L. Clewlow, "The dynamics of the S&P 500 implied volatility surface," *Review of Derivatives Research*, vol. 3, no. 3, pp. 263–282, 2000.
[42] Y. Zhu and M. Avellaneda, "An E-ARCH model for the term structure of implied volatility of FX options," *Applied Mathematical Finance*, vol. 4, no. 2, pp. 81–100, 1997.
[43] R. Hafner and M. Wallmeier, "The dynamics of DAX implied volatilities," *Quarterly International Journal of Finance*, vol. 1, pp. 1–27, 2001.
[44] M. R. Fengler, W. K. Härdle, and C. Villa, "The dynamics of implied volatilities: a common principal components approach," *Review of Derivatives Research*, vol. 6, no. 3, pp. 179–202, 2003.
[45] R. W. Lee, "Implied volatility: statistics, dynamics, and probabilistic interpretation," in *Recent Advances in Applied Probability*, pp. 1–27, Springer, New York, NY, USA, 2004.
[46] J.-P. Fouque, G. Papanicolaou, R. Sircar, and K. Solna, "Maturity cycles in implied volatility," *Finance and Stochastics*, vol. 8, no. 4, pp. 451–477, 2004.
[47] R. Cont and J. da Fonseca, "Dynamics of implied volatility surfaces," *Quantitative Finance*, vol. 2, no. 1, pp. 45–60, 2002.
[48] R. Cont and J. da Fonseca, "Deformation of implied volatility surfaces: an empirical analysis," in *Empirical Science to Financial Fluctuations*, H. Takayasu, Ed., pp. 230–239, Springer, Tokyo, Japan, 2001.
[49] R.-M. Gaspar, "Implied volatility and forward price term structures," ADVANCES Research Center, ISEG, Technical University Lisbon, Lisbon, Portugal, November 2008.
[50] E. Derman and I. Kani, "Riding on a smile," *Risk*, vol. 7, no. 2, pp. 32–39, 1994.
[51] C. Corrado and T. Su, "Skewness and kurtosis in S&P500 index returns implied by option prices," *The Journal of Financial Research*, vol. 19, no. 2, pp. 175–192, 1996.
[52] J. Hull and A. White, "The pricing of options on assets with stochastic variables," *Journal of Finance*, vol. 42, pp. 281–300, 1987.
[53] J. C. Jackwerth and M. Rubinstein, "Recovering probability distributions from option prices," *Journal of Finance*, vol. 51, no. 5, pp. 1611–1631, 1996.

Hindawi Publishing Corporation
Journal of Applied Mathematics and Decision Sciences
Volume 2009, Article ID 609196, 9 pages
doi:10.1155/2009/609196

Research Article

Discrete Analysis of Portfolio Selection with Optimal Stopping Time

Jianfeng Liang

Department of Risk Management and Insurance, Lingnan (University) College, Sun Yat-Sen University, Guangzhou 510275, China

Correspondence should be addressed to Jianfeng Liang, jfliang@mail.sysu.edu.cn

Received 28 November 2008; Accepted 5 March 2009

Recommended by Lean Yu

Most of the investments in practice are carried out without certain horizons. There are many factors to drive investment to a stop. In this paper, we consider a portfolio selection policy with market-related stopping time. Particularly, we assume that the investor exits the market once his wealth reaches a given investment target or falls below a bankruptcy threshold. Our objective is to minimize the expected time when the investment target is obtained, at the same time, we guarantee the probability that bankruptcy happens is no larger than a given level. We formulate the problem as a mix integer linear programming model and make analysis of the model by using a numerical example.

1. Introduction

Portfolio theory deals with the question of how to find an optimal policy to invest among various assets. The mean-variance analysis of Markowitz [1, 2] plays a key role in the theory of portfolio selection, which quantifies the return and the risk in computable terms. The mean-variance model is later extended to the multistage dynamic case. For this and other expected utility-maximization models in dynamic portfolio selection, one is referred to Dumas and Luciano [3], Elton and Gruber [4], Li and Ng [5], Merton [6], and Mossion [7].

An important assumption of the previous portfolio selection model is that the investment horizon is definite. That means an investor knows with certainty the exit time at the beginning of the investment. However, most of the investments in practice are carried out without certain horizons. There are many factors, related to the market or not, which can drive the investment stop. For example, sudden huge consumption, serious illness, and retirement are market-unrelated reasons. Also, those market-related reasons may more strongly affect the investment horizon. A natural example is that the investor may exit

the market once his wealth reaches an investment target, which is closely related to the market and also the investment policy itself. Because of the disparity between theory and practice, it seems sympathetic to relax the restrictive assumption that the investment horizon is preknown with certainty.

Research on this subject has been investigated in continuous setting. Yaari [8] first deals with the problem of optimal consumption for an individual with uncertain date of death, under a pure deterministic investment environment. In 2000, Karatzas and Wang [9] address the optimal dynamic investment problem in a complete market with assumption that the uncertain investment horizon is a stopping time of asset price filtration. Multiperiod mean-variance portfolio optimization problem with uncertain exit time is studied by Guo and Hu [10], where the uncertain exit time is market unrelated. A continuous time problem with minimizing the expected time to beat a benchmark is addressed in Browne [11, 12], where the exit time is a random variable related to the portfolio. Literatures of portfolio selection focus on the case that the stopping time is market-state independent. While, the state-dependent exogenous stopping time is considered by Blanchet-Scalliet et al. [13] in dynamic asset pricing theory.

In this paper, we consider a portfolio selection problem with endogenous stopping time in discrete framework, which has not been well discussed in literatures. Specially, we assume that the investor exits the market once his wealth hits an investment target or he is bankrupt. This assumption actually reflects most investors' investment behavior in real life. Our objective is to minimize the expected time that the investment target is obtained, at the same time we guarantee that the probability of which bankruptcy happens is no larger than a given threshold. The investment process is represented by a multistage scenarios tree, in which the discrete stages and notes denote the decision time points and the market states, respectively.

The rest part of the paper is organized as follows. In Section 2, we introduce the statement of the problem, including notations and the general form of the problem. Following, Section 3 is devoted to derive the deterministic formulation of the problem, in which we define a list of integer variables to indicate different states during the investment process. Finally, we make analysis of the model by using a numerical example in Section 4.

2. The Problem Statement

Consider the following investment problem. We distribute the investment budget among a set of assets, and the portfolio can be adjusted at several discrete decision time points during the investment process. At the beginning of the investment, we assign a target wealth and also a bankruptcy threshold. Our objective is to obtain this target wealth and stop the investment as soon as possible. At the same time, we also need to avoid that the bankruptcy occurs before the target wealth obtained.

The problem is based on a finite multistage scenarios tree structure. Our portfolio includes a set of m assets. The underlying dynamic scenarios tree is constructed as follows. There are T stages denoted from time 0 to T. The portfolios can be constructed at the beginning of each stage in the scenarios tree. We denote N_t to be the index set of the scenarios at time t, and S_{nt} as the nth scenario at time t, for $n \in N_t$, $t = 0, 1, \ldots, T$. For those data at this scenario, the price vector of the risky assets is denoted by $u_{nt} \in \Re^m$, and the payoff vector of the assets is denoted by $v_{nt} \in \Re^m$. The decision variables at this scenario are the number of shares of holdings of the assets $x_{nt} \in \Re^m$. We denote the wealth at this scenario to be W_{nt}, and

the initial wealth to be B. We denote $a(n)$ and $c(n)$ as the parent node and the children nodes of node n, respectively. Moreover, let $S_{a(n),t-1}$ be the parent scenario of S_{nt}, and $S_{c(n),t+1}$ be the set of immediate children of S_{nt}. The probability of scenario S_{nt} happens is p_{nt}.

We consider an objective related to the achievement of performance goal and bankruptcy. The investment stops once the goal is reached or the bankruptcy occurs, the related stopping time is denoted as t_u and t_l, respectively. Specifically, for given wealth levels l and u, with $l < B < u$, we say that the performance goal u is reached if $W_{nt} \geq u$, denoting this time as t_u, that is, $t_u = \inf\{t > 0; W_{nt} \geq u\}$; that the bankruptcy occurs if $W_{nt} < l$, denoting this time as t_l, that is, $t_l = \inf\{t > 0; W_{nt} < l\}$. Our objective is to minimize the expected time that the goal is reached, at the same time we guarantee the probability that the bankruptcy happens before the goal is reached is no more than a given level, say q, $0 < q < 1$. Thus, the investment problem can be represented in the general form

$$\begin{aligned} &\min\ E[t_u] \\ &\text{s.t. } P(W < l) \leq q \\ &\quad \text{budget constraints} \\ &\quad t \in \{0, 1, 2, \ldots, T\}, \end{aligned} \tag{M}$$

where the first constraint is a probability constraint of bankruptcy, in which W generally represents the realized wealth by investment. Moreover, the budget constraints are the wealth dynamics during the investment horizon. We will continue the discussion on the determistic formulation of the model in Section 3.

3. The Problem Formulation

In this section, we will derive the deterministic formulation of the problem (M). Most efforts are devoted to present the objective function and the probability constraint. Actually, we do this by introducing a list of indicator variables. Before we start this work, let us first consider the budget constraints first.

3.1. The Budget Constraints

Based on the previously given notations on the scenarios tree, we first have the allocation of the initial investment wealth represented as $B = u_0' x_0$. At scenario S_{nt}, $n \in N_t$, $t = 0, 1, \ldots, T$, the wealth W_{nt} should be the realized payoff during the previous period, that is,

$$W_{nt} = v_{nt}' x_{a(n),t-1}. \tag{3.1}$$

Also, for a self-financing process that we are considering here, the realized wealth will be reinvested at this decision point, which means

$$W_{nt} = u_{nt}' x_{nt}. \tag{3.2}$$

Therefore, we conclude the budget constraints at scenario $S_{nt},\ n \in N_t,\ t = 0, 1, \ldots, T$, by the set of equations as follows:

$$\begin{aligned} u_0' x_0 &= B, \\ v_{nt}' x_{a(n),t-1} &= u_{nt}' x_{nt}, \quad n \in N_t,\ t = 1, 2, \ldots, T. \end{aligned} \tag{3.3}$$

3.2. The Objective and the Probability Constraint

We come to the formulation of the objective function and the probability constraint. Let us consider the investment process. There are basically three different outputs at a given scenario S_{nt}. The first one is that we succeed to obtain the target wealth and stop the investment on this scenario, and this is really the objective. The second one is that we unfortunately fall into bankruptcy on this scenario and have to exit the investment. In either case, we cannot restart it again. In addition to the above two cases, the investment may be continued to next period.

Now, we define two 0-1 variables to describe the investment story. On scenario $S_{nt},\ n \in N_t,\ t = 0, 1, \ldots, T$, first define $\varepsilon_{nt} \in \{0, 1\}$ such that

$$\varepsilon_{nt} = \begin{cases} 1, & W_{nt} \geq u,\ 1 \leq W_{a(n),j} < u,\ \forall j < t, \\ 0, & \text{otherwise.} \end{cases} \tag{3.4}$$

Parallel to ε_{nt}, we define $\eta_{nt} \in \{0, 1\}$ such that

$$\eta_{nt} = \begin{cases} 1, & W_{nt} < l,\ l \leq W_{a(n),j} < u,\ \forall j < t, \\ 0, & \text{otherwise.} \end{cases} \tag{3.5}$$

Reading the definitions, $\varepsilon_{nt} = 1$ indicates the first case, where the investment reaches the target and stops at scenario S_{nt}, and $\eta_{nt} = 1$ represents the second case that bankruptcy happens at scenario S_{nt}. By using ε_{nt} and η_{nt}, we can write our objective as

$$E[t_u] = \sum_{t=0}^{T} \left(t \cdot \sum_{n \in N_t} p_{nt} \varepsilon_{nt} \right), \tag{3.6}$$

and the deterministic form of the probability as

$$P(W < l) = \sum_{t=0}^{T} \sum_{n \in N_t} p_{nt} \eta_{nt} \leq q. \tag{3.7}$$

We consider again the indicator variables ε_{nt} and η_{nt}. Their values on scenario S_{nt} actually depend on both the current state and also all of the ancestor states. Take ε_{nt} as an example, $\varepsilon_{nt} = 1$ holds if and only if the following two conditions are both satisfied. One is that the investment continues to the current scenario, and the other is that the payoff at the current scenario is no less than the target wealth. If either of the above conditions is not

achieved, we should get $\varepsilon_{nt} = 0$. Moreover, the case of η_{nt} is for the same logic but about the bankruptcy part. Thus, we introduce another two sets of variables to track the current state and the historical states separately.

For the current state, we define $\delta_{nt}, \xi_{nt} \in \{0,1\}$ as follows:

$$\delta_{nt} = \begin{cases} 1, & W_{nt} \geq u, \\ 0, & W_{nt} < u, \end{cases}$$
$$\xi_{nt} = \begin{cases} 1, & W_{nt} < l, \\ 0, & W_{nt} \geq l, \end{cases} \tag{3.8}$$

and for the ancestor states, we define $\phi_{nt} \in \{0,1\}$ such that

$$\phi_{nt} = \begin{cases} 1, & l \leq W_{a(n),j} < u, \ \forall j < t, \\ 0, & \text{otherwise}, \end{cases} \tag{3.9}$$

where $\phi_{nt} = 1$ means that the investment has kept going on to the current scenario and $\phi_{nt} = 0$ means that it has stopped on the parent scenario or other ancestor scenarios before.

Combine the above definitions and review ε_{nt} and η_{nt}, we realize the relations

$$\varepsilon_{nt} = \delta_{nt} \cdot \phi_{nt},$$
$$\eta_{nt} = \xi_{nt} \cdot \phi_{nt}. \tag{3.10}$$

If we replace these nonlinear constraints by a set of linear constraints, then the problem can be hopefully formulated as a linear programming problem, which will benefit for the further research on solution methods and applications. Since the indicator variables are all defined as binary 0-1 variables, we derive the transformation

$$\varepsilon_{nt} = \delta_{nt} \cdot \phi_{nt} \Longleftrightarrow \begin{cases} \delta_{nt} + \phi_{nt} - \varepsilon_{nt} \leq 1, \\ \delta_{nt} + \phi_{nt} - 2\varepsilon_{nt} \geq 0. \end{cases} \tag{3.11}$$

It is direct to check that for given values of $\{\delta_{nt}, \phi_{nt}\}$, ε_{nt} must realize the same value either by $\varepsilon_{nt} = \delta_{nt} \cdot \phi_{nt}$ or by the constraints

$$\delta_{nt} + \phi_{nt} - \varepsilon_{nt} \leq 1,$$
$$\delta_{nt} + \phi_{nt} - 2\varepsilon_{nt} \geq 0, \tag{3.12}$$

and similar case for η_{nt},

$$\eta_{nt} = \xi_{nt} \cdot \phi_{nt} \Longleftrightarrow \begin{cases} \xi_{nt} + \phi_{nt} - \eta_{nt} \leq 1, \\ \xi_{nt} + \phi_{nt} - 2\eta_{nt} \geq 0. \end{cases} \tag{3.13}$$

Therefore, we now replace (3.10) by the following set of inequalities:

$$\begin{aligned}
\delta_{nt} + \phi_{nt} - \varepsilon_{nt} &\leq 1, \\
\delta_{nt} + \phi_{nt} - 2\varepsilon_{nt} &\geq 0, \\
\xi_{nt} + \phi_{nt} - \eta_{nt} &\leq 1, \\
\xi_{nt} + \phi_{nt} - 2\eta_{nt} &\geq 0.
\end{aligned} \tag{3.14}$$

Up to now, we have almost derived out the formulation of the model based on a series of indicator variables, including ε, η, δ, ξ, ϕ. The remaining task is to construct the dynamics of ϕ_{nt} and also the constraints of δ_{nt} and ξ_{nt}, so that the definitions here can be implemented in the model.

3.3. The Dynamics of Indicator Variables

Consider the constraints of δ_{nt} and ξ_{nt} first. Given a large enough number $M_1 > u$ and a small enough number $M_2 < l$, we have for δ_{nt},

$$\begin{aligned}
\delta_{nt} &= \begin{cases} 1, & W_{nt} \geq u, \\ 0, & W_{nt} < u, \end{cases} \\
&\Longleftrightarrow \begin{cases} W_{nt} - (M_1 - u) \cdot \delta_{nt} < u, \\ W_{nt} + (u - M_2) \cdot (1 - \delta_{nt}) \geq u \end{cases}
\end{aligned} \tag{3.15}$$

and for ξ_{nt}, we have

$$\begin{aligned}
\xi_{nt} &= \begin{cases} 1, & W_{nt} < l, \\ 0, & W_{nt} \geq l, \end{cases} \\
&\Longleftrightarrow \begin{cases} W_{nt} - (M_1 - l) \cdot (1 - \xi_{nt}) < l, \\ W_{nt} + (l - M_2) \cdot \xi_{nt} \geq l. \end{cases}
\end{aligned} \tag{3.16}$$

We combine the constraints of δ_{nt} and ξ_{nt} as the constraint set

$$\begin{aligned}
W_{nt} - (M_1 - u) \cdot \delta_{nt} &< u, \\
W_{nt} + (u - M_2)(1 - \delta_{nt}) &\geq u, \\
W_{nt} - (M_1 - l)(1 - \xi_{nt}) &< l, \\
W_{nt} + (l - M_2) \cdot \xi_{nt} &\geq l.
\end{aligned} \tag{3.17}$$

Next, let us focus on the dynamics of ϕ_{nt}. At the beginning point of the investment, $\phi_0 = 1$ holds. During the investment process, we first write out the dynamics and then explain the underlying reasons:

$$\begin{aligned} \phi_0 &= 1, \\ \phi_{nt} &= \phi_{a(n),t-1} - \left(\varepsilon_{a(n),t-1} + \eta_{a(n),t-1}\right). \end{aligned} \tag{3.18}$$

The dynamic equation holds for the following reasons.

First, suppose the investment has been continued to the scenario $S_{a(n),t-1}$ and does not stop at that scenario, which means we already held $\phi_{a(n),t-1} = 1$, and $\varepsilon_{a(n),t-1} = \eta_{a(n),t-1} = 0$, then, the investment must keep going on to the current scenario S_{nt}. In this case, we should have $\phi_{nt} = 1$ base on the definition of ϕ. The recursive equation in (3.18) succeeds to realize this case and gives $\phi_{nt} = 1 - 0 = 1$.

Second, if the investment has stopped, either on the parent scenario $S_{a(n),t-1}$ or on any of the ancestor scenarios before, we should hold $\phi_{nt} = 0$. This case can also be realized by the dynamic equation (3.18). In case that the investment stopped on the parent scenario $S_{a(n),t-1}$, that is, $\phi_{a(n),t-1} = 1$, and either $\varepsilon_{a(n),t-1} = 1$ or $\eta_{a(n),t-1} = 1$, then (3.18) gives $\phi_{nt} = 0$; in the other case of stopping before the previous stage, we already had $\phi_{a(n),t-1} = 0$, also both $\varepsilon_{a(n),t-1} = 0$ and $\eta_{a(n),t-1} = 0$, the result of (3.18) is still $\phi_{nt} = 0$.

3.4. The Deterministic Formulation

Now, we have derived all the constraints of the indicator variables by (3.3), (3.14), (3.17), (3.18). Together with the objective function and the probability constraint represented by (3.6) and (3.7), respectively, the problem (M) can be finally written as a mix integer linear programming problem:

$$\begin{aligned} \min\ & \sum_{t=1}^{T}\left(t \cdot \sum_{n\in N_t} p_{nt}\varepsilon_{nt}\right) \\ \text{s.t.}\ & \sum_{t=1}^{T}\sum_{n\in N_t} p_{nt}\eta_{nt} \le q \\ & (3.3),\ (3.14),\ (3.17),\ (3.18) \\ & \varepsilon_{nt}, \eta_{nt}, \delta_{nt}, \xi_{nt}, \phi_{nt} \in \{0,1\}, \\ & n \in N_t, \quad t \in \{1,2,\ldots,T\}. \end{aligned} \tag{P}$$

Next, we construct an example to analyze the model and illustrate the solving process. The problem is input by an MATLAB program, and numerically solved by using Cplex software.

4. An Example

The investment process is represented by a 3-stage triple tree, noted from time 1 to time 4, as showed in Figure 1. The portfolio can be organized and reorganized at the beginning of each

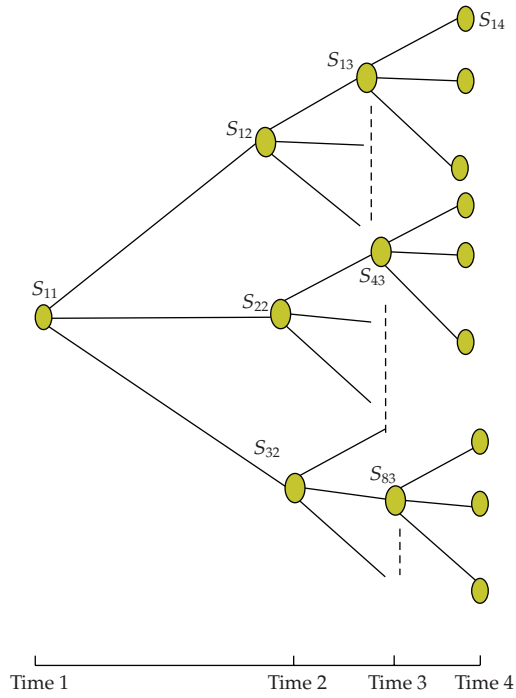


Figure 1: The scenario tree of example.

Table 1: The selected solutions of example.

		Time 1	Time 2	Time 3		Time 4	
Payoff target obtained	Scenario	None	None	$S_{1,3}$	$S_{8,3}$	None	
	Solution			$\varepsilon_{1,3} = 1$	$\varepsilon_{8,3} = 1$		
				$\delta_{1,3} = 1$	$\delta_{8,3} = 1$		
				$\phi_{1,3} = 1$	$\phi_{8,3} = 1$		
Bankruptcy happens	Scenario	None	None	$S_{4,3}$		$S_{19,4}$	$S_{27,4}$
	Solution			$\eta_{4,3} = 1$		$\eta_{19,4} = 1$	$\eta_{27,4} = 1$
				$\xi_{4,3} = 1$		$\xi_{19,4} = 1$	$\xi_{27,4} = 1$
				$\phi_{4,3} = 1$		$\phi_{19,4} = 1$	$\phi_{27,4} = 1$

stage. We simply consider a portfolio of two assets, and the prices on each decision point are given. Also, the conditional probabilities of the three notes in any single-stage subtree are $P = \{.3, .36, .34\}$ in order. For other essential constants, we assume the initial budget $B = \$100$, the target payoff $u = \$104$, and the bancruptcy banchmark $l = \$95$. In addition, we take $M_1 = 10\,000$ and $M_2 = -10\,000$ as those two large enough numbers for formulating the problem. Finally, we assign the largest accetable bancruptcy probability to be $q = 0.2$.

Cplex takes 0.41 second to optimize the problem. Reading the solution file, we find that there are chances to obtain the payoff target before the investment horizon, as clearly as in the third stage on the scenarios of $\{S_{1,3}, S_{8,3}\}$, respectively. Accordingly, the bankruptcy possibly happens on the third and the fourth stages, on the scenarios of $\{S_{4,3}, S_{19,4}, S_{27,4}\}$, which makes the total probability of bankruptcy is 0.178. Details of selected optimal solutions are shown in Table 1. Other solutions are also carefully checked, it turns out that the construction

of indicator variables does work. For example, on the children scenarios of the stopping scenarios $\{S_{1,3}, S_{8,3}, S_{4,3}\}$, the values of ϕ are all zero as the investment has been stopped before.

For two-stage problem, there are well-known algorithms such as branch-and-bound, Lagrangian relaxation, or cutting plane methods for solving it. When we extend it into the multistage case, as we are doing now, the problem becomes much more complex. As the size of the problem increases, the existing solution methods become less efficient. We will further investigate on more applicable solution methodologies. In addition to the solution methodology, another relevant research topic is to compare the investment policies under different objectives and risk constraints.

Acknowledgments

This research was supported by Guangdong Natural Science Foundation (no. 06300496), China, National Social Science Foundation (no. 08CJY064), China, and the National Science Foundation for Distinguished Young Scholars (no. 70825002), China.

References

[1] H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952.
[2] H. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*, Cowles Foundation for Research in Economics at Yale University, Monograph 16, John Wiley & Sons, New York, NY, USA, 1959.
[3] B. Dumas and E. Luciano, "An exact solution to a dynamic portfolio choice problem under transacation cost," *The Journal of Finance*, vol. 46, no. 2, pp. 577–595, 1991.
[4] E. J. Elton and M. J. Gruber, "The multi-period consumption investment problem and single period analysis," *Oxford Economic Papers*, vol. 26, no. 2, pp. 289–301, 1974.
[5] D. Li and W.-L. Ng, "Optimal dynamic portfolio selection: multiperiod mean-variance formulation," *Mathematical Finance*, vol. 10, no. 3, pp. 387–406, 2000.
[6] R. C. Merton, "Optimum consumption and portfolio rules in a continuous-time model," *Journal of Economic Theory*, vol. 3, no. 4, pp. 373–413, 1971.
[7] J. Mossion, "Optimal multiperiod portfolio policy," *The Journal of Business*, vol. 41, no. 2, pp. 215–229, 1968.
[8] M. E. Yaari, "Uncertain lifetime, life insurance, and the theory of the consumer," *The Review of Economic Studies*, vol. 32, no. 2, pp. 137–150, 1965.
[9] I. Karatzas and H. Wang, "Utility maximization with discretionary stopping," *SIAM Journal on Control and Optimization*, vol. 39, no. 1, pp. 306–329, 2000.
[10] W. Guo and Q. Hu, "Multi-period portfolio optimization when exit time is uncertain," *Journal of Management Sciences in China*, vol. 8, no. 2, pp. 14–19, 2005.
[11] S. Browne, "Beating a moving target: optimal portfolio strategies for outperforming a stochastic benchmark," *Finance and Stochastics*, vol. 3, no. 3, pp. 275–294, 1999.
[12] S. Browne, "Risk-constrained dynamic active portfolio management," *Management Science*, vol. 46, no. 9, pp. 1188–1199, 2000.
[13] C. Blanchet-Scalliet, N. El Karoui, and L. Martellini, "Dynamic asset pricing theory with uncertain time-horizon," *Journal of Economic Dynamics & Control*, vol. 29, no. 10, pp. 1737–1764, 2005.

Hindawi Publishing Corporation
Journal of Applied Mathematics and Decision Sciences
Volume 2009, Article ID 897024, 20 pages
doi:10.1155/2009/897024

Research Article

A New Decision-Making Method for Stock Portfolio Selection Based on Computing with Linguistic Assessment

Chen-Tung Chen[1] and Wei-Zhan Hung[2]

[1] *Department of Information Management, National United University, Miao-Li 36003, Taiwan*
[2] *Graduate Institute of Management, National United University, Miao-Li 36003, Taiwan*

Correspondence should be addressed to Chen-Tung Chen, ctchen@nuu.edu.tw

Received 30 November 2008; Revised 18 March 2009; Accepted 13 May 2009

Recommended by Lean Yu

The purpose of stock portfolio selection is how to allocate the capital to a large number of stocks in order to bring a most profitable return for investors. In most of past literatures, experts considered the portfolio of selection problem only based on past crisp or quantitative data. However, many qualitative and quantitative factors will influence the stock portfolio selection in real investment situation. It is very important for experts or decision-makers to use their experience or knowledge to predict the performance of each stock and make a stock portfolio. Because of the knowledge, experience, and background of each expert are different and vague, different types of 2-tuple linguistic variable are suitable used to express experts' opinions for the performance evaluation of each stock with respect to criteria. According to the linguistic evaluations of experts, the linguistic TOPSIS and linguistic ELECTRE methods are combined to present a new decision-making method for dealing with stock selection problems in this paper. Once the investment set has been determined, the risk preferences of investor are considered to calculate the investment ratio of each stock in the investment set. Finally, an example is implemented to demonstrate the practicability of the proposed method.

1. Introduction

The purpose of stock portfolio selection is how to allocate the capital to a large number of stocks in order to bring a most profitable return for investors [1]. For this point of view, stock portfolio decision problem can be divided into two questions.

(1) Which stock do you choose?

(2) Which investment ratio do you allocate your capital to this stock?

There are some literatures to handle the stock portfolio decision problem. Markowitz proposed the mean-variance method for the stock portfolio decision problem in 1952 [2].

In his method, an expected return rate of a bond is treated as a random variable. Stochastic programming is applied to solve the problem. The basic concept of his method can be expressed as follows.

(1) When the risk of stock portfolio is constant, we should pursue to maximize the return rate of stock portfolio.

(2) When the return rate of stock portfolio is constant, we should pursue to minimize the risk of stock portfolio.

The capital asset pricing model (CAPM), Sharpe-Lintner model, Black model, and two-factor model are derived from the mean-variance method [3, 4]. The capital asset pricing model (CAPM) was developed in 1960s. The concept of the CAPM is that the excepted return rate of the capital with risk is equal to the interest rate of the capital without risk and market risk premium [4]. The methods and theory of the financial decision making can be found in [5–7]. In 1980, Saaty proposed Analytic Hierarchy Process (AHP) to deal with the stock portfolio decision problem by evaluating the performance of each company in different level of criteria [8]. Edirisinghe and Zhang [9] selected the securities by using Data Envelopment Analysis (DEA). Huang [1] defined a new definition of risk and use genetic algorithm to cope with stock portfolio decision problem. Generally, in the portfolio selection problem the decision maker considers simultaneously conflicting objectives such as rate of return, liquidity, and risk. Multiobjective programming techniques such as goal programming (GP) and compromise programming (CP) are used to choose the portfolio [10–12]. Considering the uncertainty of investment environment, Tiryaki transferred experts' linguistic value into triangle fuzzy number and used a new fuzzy ranking and weighting algorithm to obtain the investment ratio of each stock [4]. In fact, the stock portfolio decision problem can be described as multiple criteria decision making (MCDM) problem.

Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method is developed by Hwang and Yoon [13], which is one of the well-known MCDM methods. The basic principle of the TOPSIS method is that the chosen alternative should have the shortest distance from the positive ideal solution (PIS) and the farthest distance from the negative ideal solution (NIS). It is an effective method to determine the total ranking order of decision alternatives.

The Elimination et choice in Translating to Reality (ELECTRE) method is a highly developed multicriteria analysis model which takes into account the uncertainty and vagueness in the decision process [14]. It is based on the axiom of partial comparability and it can simplify the evaluation procedure of alternative selection. The ELECTRE method can easily compare the degree of difference among all of alternatives.

In MCDM method, experts can express their opinions by using crisp value, triangle fuzzy numbers, trapezoidal fuzzy numbers, interval numbers, and linguistic variables. Due to imprecise information and experts' subjective opinion that often appear in stock portfolio decision process, crisp values are inadequate for solving the problems. A more realistic approach may be to use linguistic assessments instead of numerical values [15, 16]. The 2-tuple linguistic representation model is based on the concept of symbolic translation [17, 18]. Experts can apply 2-tuple linguistic variables to express their opinions and obtain the final evaluation result with appropriate linguistic variable. It is an effective method to reduce the mistakes of information translation and avoid information loss through computing with words [19]. In general, decision makers would use the different 2-tuple linguistic variables based on their knowledge or experiences to express their opinions [20]. In this paper, we use different type of 2-tuple linguistic variable to express experts' opinions and combine

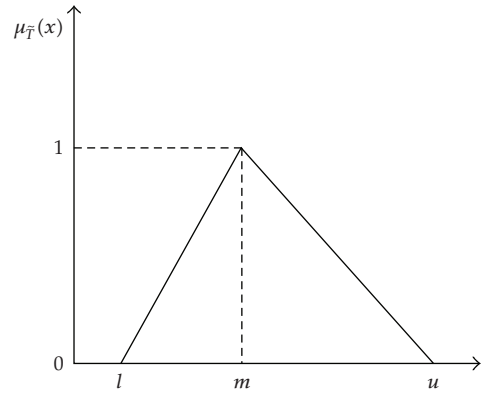


Figure 1: Triangular fuzzy number $\tilde{T}$.

linguistic ELECTRE method with TOPSIS method to obtain the final investment ratio which is reasonable in real decision environment.

This paper is organized as follows. In Section 2, we present the context of fuzzy set and the definition and operation of 2-tuple linguistic variable. In Section 3, we describe the detail of the proposed method. In Section 4, an example is implemented to demonstrate the procedure for the proposed method. Finally, the conclusion is discussed at the end of this paper.

2. The 2-Tuple Linguistic Representation

2.1. Fuzzy Set and Triangular Fuzzy Number

Fuzzy set theory is first introduced by Zadeh in 1965 [21]. Fuzzy set theory is a very feasible method to handle the imprecise and uncertain information in a real world [22]. Especially, it is more suitable for subjective judgment and qualitative assessment in the evaluation processes of decision making than other classical evaluation methods applying crisp values [23, 24].

A positive triangular fuzzy number (PTFN) $\tilde{T}$ can be defined as $\tilde{T} = (l, m, u)$, where $l \leq m \leq u$ and $l > 0$, shown in Figure 1. The membership function $\mu_{\tilde{T}}(x)$ of positive triangular fuzzy number (PTFN) $\tilde{T}$ is defined as [15]

$$\mu_{\tilde{T}}(x) = \begin{cases} \dfrac{x-l}{m-l}, & l < x < m, \\ \dfrac{u-x}{u-m}, & m < x < u, \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

A linguistic variable is a variable whose values are expressed in linguistic terms. In other words, variable whose values are not numbers but words or sentences in a nature or artificial language [25–27]. For example, "weight" is a linguistic variable whose values are very low, low, medium, high, very high, and so forth. These linguistic values can also be represented by fuzzy numbers. There are two advantages for using triangular fuzzy number to express linguistic variable [28]. First, it is a rational and simple method to use triangular

fuzzy number to express experts' opinions. Second, it is easy to do fuzzy arithmetic when using triangular fuzzy number to express the linguistic variable. It is suitable to represent the degree of subjective judgment in qualitative aspect than crisp value.

2.2. The 2-Tuple Linguistic Variable

Let $S = \{s_0, s_1, s_2, \ldots, s_g\}$ be a finite and totally ordered linguistic term set. The number of linguistic term is $g+1$ in set S. A 2-tuple linguistic variable can be expressed as (s_i, α_i), where s_i is the central value of ith linguistic term in S and α_i is a numerical value representing the difference between calculated linguistic term and the closest index label in the initial linguistic term set. The symbolic translation function Δ is presented in [29] to translate crisp value β into a 2-tuple linguistic variable. Then, the symbolic translation process is applied to translate β ($\beta \in [0,1]$) into a 2-tuple linguistic variable. The generalized translation function can be represented as [30]:

$$\begin{aligned} \Delta : [0,1] &\longrightarrow S \times \left[-\frac{1}{2g}, \frac{1}{2g}\right) \\ \Delta(\beta) &= (s_i, \alpha_i), \end{aligned} \tag{2.2}$$

where $i = \text{round}(\beta \times g)$, $\alpha_i = \beta(-i/g)$ and $\alpha_i \in [-1/2g, 1/2g)$.

A reverse function Δ^{-1} is defined to return an equivalent numerical value β from 2-tuple linguistic information (s_i, α_i). According to the symbolic translation, an equivalent numerical value β is obtained as follow [30]

$$\Delta^{-1}(s_i, \alpha_i) = \frac{i}{g} + \alpha_i = \beta. \tag{2.3}$$

Let $x = \{(r_1, \alpha_1), \ldots, (r_n, \alpha_n)\}$ be a 2-tuple linguistic variable set. The arithmetic mean $\overline{X}$ is computed as [31]

$$\overline{X} = \Delta\left(\frac{1}{n}\sum_{i=1}^{n}\Delta^{-1}(r_i, \alpha_i)\right) = (s_m, \alpha_m), \tag{2.4}$$

where n is the amount of 2-tuple linguistic variable. The (s_m, α_m) is a 2-tuple linguistic variable which is represented as the arithmetic mean.

In general, decision makers would use the different 2-tuple linguistic variables based on their knowledge or experiences to express their opinions [20]. For example, the different types of linguistic variables show as Table 1. Each 2-tuple linguistic variable can be represented as a triangle fuzzy number. A transformation function is needed to transfer these 2-tuple linguistic variables from different linguistic sets to a standard linguistic set at unique domain. In the method of Herrera and Martinez [29], the domain of the linguistic variables will increase as the number of linguistic variable is increased. To overcome this drawback, a new translation function is applied to transfer a crisp number or 2-tuple linguistic variable to a standard linguistic term at the unique domain [30]. Suppose that the interval $[0,1]$ is the unique domain. The linguistic variable sets with different semantics (or types) will be

defined by partitioning the interval $[0,1]$. Transforming a crisp number β ($\beta \in [0,1]$) into ith linguistic term $(s_i^{n(t)}, \alpha_i^{n(t)})$ of type t as

$$\Delta_t(\beta) = \left(s_i^{n(t)}, \alpha_i^{n(t)}\right), \tag{2.5}$$

where $i = \text{round}(\beta \times g_t)$, $\alpha_i^{n(t)} = \beta(-i/g_t)$, $g_t = n(t) - 1$, and $n(t)$ is the number of linguistic variable of type t.

Transforming ith linguistic term of type t into a crisp number β ($\beta \in [0,1]$) as

$$\Delta_t^{-1}\left(s_i^{n(t)}, \alpha_i^{n(t)}\right) = \frac{i}{g_t} + \alpha_i^{n(t)} = \beta, \tag{2.6}$$

where $g_t = n(t) - 1$ and $\alpha_i^{n(t)} \in [-1/2g_t, 1/2g_t)$.

Therefore, the transformation from ith linguistic term $(s_i^{n(t)}, \alpha_i^{n(t)})$ of type t to kth linguistic term $(s_k^{n(t+1)}, \alpha_k^{n(t+1)})$ of type $t+1$ at interval $[0,1]$ can be expressed as

$$\Delta_{t+1}\left(\Delta_t^{-1}\left(s_i^{n(t)}, \alpha_i^{n(t)}\right)\right) = \left(s_k^{n(t+1)}, \alpha_k^{n(t+1)}\right), \tag{2.7}$$

where $g_{t+1} = n(t+1) - 1$ and $\alpha_k^{n(t+1)} \in [-1/2g_{t+1}, 1/2g_{t+1})$.

3. Proposed Method

Because of the knowledge, experience and background of each expert is different and experts' opinions are usually uncertain and imprecise, it is difficult to use crisp value to express experts' opinions in the process of evaluating the performance of stock. Instead of crisp value, the 2-Tuple linguistic valuable which is an effective method to reduce the mistakes of information translation and avoid information loss through computing with words to express experts' opinions [19]. In this paper, different types of 2-tuple linguistic variables are used to express experts' opinions.

The TOPSIS method is one of the well-known MCDM methods. It is an effective method to determine the ranking order of decision alternatives. However, this method cannot distinguish the difference degree between two decision alternatives easily. Based on the axiom of partial comparability, the ELECTRE method can easily compare the degree of difference among of all alternatives. This method always cannot provide the total ordering of all decision alternatives. Therefore, the ELECTRE and TOPSIS methods are combined to determine the final investment ratio.

In the proposed model, the subjective opinions of experts can be expressed by different 2-tuple linguistic variables in accordance with their habitual knowledge and experience. After aggregating opinions of all experts, the linguistic TOPSIS and linguistic ELECTRE methods are applied to obtain the investment portfolio sets Ω_t and Ω_e, respectively. The strict stock portfolio set Ω_{ip} is determined by intersection Ω_t with Ω_e. In general, the risk preference of investor can be divided into three types such as risk-averter, risk-neutral, and risk-loving. Considering the risk preference of investor, we can calculate the investment ratio of each

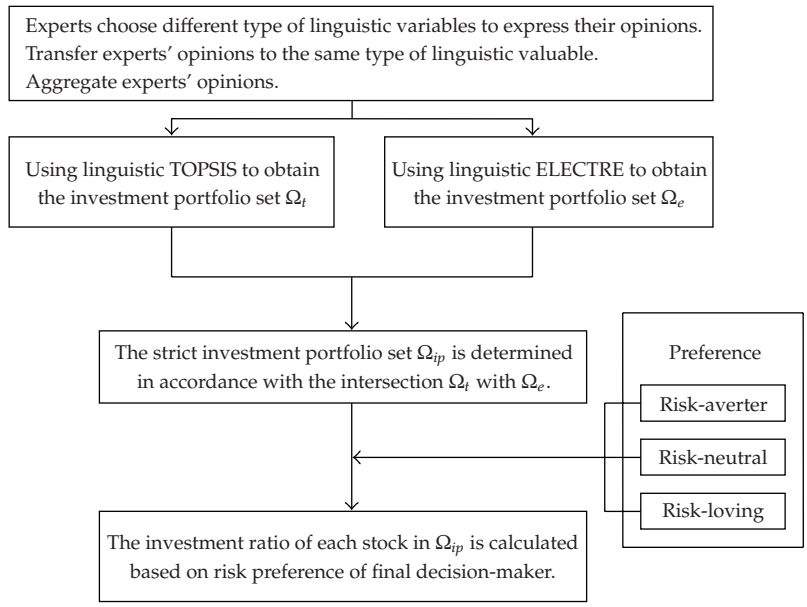


Figure 2: The decision-making process of the proposed method.

stock in strict stock portfolio set Ω_{ip}. The decision process of the proposed method is shown as in Figure 2.

In general, a stock portfolio decision may be described by means of the following sets:

(i) a set of experts or decision-makers called $E = \{E_1, E_2, \ldots, E_K\}$;

(ii) a set of stocks called $S = \{S_1, S_2, \ldots, S_m\}$;

(iii) a set of criteria $C = \{C_1, C_2, \ldots, C_n\}$ with which stock performances are measured;

(iv) a weight vector of each criterion $W = (W_1, W_2, \ldots, W_n)$;

(v) a set of performance ratings of each stock with respect to each criterion called $\widetilde{S}_{ij}$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$.

According to the aforementioned description, there are K experts, m stocks and n criteria in the decision process of stock portfolio. Experts can express their opinions by different 2-tuple linguistic variables. The kth expert's opinion about the performance rating of ith stock with respect to jth criterion can be represented as $\widetilde{S}_{ij}^k = (S_{ij}^k, \alpha_{ij}^k)$. The kth expert's opinion about the importance of jth criterion can be represented as $\widetilde{W}_{jk} = (S_{jk}^w, \alpha_{jk}^w)$.

The aggregated linguistic rating $\widetilde{S}_{ij}$ of each stock with respect to each criterion can be calculated as

$$\widetilde{S}_{ij} = \Delta\left(\frac{1}{K}\sum_{k=1}^{K}\Delta^{-1}\left(S_{ij}^k, \alpha_{ij}^k\right)\right) = \left(S_{ij}, \alpha_{ij}\right). \tag{3.1}$$

The aggregated linguistic weight $\widetilde{w}_j$ of each criterion can be calculated as

$$\widetilde{W}_j = \Delta\left(\frac{1}{K}\sum_{k=1}^{K}\Delta^{-1}\left(S_{jk}^{w}, \alpha_{jk}^{w}\right)\right) = \left(S_j^w, \alpha_j^w\right). \tag{3.2}$$

3.1. Linguistic TOPSIS Method

Considering the different importance of each criterion, the weighted linguistic decision matrix is constructed as

$$\widetilde{V}[\widetilde{v}_{ij}]_{m\times n}, \quad i = 1,2,\ldots,m,\ j = 1,2,\ldots,n, \tag{3.3}$$

where $\widetilde{v}_{ij} = \widetilde{x}_{ij}(\cdot)\widetilde{w}_j = \Delta(\Delta^{-1}(S_{ij}, \alpha_{ij}) * \Delta^{-1}(S_j^w, \alpha_j^w)) = (S_{ij}^v, \alpha_{ij}^v)$.

According to the weighted linguistic decision matrix, the linguistic positive-ideal solution (LPIS, S^*) and linguistic negative-ideal solution (LNIS, S^-) can be defined as

$$\begin{aligned} S^* &= (\widetilde{v}_1^*, \widetilde{v}_2^*, \ldots, \widetilde{v}_n^*), \\ S^- &= (\widetilde{v}_1^-, \widetilde{v}_2^-, \ldots, \widetilde{v}_n^-), \end{aligned} \tag{3.4}$$

where $\widetilde{v}_j^* = \max_i\{(S_{ij}^v, \alpha_{ij}^v)\}$ and $\widetilde{v}_j^- = \min_i\{(S_{ij}^v, \alpha_{ij}^v)\}$, $i = 1,2,\ldots,m$, $j = 1,2,\ldots,n$.

The distance of each stock S_i $(i = 1,2,\ldots,m)$ from S^* and S^-can be currently calculated as

$$\begin{aligned} d_i^* &= d(S_i, S^*) = \sqrt{\sum_{j=1}^{n} d(\widetilde{v}_{ij}, \widetilde{v}_j)} = \sqrt{\sum_{j=1}^{n}\left(\Delta^{-1}\left(\max_i\left\{\left(S_{ij}^v, \alpha_{ij}^v\right)\right\}\right) - \Delta^{-1}\left(S_{ij}^v, \alpha_{ij}^v\right)\right)^2}, \\ d_i^- &= d(S_i, S^-) = \sqrt{\sum_{j=1}^{n} d(\widetilde{v}_{ij}, \widetilde{v}_j)} = \sqrt{\sum_{j=1}^{n}\left(\Delta^{-1}\left(S_{ij}^v, \alpha_{ij}^v\right) - \Delta^{-1}\left(\min_i\left\{\left(S_{ij}^v, \alpha_{ij}^v\right)\right\}\right)\right)^2}. \end{aligned} \tag{3.5}$$

A closeness coefficient is defined to determine the ranking order of all stocks once d_i^* and d_i^- of each stock S_i $(i = 1,2,\ldots,m)$ have been calculated. The closeness coefficient represents the distances to the linguistic positive-ideal solution (S^*) and the linguistic negative-ideal solution (S^-) simultaneously by taking the relative closeness to the linguistic positive-ideal solution. The closeness coefficient (CC_i) of each stock is calculated as

$$\mathrm{CC}_i = \frac{d_i^-}{d_i^* + d_i^-}, \quad i = 1,2,\ldots,m. \tag{3.6}$$

The higher CC_i means that stock S_i relatively close to positive ideal solution, the stock S_i has more ability to compete with each others. If the closeness coefficient of stock S_i is greater than the predetermined threshold value β_t, we consider stock S_i is good enough to choose in the investment portfolio set. According to closeness coefficient of each stock, the

investment portfolio set Ω_t can be determined based on investment threshold value β_t as $\Omega_t = \{S_i \mid \mathrm{CC}_i \geq \beta_t\}$. Finally, the investment ratio of each stock in Ω_t can be calculated as

$$P_t(S_i) = \begin{cases} \dfrac{\mathrm{CC}(S_i)}{\sum_{S_i \in \Omega_t} \mathrm{CC}(S_i)}, & S_i \in \Omega_t, \\ 0, & S_i \notin \Omega_t, \end{cases} \tag{3.7}$$

where $P_t(S_i)$ is the investment ratio of each stock by linguistic TOPSIS method.

3.2. Linguistic ELECTRE Method

According to the ELECTRE method, the concordance index $C_j(S_i, S_l)$ is calculated for S_i and S_l $(i \neq l,\ i, l = 1, 2, \ldots, m)$ with respect to each criterion as

$$C_j(S_i, S_l) = \begin{cases} 1, & \Delta^{-1}(\widetilde{s}_{ij}) \geq \Delta^{-1}(\widetilde{s}_{lj}) - q_j, \\ \dfrac{\Delta^{-1}(\widetilde{s}_{ij}) - \Delta^{-1}(\widetilde{s}_{lj}) + p_j}{p_j - q_j}, & \Delta^{-1}(\widetilde{s}_{lj}) - q_j \geq \Delta^{-1}(\widetilde{s}_{ij}) \geq \Delta^{-1}(\widetilde{s}_{lj}) - p_j, \\ 0, & \Delta^{-1}(\widetilde{s}_{ij}) \leq \Delta^{-1}(\widetilde{s}_{lj}) - p_j, \end{cases} \tag{3.8}$$

where q_j and p_j are indifference and preference threshold values for criterion C_j, $p_j > q_j$.

The discordance index $D_j(S_i, S_l)$ is calculated for each pair of stocks with respect to each criterion as

$$D_j(S_i, S_l) = \begin{cases} 1, & \Delta^{-1}(\widetilde{s}_{ij}) \leq \Delta^{-1}(\widetilde{s}_{lj}) - v_j, \\ \dfrac{\Delta^{-1}(\widetilde{s}_{lj}) - p_j - \Delta^{-1}(\widetilde{s}_{ij})}{v_j - p_j}, & \Delta^{-1}(\widetilde{s}_{lj}) - p_j \geq \Delta^{-1}(\widetilde{s}_{ij}) \geq \Delta^{-1}(\widetilde{s}_{lj}) - v_j, \\ 0, & \Delta^{-1}(\widetilde{s}_{ij}) \geq \Delta^{-1}(\widetilde{s}_{lj}) - p_j, \end{cases} \tag{3.9}$$

where v_j is the veto threshold for criterion C_j, $v_j > p_j$.

Calculate the overall concordance index $C(S_i, S_l)$ as

$$C(S_i, S_l) = \sum_{j=1}^{n} \Delta^{-1}(\widetilde{w}_j) * C_j(S_i, S_l). \tag{3.10}$$

The credibility matrix $S(S_i, S_l)$ of each pair of the stocks is calculated as

$$S(S_i, S_l) = \begin{cases} C(S_i, S_l), & \text{if } D_j(S_i, S_l) \leq C(S_i, S_l)\ \forall j, \\ C(S_i, S_l) \displaystyle\prod_{j \in J(S_i, S_l)} \frac{1 - D_j(S_i, S_l)}{1 - C(S_i, S_l)}, & \text{otherwise}, \end{cases} \tag{3.11}$$

where $J(S_i, S_l)$ is the set of criteria for which $D_j(S_i, S_l) > C(S_i, S_l)$, $i \neq l$, $i, l = 1, 2, \ldots, m$.

The concordance credibility and discordance credibility degrees are defined as [32]

$$\begin{aligned}\phi^{+}(S_i) &= \sum_{i \neq l} S(S_i, S_l),\\ \phi^{-}(S_i) &= \sum_{i \neq l} S(S_l, S_i).\end{aligned} \tag{3.12}$$

The concordance credibility degree represents that the degree of stock S_i is at least as good as all the other stocks. The discordance credibility degree represents that the degree of all the other stocks is at least as good as stock S_i.

Then, the net credibility degree is defined as $\phi(S_i) = \phi^{+}(S_i) - \phi^{-}(S_i)$. If the net credibility degree of stock S_i is higher, then it represents a higher attractiveness of stock S_i. In order to determine the investment ratio, the outranking index of stock S_i can be defined as

$$\mathrm{OTI}(S_i) = \frac{\phi(S_i)/(m-1)+1}{2}. \tag{3.13}$$

Property 3.1. According to the definition of $\mathrm{OTI}(S_i)$, we can find $0 \leq \mathrm{OTI}(S_i) \leq 1$.

Proof. Because $\phi(S_i) = \phi^{+}(S_i) - \phi^{-}(S_i) = \sum_{i \neq l} S(S_i, S_l) - \sum_{i \neq l}(S_l \cdot S_i)$, $i \neq l$, $i, l = 1, 2, \ldots, m$.

If the stock S_i is better than S_l with respect to each criterion, the best case is

$$\sum_{i \neq l} S(S_i, S_l) - \sum_{i \neq l}(S_l, S_i) = m - 1. \tag{3.14}$$

If the stock S_i is worse than S_l with respect to each criterion, the worst case is

$$\sum_{i \neq l} S(S_i, S_l) - \sum_{i \neq l}(S_l, S_i) = -(m-1). \tag{3.15}$$

Therefore, $-(m-1) \leq \phi(S_i) \leq m - 1$.

Then, $-1 \leq \phi(S_i)/(m-1) \leq 1$. Finally, we can prove $0 \leq (\phi(S_i)/(m-1)+1)/2 = \mathrm{OTI}(S_i) \leq 1$.

The $\mathrm{OTI}(S_i)$ denotes the standardization result of the net credibility degree. According to the definition, it is easy to understand and transform the net credibility degree into interval $[0, 1]$.

If the outranking index of stock S_i is greater than the predetermined threshold value β_e, we consider stock S_i is good enough to choose in the investment portfolio set. According to the outranking index of each stock, the investment portfolio set Ω_e can be determined based on investment threshold value β_e as $\Omega_e = \{S_i \mid \mathrm{OTI}(S_i) \geq \beta_e\}$. Finally, the investment ratio of each stock in Ω_e can be calculated as

$$P_e(S_i) = \begin{cases} \dfrac{\mathrm{OTI}(S_i)}{\sum_{S_i \in \Omega_e} \mathrm{OTI}(S_i)}, & S_i \in \Omega_e,\\ 0, & S_i \notin \Omega_e, \end{cases} \tag{3.16}$$

where $P_e(S_i)$ is the investment ratio of each stock by using linguistic ELECTRE method. □

3.3. Stock Portfolio Decision

We can consider Linguistic TOPSIS and Linguistic ELECTRE methods as two financial experts to provide investment ratio of each stock, respectively. Smart investor will make a stock portfolio decision by considering the suggestions of investment ratio of each stock simultaneously. Therefore, the portfolio set Ω_{ip} is defined as strict stock portfolio set $\Omega_{ip} = \Omega_t \cap \Omega_e$.

According to the closeness coefficient, the investment ratio of each stock in strict stock portfolio set Ω_{ip} can be calculated as

$$P_{t\_ip}(S_i) = \begin{cases} \dfrac{\mathrm{CC}(S_i)}{\sum_{S_i \in \Omega_{ip}} \mathrm{CC}(S_i)}, & S_i \in \Omega_{ip}, \\ 0, & S_i \notin \Omega_{ip}. \end{cases} \tag{3.17}$$

According to the outranking index, the investment ratio of each stock in strict stock portfolio set Ω_{ip} can be calculated as

$$P_{e\_ip}(S_i) = \begin{cases} \dfrac{\mathrm{OTI}(S_i)}{\sum_{S_i \in \Omega_{ip}} \mathrm{OTI}(S_i)}, & S_i \in \Omega_{ip}, \\ 0, & S_i \notin \Omega_{ip}. \end{cases} \tag{3.18}$$

In general, the investment preference of investors can be divided into three types such as risk-averter (RA), risk-neutral (RN), and risk-loving (RL). If a person is risk-averter, he/she will consider the smaller investment rates between $P_{t\_ip}(S_i)$ and $P_{e\_ip}(S_i)$. Therefore, the final ratio of each stock in strict portfolio set can be calculated as

$$P_{\mathrm{RA}}(S_i) = \frac{\min\left(P_{t\_ip}(S_i), P_{e\_ip}(S_i)\right)}{\sum_{S_i \in \Omega_{ip}} \min\left(P_{t\_ip}(S_i), P_{e\_ip}(S_i)\right)}. \tag{3.19}$$

If a person is risk-neutral, he/she will consider the average investment rates between $P_{t\_ip}(S_i)$ and $P_{e\_ip}(S_i)$. Therefore, the final ratio of each stock in strict portfolio set can be calculated as

$$P_{\mathrm{RN}}(S_i) = \frac{\left(P_{t\_ip}(S_i) + P_{e\_ip}(S_i)\right)/2}{\sum_{S_i \in \Omega_{ip}} \left(\left(P_{t\_ip}(S_i) + P_{e\_ip}(S_i)\right)/2\right)}. \tag{3.20}$$

If a person is risk-loving, he/she will consider the bigger investment rates between $P_{t\_ip}(S_i)$ and $P_{e\_ip}(S_i)$. Therefore, the final ratio of each stock in portfolio set can be calculated as

$$P_{\mathrm{RL}}(S_i) = \frac{\max\left(P_{t\_ip}(S_i), P_{e\_ip}(S_i)\right)}{\sum_{S_i \in \Omega_{ip}} \max\left(P_{t\_ip}(S_i), P_{e\_ip}(S_i)\right)}. \tag{3.21}$$

Table 1: Ten stocks of semiconduct industry in Taiwan.

S_1	Taiwan Semiconductor Manufacturing Co. Ltd.	S_2	United Microelectronics Corp.
S_3	Advanced Semiconductor Engineering, Inc.	S_4	Via Technologies, Inc.
S_5	MediaTek Inc.	S_6	King Yuan Electronics Co. Ltd.
S_7	Taiwan Mask Corp.	S_8	Winbond Electronics Corp.
S_9	SunPlus Technology Co. Ltd.	S_{10}	Nanya Technology Corporation

4. Numerical Example

An example with ten stocks of semiconduct industry in placecountry-region, Taiwan, will be considered to determine the investment ratio of each stock in this paper. Ten stocks are shown as Table 1. A committee of three financial experts $E = \{E_1, E_2, E_3\}$ has been formed to evaluate the performance of each stock. They are famous professors of a department of finance at well-known university in country-regionplace, Taiwan. Their knowledge and experiences are enough to evaluate the stock performance of each company for this example. In the process of criteria selection, they considered the quantitative and qualitative factors to deal with the portfolio selection. After the serious discussion and selection by three financial experts, six criteria are considered to determined the investment ratio of each stock such as profitability (C_1), asset utilization (C_2), liquidity (C_3), leverage (C_4), valuation (C_5), growth (C_6).

Profitability (C_1)

The goal of enterprise is tomakeaprofit. There are some indexes to evaluate the profitability of a company such as earnings per share (EPS), net profit margin, return on assets (ROA), and return on equity (ROE). The profitability of a company will influence the performance of each stock.

Asset Utilization (C_2)

Asset utilization means the efficiency of using company's resource in a period. A good company will promote the resource using efficiency as more as possible. Experts evaluate the asset utilization of the company based on receivables turnover, inventory turnover, and asset turnover.

Liquidity (C_3)

Liquidity will focus on cash flow generation and a company's ability to meet its financial obligations. When company's transfer assets (1 and, factory buildings, equipment, patent, goodwill) to currency in a short period, there will have some loss because the company's manager do not have enough time to find out the buyer who provide the highest price. An appropriate liquidity ratio (debt to equity ratio, current ratio, quick ratio) will both prevent liquidity risk and minimize the working capital.

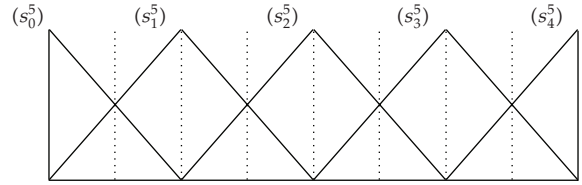


Figure 3: Membership functions of linguistic variables at type 1 ($t = 1$).

Table 2: Different types of linguistic variables.

Type		Linguistic variable	Figure
1	Performance	Extremely Poor (s_0^5), Poor (s_1^5), Fair (s_2^5), Good (s_3^5), Extremely Good (s_4^5)	Figure 3
	Weight	Extremely Low (s_0^5), Low (s_1^5), Fair (s_2^5), High (s_3^5), Extremely High (s_4^5)	
2	Performance	Extremely Poor (s_0^7), Poor (s_1^7), Medium Poor (s_2^7), Fair (s_3^7), Medium Good (s_4^7), Good (s_5^7), Extremely Good (s_6^7)	Figure 4
	Weight	Extremely Low (s_0^7), Low (s_1^7), Medium Low (s_2^7), Fair (s_3^7), Medium High (s_4^7), High (s_5^7), Extremely High (s_6^7)	
3	Performance	Extremely Poor (s_0^9), Very Poor (s_1^9), Poor (s_2^9), Medium Poor (s_3^9), Fair (s_4^9), Medium Good (s_5^9), Good (s_6^9), Very Good (s_7^9), Extremely Good (s_8^9)	Figure 5
	Weight	Extremely Low (s_0^9), Very Low (s_1^9), Low (s_2^9), Medium Low (s_3^9), Fair (s_4^9), Medium High (s_5^9), High (s_6^9), Very High (s_7^9), Extremely High (s_8^9)	

Leverage (C_4)

When the return on assets is greater than lending rate, it is time for a company to lend money to operate. But increasing the company's debt will increase risk if the company does not earn enough money to pay the debt in the future. A suitable leverage ratio is one of the criteria to evaluate the performance of each stock.

Valuation (C_5)

Book value means the currency which all of the company's assets transfer to, stock value means the price if you want to buy now, earnings before amortization, interest and taxes ratio (EBAIT) means the company earns in this year, expert must consider the best time point to buy the stock by Technical Analysis (TA) and Time Series Analysis (TSA). So, valuation is also one of the criteria to evaluate the performance of each stock.

Growth (C_6)

If the scale of a company was expanded year by year, EBAIT will increase which is like "compound interest." Because of economies of scale, the growth of the company will promote asset utilization and then raise the EBAIT and EPS.

According to the proposed method, the computational procedures of the problem are summarized as follows.

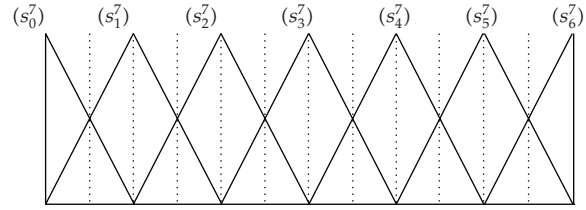


Figure 4: Membership functions of linguistic variables at type 2 ($t = 2$).

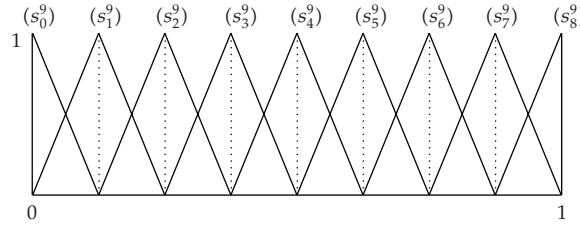


Figure 5: Membership functions of linguistic variables at type 3 ($t = 3$).

Table 3: Evaluation decisions (the ratings of the all stocks under all criteria) by three experts.

	C_1			C_2			C_3			C_4			C_5			C_6		
	E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3
S_1	F	F	G	EG	MG	VG	G	MG	VG	P	F	EG	F	G	VG	P	MG	VG
S_2	P	F	MG	F	MG	G	F	F	G	EP	F	MG	P	MG	VG	P	MG	G
S_3	F	F	G	G	F	MG	F	MG	G	F	F	MG	P	MG	G	F	MG	MG
S_4	F	G	MG	G	G	G	F	MG	MG	F	G	G	F	MG	MG	P	EG	MG
S_5	F	MG	EG	G	MG	VG	F	G	G	G	MG	VG	G	MG	VG	P	G	G
S_6	P	F	G	G	F	VG	F	F	VG	P	MG	VG	F	F	G	F	F	G
S_7	G	F	G	P	MG	VG	F	F	G	F	F	VG	P	MG	VG	F	MG	VG
S_8	EP	MG	G	F	F	VG	EP	F	VG	EP	MG	EG	EP	MG	VG	P	MG	VG
S_9	G	MG	VG	F	MG	G	F	F	VG	F	MG	VG	F	MG	VG	F	G	G
S_{10}	EP	G	G	F	G	G	F	MG	MG	EP	MG	G	EP	F	MG	EP	MG	MG

Step 1. Each expert selects the suitable 2-tuple linguistic variables to express their opinions. Expert 1 uses linguistic variables with 5 scale of linguistic term set to express his opinion, expert 2 uses linguistic variables with 7 scale of linguistic term set and expert 3 uses linguistic variables with 9 scale of linguistic term set, respectively (see Table 2).

Step 2. Each expert expresses his opinion about the performance of each stock with respect to each criterion as shown in Table 3.

Step 3. Each expert expresses his opinion about the importance of each criterion as shown in Table 4.

Step 4. Transform the linguistic ratings into the linguistic variables of type 2 and aggregate the linguistic ratings of each stock with respect to criteria as Table 5.

Table 4: Evaluation decisions (the weightings of all criteria) by three experts.

	C_1	C_2	C_3	C_4	C_5	C_6
E_1	EH	H	H	H	EH	F
E_2	EH	H	H	MH	H	H
E_3	EH	EH	VH	EH	VH	H

Table 5: Transfer to the linguistic variable of type 2.

*S*tock	Criterion	E_1	E_1	E_1	Average
C_1	S_1	S_3^7, 0.0000	S_3^7, 0.0000	S_5^7, −0.0833	S_4^7, −0.0833
	S_2	S_2^7, −0.0833	S_3^7, 0.0000	S_4^7, −0.0417	S_3^7, −0.0417
	S_3	S_3^7, 0.0000	S_3^7, 0.0000	S_5^7, −0.0833	S_4^7, −0.0833
	S_4	S_3^7, 0.0000	S_5^7, 0.0000	S_4^7, −0.0417	S_4^7, −0.0139
	S_5	S_3^7, 0.0000	S_4^7, 0.0000	S_6^7, 0.0000	S_4^7, 0.0556
	S_6	S_2^7, −0.0833	S_3^7, 0.0000	S_5^7, −0.0833	S_3^7, 0.0000
	S_7	S_5^7, −0.0833	S_3^7, 0.0000	S_5^7, −0.0833	S_4^7, 0.0000
	S_8	S_0^7, 0.0000	S_4^7, 0.0000	S_5^7, −0.0833	S_3^7, −0.0278
	S_9	S_5^7, −0.0833	S_4^7, 0.0000	S_5^7, 0.0417	S_5^7, −0.0694
	S_{10}	S_0^7, 0.0000	S_5^7, 0.0000	S_5^7, −0.0833	S_3^7, 0.0278
C_2	S_1	S_6^7, 0.0000	S_4^7, 0.0000	S_5^7, 0.0417	S_5^7, 0.0139
	S_2	S_3^7, 0.0000	S_4^7, 0.0000	S_5^7, −0.0833	S_4^7, −0.0278
	S_3	S_5^7, −0.0833	S_3^7, 0.0000	S_4^7, −0.0417	S_4^7, −0.0417
	S_4	S_5^7, −0.0833	S_5^7, 0.0000	S_5^7, −0.0833	S_5^7, −0.0556
	S_5	S_5^7, −0.0833	S_4, 0.0000	S_5^7, 0.0417	S_5^7, −0.0694
	S_6	S_5^7, −0.0833	S_3^7, 0.0000	S_5^7, 0.0417	S_4^7, 0.0417
	S_7	S_2^7, −0.0833	S_4^7, 0.0000	S_5^7, 0.0417	S_4^7, −0.0694
	S_8	S_3^7, 0.0000	S_3^7, 0.0000	S_5^7, 0.0417	S_4^7, −0.0417
	S_9	S_3^7, 0.0000	S_4, 0.0000	S_5^7, −0.0833	S_4^7, −0.0278
	S_{10}	S_3^7, 0.0000	S_5^7, 0.0000	S_5^7, −0.0833	S_4^7, 0.0278
C_3	S_1	S_5^7, −0.0833	S_4^7, 0.0000	S_5^7, 0.0417	S_5^7, −0.0694
	S_2	S_3^7, 0.0000	S_3^7, 0.0000	S_5^7, −0.0833	S_4^7, −0.0833
	S_3	S_3^7, 0.0000	S_4^7, 0.0000	S_5^7, −0.0833	S_4^7, −0.0278
	S_4	S_3^7, 0.0000	S_4^7, 0.0000	S_4^7, −0.0417	S_4^7, −0.0694
	S_5	S_3^7, 0.0000	S_5^7, 0.0000	S_5^7, −0.0833	S_4^7, 0.0278
	S_6	S_3^7, 0.0000	S_3^7, 0.0000	S_5^7, 0.0417	S_4^7, −0.0417
	S_7	S_3^7, 0.0000	S_3^7, 0.0000	S_5^7, −0.0833	S_4^7, −0.0833
	S_8	S_0^7, 0.0000	S_3^7, 0.0000	S_5^7, 0.0417	S_3^7, −0.0417
	S_9	S_3^7, 0.0000	S_3^7, 0.0000	S_5^7, 0.0417	S_4^7, −0.0417
	S_{10}	S_3^7, 0.0000	S_4^7, 0.0000	S_4^7, −0.0417	S_4^7, −0.0694

Table 5: Continued.

Stock	Criterion	E_1	E_1	E_1	Average
	S_1	S_2^7, −0.0833	S_3^7, 0.0000	S_6^7, 0.0000	S_4^7, −0.0833
	S_2	S_0^7, 0.0000	S_3^7, 0.0000	S_4^7, −0.0417	S_2^7, 0.0417
	S_3	S_3^7, 0.0000	S_3^7, 0.0000	S_4^7, −0.0417	S_3^7, 0.0417
	S_4	S_3^7, 0.0000	S_5^7, 0.0000	S_5^7, −0.0833	S_4^7, 0.0278
C_4	S_5	S_5^7, −0.0833	S_4^7, 0.0000	S_5^7, 0.0417	S_5^7, −0.0694
	S_6	S_2^7, −0.0833	S_4^7, 0.0000	S_5^7, 0.0417	S_4^7, −0.0694
	S_7	S_3^7, 0.0000	S_3^7, 0.0000	S_5^7, 0.0417	S_4^7, −0.0417
	S_8	S_0^7, 0.0000	S_4^7, 0.0000	S_6^7, 0.0000	S_3^7, 0.0556
	S_9	S_3^7, 0.0000	S_4^7, 0.0000	S_5^7, 0.0417	S_4^7, 0.0139
	S_{10}	S_0^7, 0.0000	S_4^7, 0.0000	S_5^7, −0.0833	S_3^7, −0.0278
	S_1	S_3^7, 0.0000	S_5^7, 0.0000	S_5^7, 0.0417	S_4^7, 0.0694
	S_2	S_2^7, −0.0833	S_4^7, 0.0000	S_5^7, 0.0417	S_4^7, −0.0694
	S_3	S_2^7, −0.0833	S_4^7, 0.0000	S_5^7, −0.0833	S_3^7, 0.0556
	S_4	S_3^7, 0.0000	S_4^7, 0.0000	S_4^7, −0.0417	S_4^7, −0.0694
C_5	S_5	S_5^7, −0.0833	S_4^7, 0.0000	S_5^7, 0.0417	S_5^7, −0.0694
	S_6	S_3^7, 0.0000	S_3^7, 0.0000	S_5^7, −0.0833	S_4^7, −0.0833
	S_7	S_2^7, −0.0833	S_4^7, 0.0000	S_5^7, 0.0417	S_4^7, −0.0694
	S_8	S_0^7, 0.0000	S_4^7, 0.0000	S_5^7, 0.0417	S_3^7, 0.0139
	S_9	S_3^7, 0.0000	S_4^7, 0.0000	S_5^7, 0.0417	S_4^7, 0.0139
	S_{10}	S_0^7, 0.0000	S_3^7, 0.0000	S_4^7, −0.0417	S_2^7, 0.0417
	S_1	S_2^7, −0.0833	S_4^7, 0.0000	S_5^7, 0.0417	S_4^7, −0.0694
	S_2	S_2^7, −0.0833	S_4^7, 0.0000	S_5^7, −0.0833	S_3^7, 0.0556
	S_3	S_3^7, 0.0000	S_4^7, 0.0000	S_4^7, −0.0417	S_4^7, −0.0694
	S_4	S_2^7, −0.0833	S_6^7, 0.0000	S_4^7, −0.0417	S_4^7, −0.0417
C_6	S_5	S_2^7, −0.0833	S_5^7, 0.0000	S_5^7, −0.0833	S_4^7, −0.0556
	S_6	S_3^7, 0.0000	S_3^7, 0.0000	S_5^7, −0.0833	S_4^7, −0.0833
	S_7	S_3^7, 0.0000	S_4^7, 0.0000	S_5^7, 0.0417	S_4^7, 0.0139
	S_8	S_2^7, −0.0833	S_4^7, 0.0000	S_5^7, 0.0417	S_4^7, −0.0694
	S_9	S_3^7, 0.0000	S_5^7, 0.0000	S_5^7, −0.0833	S_4^7, 0.0278
	S_{10}	S_0^7, 0.0000	S_4^7, 0.0000	S_4^7, −0.0417	S_3^7, −0.0694

Table 6: Transfer to the linguistic variable of type 2.

Criterion	E_1	E_2	E_3	Average
C_1	S_6^7, 0.0000	S_6^7, 0.0000	S_6^7, 0.0000	S_6^7, 0.0000
C_2	S_5^7, −0.0833	S_5^7, 0.0000	S_6^7, 0.0000	S_5^7, 0.0278
C_3	S_5^7, −0.0833	S_5^7, 0.0000	S_5^7, 0.0417	S_5^7, −0.0139
C_4	S_5^7, −0.0833	S_4^7, 0.0000	S_6^7, 0.0000	S_5^7, −0.0278
C_5	S_6^7, 0.0000	S_5^7, 0.0000	S_5^7, 0.0417	S_5^7, 0.0694
C_6	S_3^7, 0.0000	S_5^7, 0.0000	S_5^7, −0.0833	S_4^7, 0.0278

Table 7: The weighted linguistic decision matrix.

	C_1	C_2	C_3	C_4	C_5	C_6
S_1	0.1148	0.1435	0.1231	0.0924	0.1307	0.0816
S_2	0.0902	0.1082	0.0940	0.0594	0.1061	0.0759
S_3	0.1148	0.1059	0.1030	0.0858	0.0987	0.0816
S_4	0.1284	0.1318	0.0963	0.1100	0.1061	0.0854
S_5	0.1421	0.1294	0.1119	0.1211	0.1357	0.0835
S_6	0.0984	0.1200	0.1008	0.0946	0.1036	0.0797
S_7	0.1311	0.1012	0.0940	0.0990	0.1061	0.0930
S_8	0.0929	0.1059	0.0739	0.0880	0.0913	0.0816
S_9	0.1503	0.1082	0.1008	0.1078	0.1209	0.0949
S_{10}	0.1038	0.1176	0.0963	0.0748	0.0666	0.0588

Table 8: Linguistic positive-ideal solution (LPIS, S^*) and linguistic negative-ideal solution (LNIS, S^-).

	C_1	C_2	C_3	C_4	C_5	C_6
S^*	0.1503	0.1435	0.1231	0.1211	0.1357	0.0949
S^-	0.0902	0.1012	0.0739	0.0594	0.0666	0.0588

Table 9: Calculate the distance from S^* and the distance from S^-, the closeness coefficient of each stock.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
d^*	0.0478	0.1036	0.0766	0.0492	0.0228	0.0755	0.0661	0.1018	0.0463	0.1084
d^-	0.1027	0.0480	0.0610	0.0879	0.1188	0.0647	0.0799	0.0444	0.1048	0.0346
CC	0.6826	0.3166	0.4432	0.6409	0.8388	0.4614	0.5471	0.3038	0.6937	0.2419

Table 10: The overall concordance matrix.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
S_1	1.0000	1.0000	1.0000	1.0000	0.9868	1.0000	1.0000	1.0000	0.9836	1.0000
S_2	0.9046	1.0000	1.0000	0.8219	0.7268	0.9472	0.8716	0.9868	0.7040	1.0000
S_3	0.9287	1.0000	1.0000	1.0000	0.9028	1.0000	1.0000	1.0000	0.9836	1.0000
S_4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
S_5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
S_6	1.0000	1.0000	1.0000	1.0000	0.9196	1.0000	1.0000	1.0000	0.8852	1.0000
S_7	0.9019	1.0000	1.0000	0.9859	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
S_8	0.7500	1.0000	0.9866	0.9836	0.7061	1.0000	0.9672	1.0000	0.8525	1.0000
S_9	0.9577	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
S_{10}	0.6257	0.7441	0.7885	0.6685	0.4954	0.7589	0.6758	0.8033	0.5360	0.8033

Step 5. Transform the linguistic evaluations of weight of each criterion into the linguistic variables of type 2 and aggregate the linguistic weight of each criterion as Table 6.

Step 6. Calculate the weighted linguistic decision matrix $V = [v_{ij}]_{m*n}$ as Table 7.

Step 7. Calculate the linguistic positive-ideal solution (LPIS, S^*) and linguistic negative-ideal solution (LNIS, S^-) as Table 8.

Table 11: The credibility matrix.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
S_1	1.0000	1.0000	1.0000	1.0000	0.9868	1.0000	1.0000	1.0000	0.9836	1.0000
S_2	0.9046	1.0000	1.0000	0.8219	0.4845	0.9472	0.8716	0.9868	0.7040	1.0000
S_3	0.9287	1.0000	1.0000	1.0000	0.9028	1.0000	1.0000	1.0000	0.9836	1.0000
S_4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
S_5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
S_6	1.0000	1.0000	1.0000	1.0000	0.9196	1.0000	1.0000	1.0000	0.8852	1.0000
S_7	0.9019	1.0000	1.0000	0.9859	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
S_8	0.7500	1.0000	0.9866	0.9836	0.7061	1.0000	0.9672	1.0000	0.8525	1.0000
S_9	0.9577	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
S_{10}	0.5213	0.7440	0.7884	0.6684	0.3302	0.7588	0.6757	0.8032	0.5359	0.8032

Table 12: The concordance credibility degree, the discordance credibility degree, the net credibility degree, and the outranking index.

Stock	$\phi^+(S_i)$	$\phi^-(S_i)$	$\phi(S_i)$	OTI
S_1	9.9704	8.9641	1.0063	0.5559
S_2	8.7206	9.7440	–1.0234	0.4431
S_3	9.8151	9.7750	0.0402	0.5022
S_4	10.0000	9.4599	0.5401	0.5300
S_5	10.0000	8.3300	1.6700	0.5928
S_6	9.8049	9.7060	0.0989	0.5055
S_7	9.8878	9.5145	0.3732	0.5207
S_8	9.2459	9.7900	–0.5441	0.4698
S_9	9.9577	8.9448	1.0128	0.5563
S_{10}	6.6292	9.8032	–3.1740	0.3237

Table 13: Compute the ratio of investment in accordance with the risk preference.

Rank	$P_t(S_i)$	$P_e(S_i)$	$P_{\mathrm{RA}}(S_i)$	$P_{\mathrm{RN}}(S_i)$	$P_{\mathrm{RL}}(s_i)$
1	S_5, 0.2465	S_5, 0.1575	S_5, 0.2225	S_5, 0.2308	S_5, 0.2385
2	S_9, 0.2038	S_9, 0.1478	S_9, 0.2088	S_9, 0.2028	S_9, 0.1973
3	S_1, 0.2006	S_1, 0.1477	S_1, 0.2075	S_1, 0.2012	S_1, 0.1952
4	S_4, 0.1883	S_4, 0.1408	S_4, 0.1948	S_4, 0.1903	S_4, 0.1861
5	S_7, 0.1608	S_7, 0.1384	S_7, 0.1663	S_7, 0.1749	S_7, 0.1829
6		S_6, 0.1343			
7		S_3, 0.1335			

Step 8. Calculate the distance of each stock from S^* and the distance from S^-, and the closeness coefficient of each stock as Table 9.

Step 9. Define investment threshold value as the average of the closeness coefficient $\beta_t = \sum_{i=1}^{n} \mathrm{CC}(S_i)/n$, so the investment portfolio set is $\Omega_t = \{S_1, S_4, S_5, S_7, S_9\}$ in accordance with TOPSIS. The ratio of investment based on TOPSISmethod is shown as Table 13.

Step 10. The indifference threshold, preference threshold, and veto threshold values of each criterion can be determined in accordance with the linguistic variables of type 2 as

$$q_j = \Delta^{-1}\left(S_1^7\right) - \Delta^{-1}\left(S_0^7\right) = \frac{1}{6}, \qquad p_j = \Delta^{-1}\left(S_2^7\right) - \Delta^{-1}\left(S_0^7\right) = \frac{2}{6},$$
$$v_j = \Delta^{-1}\left(S_3^7\right) - \Delta^{-1}\left(S_0^7\right) = \frac{3}{6}, \quad j = 1, \ldots, 6. \tag{4.1}$$

Step 11. Calculate the concordance matrix and the discordance matrix of each pair stock with respect to each criterion.Then, calculate the overall concordance matrix as Table 10 and the credibility matrix as Table 11.

Step 12. Calculate the concordance credibility degree, the discordance credibility degree, the net credibility degree, and the outranking index as Table 12.

Step 13. Define investment threshold value as the average of the outranking index $\beta_e = \sum_{i=1}^{n} \mathrm{OTI}(S_i)/n$, so the investment portfolio set is $\Omega_e = \{S_1, S_3, S_4, S_5, S_6, S_7, S_9\}$ in accordance with ELECTRE method. The ratio of investment based on ELECTRE method is shown as Table 13.

Step 14. Compute strict stock portfolio set as $\Omega_{ip} = \Omega_t \cap \Omega_e = \{S_1, S_4, S_5, S_7, S_9\}$.

Step 15. According to the investment preference of investor, the result of the ratio of investment based on combining linguistic ELECTRE with TOPSIS can be calculated as Table 13.

According to the result of numerical example, experts considered that the proposed method is useful to help investor determine the stock portfolio.

5. Conclusion

In general, the stock portfolio decision problem adheres to uncertain and imprecise data, and fuzzy set theory is adequate to deal with it. In this proposed model, different types of 2-tuple linguistic variables are applied to express the subjective judgment of each expert. Expert can easily express his opinion by different types of 2-tuple linguistic variables. The generalized translation method of different types of 2-tuple linguistic variables is applied to aggregate the subjective judgment of each expert. It is a flexible way to aggregate the opinions of all experts. Then, a new decision-making method has been presented in this paper by combining the advantages of ELECTRE with TOPSIS methods. According to the experts' opinions, the linguistic ELECTRE method and linguistic TOPSIS method are used to derive the closeness coefficient and the outranking index of each stock, respectively. Based on the closeness coefficient, the outranking index, and selection threshold, we can easily obtain three type of the investment ratio in accordance with different investment preference of final decision-maker. It is a reasonable way in real decision environment. In other words, the proposed method provides a flexible way to determine the stock portfolio under the uncertain environment. In the future, the concept of combing different decision methods for deciding stock portfolio will be applied to different fields such as R&D projects investment, bonus distribution in a company. A decision support system will be developed based on the proposed method for dealing with the stock selection problems in the future.

References

[1] X. Huang, "Portfolio selection with a new definition of risk," *European Journal of Operational Research*, vol. 186, no. 1, pp. 351–357, 2008.

[2] H. Markowitz, "Portfolio selection," *Journal of Finance*, pp. 77–91, 1952.

[3] C. A. Magni, "Correct or incorrect application of CAPM? Correct or incorrect decisions with CAPM?" *European Journal of Operational Research*, vol. 192, no. 2, pp. 549–560, 2009.

[4] F. Tiryaki and M. Ahlatcioglu, "Fuzzy stock selection using a new fuzzy ranking and weighting algorithm," *Applied Mathematics and Computation*, vol. 170, no. 1, pp. 144–157, 2005.

[5] E. F. Fama and K. R. French, "The cross-section of expected stock returns," *Journal of Finance*, vol. 47, pp. 427–465, 1992.

[6] C. D. Feinstein and M. N. Thapa, "Notes: a reformulation of a mean-absolute deviation portfolio optimization model," *Management Sciences*, vol. 39, no. 12, pp. 1552–1553, 1993.

[7] W. F. Sharpe, G. J. Alexander, and J. F. Bailey, *Investments*, Prentice-Hall, Englewood Cliffs, NJ, USA, 6th edition, 1999.

[8] T. L. Saaty, P. C. Rogers, and R. Bell, "Portfolio selection through hierarchies," *Journal of Portfolio Manage*, pp. 16–21, 1980.

[9] N. C. P. Edirisinghe and X. Zhang, "Generalized DEA model of fundamental analysis and its application to portfolio optimization," *Journal of Banking & Finance*, vol. 31, no. 11, pp. 3311–3335, 2007.

[10] F. Ben Abdelaziz, B. Aouni, and R. El Fayedh, "Multi-objective stochastic programming for portfolio selection," *European Journal of Operational Research*, vol. 177, no. 3, pp. 1811–1823, 2007.

[11] E. Ballestero, "Stochastic goal programming: a mean-variance approach," *European Journal of Operational Research*, vol. 131, no. 3, pp. 476–481, 2001.

[12] B. Aouni, F. Ben Abdelaziz, and J.-M. Martel, "Decision-maker's preferences modeling in the stochastic goal programming," *European Journal of Operational Research*, vol. 162, no. 3, pp. 610–618, 2005.

[13] C. L. Hwang and K. Yoon, *Multiple Attribute Decision Making Methods and Applications*, vol. 186 of *Lecture Notes in Economics and Mathematical Systems*, Springer, Berlin, Germany, 1981.

[14] A. Papadopoulos and A. Karagiannidis, "Application of the multi-criteria analysis method Electre III for the optimisation of decentralised energy systems," *Omega*, vol. 36, no. 5, pp. 766–776, 2008.

[15] C.-T. Chen, "Extensions of the TOPSIS for group decision-making under fuzzy environment," *Fuzzy Sets and Systems*, vol. 114, no. 1, pp. 1–9, 2000.

[16] F. Herrera and E. Herrera-Viedma, "Linguistic decision analysis: steps for solving decision problems under linguistic information," *Fuzzy Sets and Systems*, vol. 115, no. 1, pp. 67–82, 2000.

[17] F. Herrera and L. Martinez, "A 2-tuple fuzzy linguistic representation model for computing with words," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 6, pp. 746–752, 2000.

[18] Z. Xu, "Deviation measures of linguistic preference relations in group decision making," *Omega*, vol. 33, no. 3, pp. 249–254, 2005.

[19] E. Herrera-Viedma, O. Cordón, M. Luque, A. G. Lopez, and A. M. Muñoz, "A model of fuzzy linguistic IRS based on multi-granular linguistic information," *International Journal of Approximate Reasoning*, vol. 34, no. 2-3, pp. 221–239, 2003.

[20] F. Herrera, L. Martinez, and P. J. Sanchez, "Managing non-homogeneous information in group decision making," *European Journal of Operational Research*, vol. 166, no. 1, pp. 115–132, 2005.

[21] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[22] R. R. Yager, "An approach to ordinal decision making," *International Journal of Approximate Reasoning*, vol. 12, no. 3-4, pp. 237–261, 1995.

[23] C.-T. Lin and Y.-T. Chen, "Bid/no-bid decision-making—a fuzzy linguistic approach," *International Journal of Project Management*, vol. 22, no. 7, pp. 585–593, 2004.

[24] R.-C. Wang and S.-J. Chuu, "Group decision-making using a fuzzy linguistic approach for evaluating the flexibility in a manufacturing system," *European Journal of Operational Research*, vol. 154, no. 3, pp. 563–572, 2004.

[25] E. Herrera-Viedma and E. Peis, "Evaluating the informative quality of documents in SGML format from judgements by means of fuzzy linguistic techniques based on computing with words," *Information Processing and Management*, vol. 39, no. 2, pp. 233–249, 2003.

[26] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I," *Information Sciences*, vol. 8, pp. 199–249, 1975.

[27] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—II," *Information Sciences*, vol. 8, pp. 301–357, 1975.

[28] A. Kaufmann and M. M. Gupta, *Introduction to Fuzzy Arithmetic: Theory and Applications*, Van Nostrand Reinhold, New York, NY, USA, 1991.

[29] F. Herrera and L. Martinez, "A model based on linguistic 2-tuples for dealing with multigranular hierarchical linguistic contexts in multi-expert decision-making," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 31, no. 2, pp. 227–234, 2001.

[30] W.-S. Tai and C.-T. Chen, "A new evaluation model for intellectual capital based on computing with linguistic variable," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3483–3488, 2009.

[31] E. Herrera-Viedma, F. Herrera, L. Martínez, J. C. Herrera, and A. G. López, "Incorporating filtering techniques in a fuzzy linguistic multi-agent model for information gathering on the web," *Fuzzy Sets and Systems*, vol. 148, no. 1, pp. 61–83, 2004.

[32] H.-F. Li and J.-J. Wang, "An improved ranking method for ELECTRE III," in *International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '07)*, pp. 6659–6662, 2007.

Hindawi Publishing Corporation
Journal of Applied Mathematics and Decision Sciences
Volume 2009, Article ID 238196, 14 pages
doi:10.1155/2009/238196

Research Article

A Fuzzy Pay-Off Method for Real Option Valuation

Mikael Collan,[1] Robert Fullér,[1] and József Mezei[2]

[1] *IAMSR, Åbo Akademi University, Joukahaisenkatu 3-5 A, 20520 Turku, Finland*
[2] *Turku Centre for Computer Science, Joukahaisenkatu 3-5 B, 20520 Turku, Finland*

Correspondence should be addressed to Mikael Collan, mikael.collan@abo.fi

Received 20 November 2008; Revised 22 February 2009; Accepted 19 March 2009

Recommended by Lean Yu

Real option analysis offers interesting insights on the value of assets and on the profitability of investments, which has made real options a growing field of academic research and practical application. Real option valuation is, however, often found to be difficult to understand and to implement due to the quite complex mathematics involved. Recent advances in modeling and analysis methods have made real option valuation easier to understand and to implement. This paper presents a new method (fuzzy pay-off method) for real option valuation using fuzzy numbers that is based on findings from earlier real option valuation methods and from fuzzy real option valuation. The method is intuitive to understand and far less complicated than any previous real option valuation model to date. The paper also presents the use of number of different types of fuzzy numbers with the method and an application of the new method in an industry setting.

1. Introduction

Real option valuation is based on the observation that the possibilities financial options give their holder resemble the possibilities to invest in real investments and possibilities found within real investments, that is, managerial flexibility: "an irreversible investment opportunity is much like a financial call option" [1]. In other words, real option valuation is treating investment opportunities and the different types of managerial flexibility as options and valuing them with option valuation models. Real options are useful both, as a mental model for strategic and operational decision-making, and as a valuation and numerical analysis tool. This paper concentrates on the use of real options in numerical analysis, and particularly on the derivation of the real option value for a given investment opportunity, or identified managerial flexibility.

Real options are commonly valued with the same methods that have been used to value financial options, that is, with Black-Scholes option pricing formula [2], with the

binomial option valuation method [3], with Monte-Carlo-based methods [4], and with a number of later methods based on these. Most of the methods are complex and demand a good understanding of the underlying mathematics, issues that make their use difficult in practice. In addition these models are based on the assumption that they can quite accurately mimic the underlying markets as a process, an assumption that may hold for some quite efficiently traded financial securities, but may not hold for real investments that do not have existing markets or have markets that can by no means be said to exhibit even weak market efficiency.

Recently, a novel approach to real option valuation, called the Datar-Mathews method (DMM) was presented in [5–7], where the real option value is calculated from a pay-off distribution, derived from a probability distribution of the net present value (NPV) for a project that is generated with a (Monte-Carlo) simulation. The authors show that the results from the method converge to the results from the analytical Black-Scholes method. The method presented greatly simplifies the calculation of the real option value, making it more transparent and brings real option valuation as a method a big leap closer to practitioners. The most positive issue in the DMM is that it does not suffer from the problems associated with the assumptions connected to the market processes connected to the Black-Scholes and the binomial option valuation methods. The DMM utilizes cash-flow scenarios as an input to a Monte Carlo simulation to derive a distribution for the future investment outcomes. This distribution is then used to create a pay-off distribution for the investment. The DMM is highly compatible with the way cash-flow-based profitability analysis is commonly done in companies, because it can use the same type of inputs as NPV analysis.

All of the afore-mentioned models and methods use probability theory in their treatment of uncertainty, there are, however, other ways than probability to treat uncertainty, or imprecision in future estimates, namely, fuzzy logic and fuzzy sets. In classical set theory an element either (fully) belongs to a set or does not belong to a set at all. This type of bivalue, or true/false, logic is commonly used in financial applications (and is a basic assumption of probability theory). Bivalue logic, however, presents a problem, because financial decisions are generally made under uncertainty. Uncertainty in the financial investment context means that it is in practice impossible, exante to give absolutely correct precise estimates of, for example, future cash-flows. There may be a number of reasons for this, see, for example, [8], however, the bottom line is that our estimations about the future are imprecise.

Fuzzy sets are sets that allow (have) gradation of belonging, such as "a future cash flow at year ten is about x euro". This means that fuzzy sets can be used to formalize inaccuracy that exists in human decision making and as a representation of vague, uncertain, or imprecise knowledge, for example, future cash-flow estimation, which human reasoning is especially adaptive to. "Fuzzy set-based methodologies blur the traditional line between qualitative and quantitative analysis, since the modeling may reflect more the type of information that is available rather than researchers' preferences" [9], and indeed in economics "the use of fuzzy subsets theory leads to results that could not be obtained by classical methods" [10]. The origins of fuzzy sets date back to an article by Lotfi Zadeh [11] where he developed an algebra for what he called fuzzy sets. This algebra was created to handle imprecise elements in our decision-making processes, and is the formal body of theory that allows the treatment of practically all decisions in an uncertain environment. "Informally, a fuzzy set is a class of objects in which there is no sharp boundary between those objects that belong to the class and those that do not" [12].

In the following subsection we will shortly present fuzzy sets and fuzzy numbers and continue shortly on using fuzzy numbers in option valuation. We will then present a

new method for valuation of real options from fuzzy numbers that is based on the previous literature on real option valuation, especially the findings presented in [5] and on fuzzy real option valuation methods, we continue by illustrating the use of the method with a selection of different types of fuzzy numbers and with a case application of the new method in an industry setting, and close with a discussion and conclusions.

1.1. Fuzzy Sets and Fuzzy Numbers

A *fuzzy subset* A of a nonempty X set can be defined as a set of ordered pairs, each with the first element from X, and the second element from the interval $[0,1]$, with exactly one-ordered pair presents for each element of X. This defines a mapping,

$$\mu_A : A \rightarrow [0,1], \tag{1.1}$$

between elements of the set X and values in the interval $[0,1]$. The value zero is used to represent complete nonmembership, the value one is used to represent complete membership, and values in between are used to represent intermediate degrees of membership. The set X is referred to as the universe of discourse for the fuzzy subset A. Frequently, the mapping μ_A is described as a function, the membership function of A. The degree to which the statement $x \ \ is \ \ in \ \ A$ is true is determined by finding the ordered pair $(x, \mu_A(x))$. The degree of truth of the statement is the second element of the ordered pair. It is clear that A is completely determined by the set of tuples

$$A = \{(x, \mu_A(x)) \mid x \in X\}. \tag{1.2}$$

It should be noted that the terms *membership function* and *fuzzy subset* get used interchangeably and frequently we will write simply $A(x)$ instead of $\mu_A(x)$. A γ-level set (or γ-cut) of a fuzzy set A of X is a nonfuzzy set denoted by $[A]^\gamma$ and defined by

$$[A]^\gamma = \{t \in X \mid A(t) \geq \gamma\}, \tag{1.3}$$

if $\gamma > 0$ and cl(supp A) if $\gamma = 0$, where cl(supp A) denotes the closure of the support of A. A fuzzy set A of X is called *convex* if $[A]^\gamma$ is a convex subset of X for all $\gamma \in [0,1]$. A fuzzy number A is a fuzzy set of the real line with a normal, (fuzzy) convex, and continuous membership function of bounded support [13]. Fuzzy numbers can be considered as possibility distributions.

Definition 1.1. Let A be a fuzzy number. Then $[A]^\gamma$ is a closed convex (compact) subset of $\mathbb{R}$ for all $\gamma \in [0,1]$. Let us introduce the notations

$$a_1(\gamma) = \min[A]^\gamma, \quad a_2(\gamma) = \max[A]^\gamma \tag{1.4}$$

In other words, $a_1(\gamma)$ denotes the left-hand side and $a_2(\gamma)$ denotes the right-hand side of the γ-cut, $\gamma \in [0,1]$.

Definition 1.2. A fuzzy set A is called triangular fuzzy number with peak (or center) a, left width $\alpha > 0$ and right width $\beta > 0$ if its membership function has the following form

$$A(t) = \begin{cases} 1 - \dfrac{a-t}{\alpha} & \text{if } a-\alpha \le t \le a, \\ 1 - \dfrac{t-a}{\beta} & \text{if } a \le t \le \mathrm{a}+\beta, \\ 0 & \text{otherwise,} \end{cases} \tag{1.5}$$

and we use the notation $A = (a, \alpha, \beta)$. It can easily be verified that

$$[A]^{\gamma} = \left[a - (1-\gamma)\alpha, a + (1-\gamma)\beta\right], \quad \forall \gamma \in [0,1]. \tag{1.6}$$

The support of A is $(a-\alpha, b+\beta)$. A triangular fuzzy number with center a may be seen as a fuzzy quantity "x is approximately equal to a".

Definition 1.3. The *possibilistic (or fuzzy) mean value* of fuzzy number A with $[A]^{\gamma} = [a_1(\gamma), a_2(\gamma)]$ is defined in [13] by

$$\begin{aligned} E(A) &= \int_0^1 \frac{a_1(\gamma) + a_2(\gamma)}{2} 2\gamma \, \mathrm{d}\gamma \\ &= \int_0^1 \left(a_1(\gamma) + a_2(\gamma)\right)\gamma \, \mathrm{d}\gamma. \end{aligned} \tag{1.7}$$

Definition 1.4. A fuzzy set A is called trapezoidal fuzzy number with tolerance interval $[a, b]$, left width α, and right width β if its membership function has the following form:

$$A(t) = \begin{cases} 1 - \dfrac{a-t}{\alpha} & \text{if } a-\alpha \le t \le a, \\ 1 & \text{if } a \le t \le b, \\ 1 - \dfrac{t-b}{\beta} & \text{if } a \le t \le b+\beta, \\ 0 & \text{otherwise,} \end{cases} \tag{1.8}$$

and we use the notation

$$A = (a, b, \alpha, \beta). \tag{1.9}$$

It can easily be shown that $[A]^{\gamma} = [a - (1-\gamma)\alpha, b + (1-\gamma)\beta]$ for all $\gamma \in [0,1]$. The support of A is $(a-\alpha, b+\beta)$.

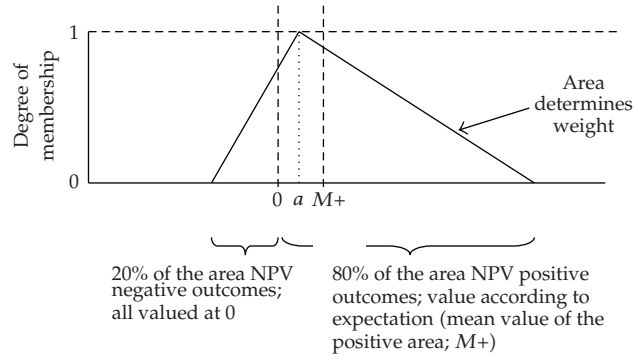


Figure 1: A triangular fuzzy number A, defined by three points $\{a, \alpha, \beta\}$ describing the NPV of a prospective project; (percentages 20% and 80% are for illustration purposes only).

Fuzzy set theory uses fuzzy numbers to quantify subjective fuzzy observations or estimates. Such subjective observations or estimates can be, for example, estimates of future cash flows from an investment. To estimate future cash flows and discount rates "one usually employs educated guesses, based on expected values or other statistical techniques" [14], which is consistent with the use of fuzzy numbers. In practical applications the most used fuzzy numbers are trapezoidal and triangular fuzzy numbers. They are used because they make many operations possible and are intuitively understandable and interpretable.

When we replace nonfuzzy numbers (crisp, single) numbers that are commonly used in financial models with fuzzy numbers, we can construct models that include the inaccuracy of human perception, or ability to forecast, within the (fuzzy) numbers. This makes these models more in line with reality, as they do not simplify uncertain distribution-like observations to a single-point estimate that conveys the sensation of no-uncertainty. Replacing nonfuzzy numbers with fuzzy numbers means that the models that are built must also follow the rules of fuzzy arithmetic.

1.2. Fuzzy Numbers in Option Valuation

Fuzzy numbers (fuzzy logic) have been adopted to option valuation models in (binomial) pricing an option with a fuzzy pay-off, for example, in [15], and in Black-Scholes valuation of financial options in, for example, [16]. There are also some option valuation models that present a combination of probability theory and fuzzy sets, for example, [17]. Fuzzy numbers have also been applied to the valuation of real options in, for example, [18–20]. More recently there are a number of papers that present the application of fuzzy real option models in the industry setting, for example, [21, 22]. There are also specific fuzzy models for the analysis of the value of optionality for very large industrial real investments, for example, [23].

2. New Fuzzy Pay-Off Method for Valuation of Real Options from Fuzzy Numbers

Two recent papers [5, 6] present a practical probability theory-based Datar-Mathews method for the calculation of real option value and show that the method and results from the method

are mathematically equivalent to the Black-Sholes formula [2]. The method is based on simulation-generated probability distributions for the NPV of future project outcomes. The project outcome probability distributions are used to generate a pay-off distribution, where the negative outcomes (subject to terminating the project) are truncated into one chunk that will cause a zero pay-off, and where the probability-weighted average value of the resulting pay-off distribution is the real option value. The DMM shows that the real-option value can be understood as the probability-weighted average of the pay-off distribution. We use fuzzy numbers in representing the expected future distribution of possible project costs and revenues, and hence also the profitability (NPV) outcomes. The fuzzy NPV, a fuzzy number, is the pay-off distribution from the project.

The method presented in [5] implies that the weighted average of the positive outcomes of the pay-off distribution is the real option value; in the case with fuzzy numbers the weighted average is the fuzzy mean value of the positive NPV outcomes. Derivation of the fuzzy mean value is presented in [13]. This means that calculating the ROV from a fuzzy NPV (distribution) is straightforward, it is the fuzzy mean of the possibility distribution with values below zero counted as zero, that is, the area-weighted average of the fuzzy mean of the positive values of the distribution and zero (for negative values)

Definition 2.1. We calculate the real option value from the fuzzy NPV as follows:

$$\mathrm{ROV} = \frac{\int_0^\infty A(x)\mathrm{d}x}{\int_{-\infty}^\infty A(x)\mathrm{d}x} \times E(A_+), \tag{2.1}$$

where A stands for the fuzzy NPV, $E(A_+)$ denotes the fuzzy mean value of the positive side of the NPV, and $\int_{-\infty}^\infty A(x)\mathrm{d}x$ computes the area below the whole fuzzy number A while $\int_0^\infty A(x)\mathrm{d}x$ computes the area below the positive part of A.

It is easy to see that when the whole fuzzy number is above zero, then ROV is the fuzzy mean of the fuzzy number, and when the whole fuzzy number is below zero, the ROV is zero.

The components of the new method are simply the observation that real option value is the probability-weighted average of the positive values of a pay-off distribution of a project, which is the fuzzy NPV of the project, and that for fuzzy numbers, the probability-weighted average of the positive values of the pay-off distribution is the weighted fuzzy mean of the positive values of the fuzzy NPV, when we use fuzzy numbers.

2.1. Calculating the ROV with the Fuzzy Pay-Off Method with a Selection of Different Types of Fuzzy Numbers

As the form of a fuzzy number may vary, the most used forms are the triangular and trapezoidal fuzzy numbers. These are very usable forms, as they are easy to understand and can be simply defined by three (triangular) and four (trapezoidal) values.

We should calculate the positive area and the fuzzy mean of the positive area of a triangular fuzzy pay-off $A = (a, \alpha, \beta)$ in the case of $a - \alpha < 0 < a$. Variable z, where $0 \le z \le \alpha$, represents the distance of a general cut point from $a - \alpha$ at which we separate the triangular fuzzy number (distribution) into two parts—for our purposes the variable z gets the value

$\alpha - a$ (we are interested in the positive part of A). Let us introduce the notation

$$(A \mid z)(\mathrm{t}) = \begin{cases} 0 & \text{if } t \leq a - \alpha + z, \\ A(t) & \text{otherwise,} \end{cases} \tag{2.2}$$

for the membership function of the right-hand side of a triangular fuzzy number truncated at point $a - \alpha + z$, where $0 \leq z \leq \alpha$.

Then we can compute the expected value of this truncated triangular fuzzy number:

$$E(A \mid z) = I_1 + I_2 = \int_0^{z_1} \gamma(a - \alpha + z + a + (1-\gamma)\beta)d\gamma + \int_{z_1}^{1} \gamma(a - (1-\gamma)\alpha + a + (1-\gamma)\beta)d\gamma, \tag{2.3}$$

where

$$z_1 = 1 - \frac{\alpha - z}{\alpha} = \frac{z}{\alpha}, \tag{2.4}$$

and the integrals are computed by

$$\begin{aligned} I_1 &= \int_0^{z_1} \left[(2a - \alpha + z + \beta)\gamma - \beta\gamma^2\right]d\gamma \\ &= (2a - \alpha + z + \beta)\frac{z^2}{2\alpha^2} - \beta\frac{z^3}{3\alpha^3}, \\ I_2 &= \int_{z_1}^{1} \left[(2a + \beta - \alpha)\gamma - \gamma^2(\beta - \alpha)\right]d\gamma \\ &= (2a + \beta - \alpha)\left(\frac{1}{2} - \frac{z^2}{2\alpha^2}\right) - g(\beta - \alpha)\left(\frac{1}{3} - \frac{z^3}{3\alpha^3}\right), \end{aligned} \tag{2.5}$$

that is,

$$\begin{aligned} I_1 + I_2 &= (2a - \alpha + z + \beta) \times \frac{z^2}{2\alpha^2} - \beta \times \frac{z^3}{3\alpha^3} + (2a + \beta - \alpha) \times \left(\frac{1}{2} - \frac{z^2}{2\alpha^2}\right) \\ &\quad - (\beta - \alpha) \times \left(\frac{1}{3} - \frac{z^3}{3\alpha^3}\right) = \frac{z^3}{2\alpha^2} + \frac{2a - \alpha + \beta}{2} + \frac{\alpha - \beta}{3} - \alpha \times \frac{z^3}{3\alpha^3}, \end{aligned} \tag{2.6}$$

and we get,

$$E(A \mid z) = \frac{z^3}{6\alpha^2} + a + \frac{\beta - \alpha}{6}. \tag{2.7}$$

If $z = \alpha - a$, then $A \mid z$ becomes A_+, the positive side of A, and therefore, we get

$$E(A_+) = \frac{(\alpha - a)^3}{6\alpha^2} + a + \frac{\beta - \alpha}{6}. \tag{2.8}$$

To compute the real option value with the afore-mentioned formulas we must calculate the ratio between the positive area of the triangular fuzzy number and the total area of the same number and multiply this by $E(A_+)$, the fuzzy mean value of the positive part of the fuzzy number A, according to (2.1).

For computing the real option value from an NPV (pay-off) distribution of a trapezoidal form we must consider a trapezoidal fuzzy pay-off distribution A defined by

$$A(u) = \begin{cases} \dfrac{u}{\alpha} - \dfrac{a_1 - \alpha}{\alpha} & \text{if } a_1 - \alpha \le u \le a_1, \\ 1 & \text{if } a_1 \le u \le a_2, \\ \dfrac{u}{-\beta} + \dfrac{a_2 + \beta}{\beta} & \text{if } a_2 \le u \le a_2 + \beta, \\ 0 & \text{otherwise,} \end{cases} \tag{2.9}$$

where the γ-level of A is defined by $[\mathrm{A}]^\gamma = [\gamma\alpha + a_1 - \alpha, -\gamma\beta + a_2 + \beta]$ and its expected value is caculated by

$$E(A) = \frac{a_1 + a_2}{2} + \frac{\beta - \alpha}{6}. \tag{2.10}$$

Then we have the following five cases.

Case 1. $z < a_1 - \alpha$. In this case we have $E(A \mid z) = E(A)$.

Case 2. $a_1 - \alpha < z < a_1$. Then introducing the notation

$$\gamma_z = \frac{z}{\alpha} - \frac{a_1 - \alpha}{\alpha}, \tag{2.11}$$

we find

$$[A]^\gamma = \begin{cases} (z, -\gamma\beta + a_2 + \beta) & \text{if } \gamma \le \gamma_z, \\ (\gamma\alpha + a_1 - \alpha, -\gamma\beta + a_2 + \beta) & \text{if } \gamma_z \le \gamma \le 1, \end{cases} \tag{2.12}$$

$$\begin{aligned} E(A \mid z) &= \int_0^{\gamma_z} \gamma\left(z - \gamma\beta + a_2 + \beta\right)d\gamma + \int_{\gamma_z}^1 \gamma\left(\gamma\alpha + a_1 - \alpha - \gamma\beta + a_2 + \beta\right)d\gamma \\ &= \frac{a_1 + a_2}{2} + \frac{\beta - \alpha}{6} + (z - a_1 + \alpha)\frac{\gamma_z^2}{2} - \alpha\frac{\gamma_z^3}{3}. \end{aligned} \tag{2.13}$$

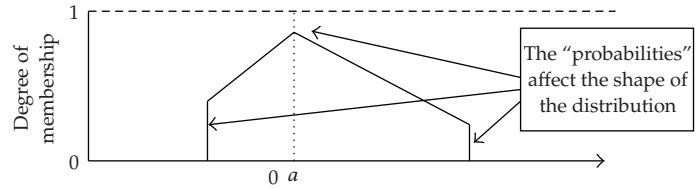


Figure 2: Calculation of the fuzzy mean for the positive part of a fuzzy pay-off distribution of the form of special case.

Case 3. $a_1 < z < a_2$. In this case $\gamma_z = 1$ and

$$[A]^{\gamma} = \left[z, -\gamma\beta + a_2 + \beta\right], \tag{2.14}$$

and we get

$$\begin{aligned} E(A \mid z) &= \int_0^1 \gamma\left(z - \gamma\beta + a_2 + \beta\right) d\gamma \\ &= \frac{z + a_2}{2} + \frac{\beta}{6}. \end{aligned} \tag{2.15}$$

Case 4. $a_2 < z < a_2 + \beta$. In this case we have

$$\gamma_z = \frac{z}{-\beta} + c\frac{a_2 + \beta}{\beta}, \tag{2.16}$$

$$[A]^{\gamma} = \left[z, -\gamma\beta + a_2 + \beta\right], \tag{2.17}$$

if $\gamma < \gamma_z$ and we find,

$$\begin{aligned} E(A \mid z) &= \int_0^{\gamma_z} \gamma\left(z - \gamma\beta + a_2 + \beta\right) d\gamma \\ &= \left(z + a_2 + \beta\right)\frac{\gamma_z^2}{2} - \beta\frac{\gamma_z^3}{3}. \end{aligned} \tag{2.18}$$

Case 5. $a_2 + \beta < z$. Then it is easy to see that $E(A \mid z) = 0$.

In the following special case, we expect that the managers will have already performed the construction of three cash-flow scenarios and have assigned estimated probabilities to each scenario (adding up to 100%). We want to use all this information and hence will assign the estimated "probabilities" to the scenarios resulting in a fuzzy number that has a graphical

presentation of the type presented in Figure 2 (not in scale):

$$A(u) = \begin{cases} (\gamma_3 - \gamma_1)\dfrac{u}{\alpha} - (\gamma_3 - \gamma_1)\dfrac{a-\alpha}{\alpha} + \gamma_1 & \text{if } a-\alpha \le u \le a, \\ \gamma_3 & \text{if } u = a, \\ (\gamma_2 - \gamma_3)\dfrac{u}{\beta} - (\gamma_2 - \gamma_3)\dfrac{a}{\beta} + \gamma_3 & \text{if } a \le u \le a+\beta, \\ 0 & \text{otherwise,} \end{cases}$$

$$\begin{aligned} E(A) &= \int_0^1 \gamma\big(a_1(\gamma) + a_2(\gamma)\big)d\gamma \\ &= \int_0^1 \gamma a_1(\gamma)d\gamma + \int_0^1 \gamma a_2(\gamma)d\gamma, \end{aligned}$$

$$\begin{aligned} \int_0^1 \gamma a_1(\gamma)d\gamma &= \int_0^{\gamma_1} \gamma(a-\alpha)d\gamma + \int_{\gamma_1}^{\gamma_3} \gamma\left(\frac{\gamma-\gamma_1}{\gamma_3-\gamma_1}\alpha + a - \alpha\right)d\gamma \\ &= (a-\alpha)\frac{\gamma_1^2}{2} + \left(a - \alpha - \frac{\alpha\gamma_1}{\gamma_3-\gamma_1}\right)\left(\frac{\gamma_3^2}{2} - \frac{\gamma_1^2}{2}\right) + \frac{\alpha}{\gamma_3-\gamma_1}\left(\frac{\gamma_3^3}{3} - \frac{\gamma_1^3}{3}\right), \end{aligned} \tag{2.19}$$

$$\begin{aligned} \int_0^1 \gamma a_2(\gamma)d\gamma &= \int_0^{\gamma_2} \gamma(a+\beta)d\gamma + \int_{\gamma_2}^{\gamma_3} \gamma\left(\frac{\gamma-\gamma_3}{\gamma_2-\gamma_3}\beta + a\right)d\gamma \\ &= (a+\beta)\frac{\gamma_2^2}{2} + \left(a - \frac{\beta\gamma_3}{\gamma_2-\gamma_3}\right)\left(\frac{\gamma_3^2}{2} - \frac{\gamma_2^2}{2}\right) + \frac{\beta}{\gamma_2-\gamma_3}\left(\frac{\gamma_3^3}{3} - \frac{\gamma_2^3}{3}\right), \end{aligned}$$

$$\begin{aligned} E(A) = {} & \frac{\gamma_1^2}{2}\frac{\alpha\gamma_1}{\gamma_3-\gamma_1} + \frac{\gamma_2^2}{2}\left(\beta + \frac{\beta\gamma_3}{\gamma_2-\gamma_3}\right) + \frac{\gamma_3^2}{2}\left(2a - \alpha - \frac{\alpha\gamma_1}{\gamma_3-\gamma_1} - \frac{\beta\gamma_3}{\gamma_2-\gamma_3}\right) \\ & - \frac{\gamma_1^3}{3}\frac{\alpha}{\gamma_3-\gamma_1} - \frac{\gamma_2^3}{3}\frac{\beta}{\gamma_2-\gamma_3} + \frac{\gamma_3^3}{3}\left(\frac{\alpha}{\gamma_3-\gamma_1} + \frac{\beta}{\gamma_2-\gamma_3}\right); \end{aligned}$$

(1) $z < a-\alpha : E(A \mid z) = E(A)$,

(2) $a-\alpha < z < a$: $\gamma_z = (\gamma_3 - \gamma_1)\dfrac{z}{\alpha} - (\gamma_3 - \gamma_1)\dfrac{a-\alpha}{\alpha} + \gamma_1$,

$$\begin{aligned} E(A \mid z) = {} & \frac{\gamma_z^2}{2}\left(z - a + \alpha + \frac{\alpha\gamma_1}{\gamma_3-\gamma_1}\right) + \frac{\gamma_2^2}{2}\left(\beta + \frac{\beta\gamma_3}{\gamma_2-\gamma_3}\right) \\ & + \frac{\gamma_3^2}{2}\left(2a - \alpha - \frac{\alpha\gamma_1}{\gamma_3-\gamma_1} - \frac{\beta\gamma_3}{\gamma_2-\gamma_3}\right) - \frac{\gamma_z^3}{3}\frac{\alpha}{\gamma_3-\gamma_1} \\ & - \frac{\gamma_2^3}{3}\frac{\beta}{\gamma_2-\gamma_3} + \frac{\gamma_3^3}{3}\left(\frac{\alpha}{\gamma_3-\gamma_1} + \frac{\beta}{\gamma_2-\gamma_3}\right), \end{aligned} \tag{2.20}$$

(3) $a < z < a + \beta : \gamma_z = (\gamma_2 - \gamma_3)\dfrac{z}{\beta} - (\gamma_2 - \gamma_3)\dfrac{a}{\beta} + \gamma_3$,

$$E(A \mid z) = \frac{\gamma_z^2}{2}\left(z + a - \frac{\beta}{\gamma_2 - \gamma_3}\right) + \frac{\gamma_2^2}{2}\left(\beta + \frac{\beta\gamma_3}{\gamma_2 - \gamma_3}\right) + \frac{\gamma_z^3}{3}\frac{\beta\gamma_3}{\gamma_2 - \gamma_3} - \frac{\gamma_2^3}{3}\frac{\beta}{\gamma_2 - \gamma_3}, \tag{2.21}$$

(4) $a + \beta < z : E(A \mid z) = 0$.

In the same way as was discussed earlier in connection to the triangular NPV, to compute the real option value with the afore-mentioned formulas we must calculate the ratio between the positive area of the fuzzy number (NPV) and the total area of the same number according to the formula (2.1).

3. A Simple Case: Using the New Method in Analyzing a Business Case

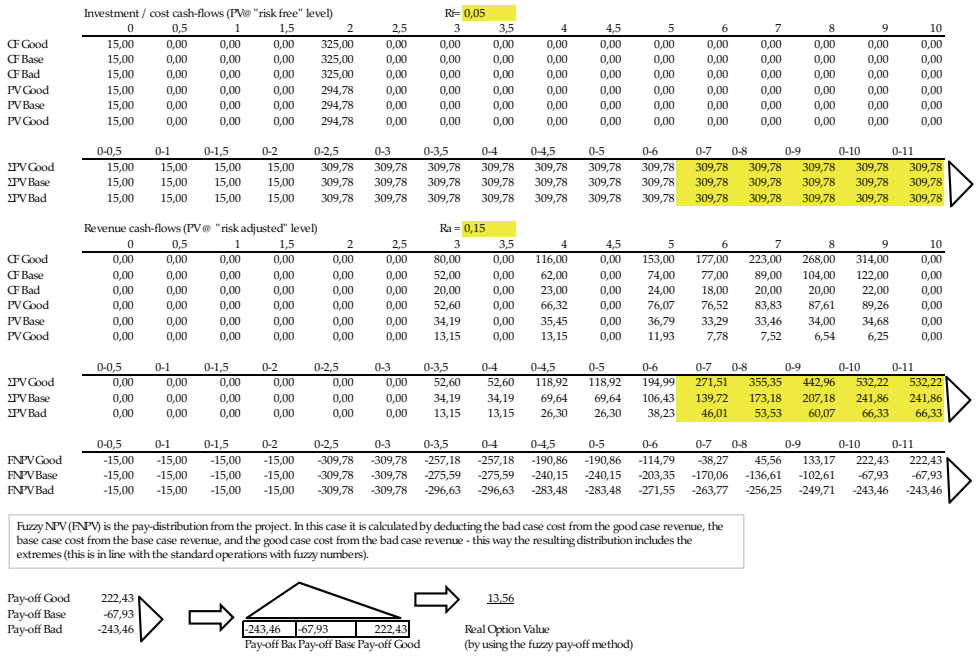
The problem at hand is to evaluate the value of uncertain cash-flows from a business case. The input information available is in the form of three future cash-flow scenarios, good (optimistic), most likely, and bad (pessimistic). The same business case with the same numbers has been earlier presented in [7] and is presented here to allow superficial comparison with the Datar-Mathews method—we are using the same numbers with the fuzzy pay-off method.

The scenario values are given by managers as nonfuzzy numbers, they can, in general, have used any type of analysis tools, or models to reach these scenarios. For more accurate information on the generation of the numbers in this case, see [7] for reference. From the cost and benefit scenarios three scenarios for the NPV are combined (PV benefits - PV investment costs), where the cost cash-flows (CF) are discounted at the risk-free rate and the benefit CF discount rate is selected according to the risk (risk adjusted discount rate). The NPV is calculated for each of the three scenarios separately, see Figures 3 and 4. The resulting fuzzy NPV is the fuzzy pay-off distribution for the investment. To reach a similar probability distribution [7] use Monte Carlo simulation. They point out that a triangular distribution can also be used. The real option value for the investment can be calculated from the resulting fuzzy NPV, which is the pay-off distribution for the project, according to the formula presented in (2.1). We use the formula described in Section 2.1. to calculate the real option value for this business case. We reach the value ROV = 13.56. The work in [7] shows that the value with the same inputs is 8. The difference is caused by the difference in the distributions generated from the inputs.

It is usual that managers are asked to give cash-flow information in the form of scenarios (usually three) and they often have a preselected set of methods for building the scenarios. Usually the scenarios are constructed by trusting past experience and based on looking at, for example, the variables that most contribute to cash-flows and the future market outlook; similar approaches are also reported in [7].

With the fuzzy pay-off method, the scenario approach can be fully omitted and the future cash-flow forecasting can be done fully with fuzzy numbers. The end result will be a fuzzy NPV that is the pay-off distribution for the project. This is the same result that we get if we use scenarios, however, it does not require us to simplify the future to three alternative scenarios.

The detailed calculation used in the case includes present value calculation for the three scenarios of investment cost and revenue cash-flows and then integrates these to form



Investment / cost cash-flows (PV@ "risk free" level) Rf= 0,05

	0	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	6	7	8	9	10
CF Good	15,00	0,00	0,00	0,00	325,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
CF Base	15,00	0,00	0,00	0,00	325,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
CF Bad	15,00	0,00	0,00	0,00	325,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
PV Good	15,00	0,00	0,00	0,00	294,78	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
PV Base	15,00	0,00	0,00	0,00	294,78	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
PV Good	15,00	0,00	0,00	0,00	294,78	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

	0-0,5	0-1	0-1,5	0-2	0-2,5	0-3	0-3,5	0-4	0-4,5	0-5	0-6	0-7	0-8	0-9	0-10	0-11
ΣPV Good	15,00	15,00	15,00	15,00	309,78	309,78	309,78	309,78	309,78	309,78	309,78	309,78	309,78	309,78	309,78	309,78
ΣPV Base	15,00	15,00	15,00	15,00	309,78	309,78	309,78	309,78	309,78	309,78	309,78	309,78	309,78	309,78	309,78	309,78
ΣPV Bad	15,00	15,00	15,00	15,00	309,78	309,78	309,78	309,78	309,78	309,78	309,78	309,78	309,78	309,78	309,78	309,78

Revenue cash-flows (PV@ "risk adjusted" level) Ra = 0,15

	0	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	6	7	8	9	10
CF Good	0,00	0,00	0,00	0,00	0,00	0,00	80,00	0,00	116,00	0,00	153,00	177,00	223,00	268,00	314,00	0,00
CF Base	0,00	0,00	0,00	0,00	0,00	0,00	52,00	0,00	62,00	0,00	74,00	77,00	89,00	104,00	122,00	0,00
CF Bad	0,00	0,00	0,00	0,00	0,00	0,00	20,00	0,00	23,00	0,00	24,00	18,00	20,00	20,00	22,00	0,00
PV Good	0,00	0,00	0,00	0,00	0,00	0,00	52,60	0,00	66,32	0,00	76,07	76,52	83,83	87,61	89,26	0,00
PV Base	0,00	0,00	0,00	0,00	0,00	0,00	34,19	0,00	35,45	0,00	36,79	33,29	33,46	34,00	34,68	0,00
PV Good	0,00	0,00	0,00	0,00	0,00	0,00	13,15	0,00	13,15	0,00	11,93	7,78	7,52	6,54	6,25	0,00

	0-0,5	0-1	0-1,5	0-2	0-2,5	0-3	0-3,5	0-4	0-4,5	0-5	0-6	0-7	0-8	0-9	0-10	0-11
ΣPV Good	0,00	0,00	0,00	0,00	0,00	0,00	52,60	52,60	118,92	118,92	194,99	271,51	355,35	442,96	532,22	532,22
ΣPV Base	0,00	0,00	0,00	0,00	0,00	0,00	34,19	34,19	69,64	69,64	106,43	139,72	173,18	207,18	241,86	241,86
ΣPV Bad	0,00	0,00	0,00	0,00	0,00	0,00	13,15	13,15	26,30	26,30	38,23	46,01	53,53	60,07	66,33	66,33

	0-0,5	0-1	0-1,5	0-2	0-2,5	0-3	0-3,5	0-4	0-4,5	0-5	0-6	0-7	0-8	0-9	0-10	0-11
FNPV Good	-15,00	-15,00	-15,00	-15,00	-309,78	-309,78	-257,18	-257,18	-190,86	-190,86	-114,79	-38,27	45,56	133,17	222,43	222,43
FNPV Base	-15,00	-15,00	-15,00	-15,00	-309,78	-309,78	-275,59	-275,59	-240,15	-240,15	-203,35	-170,06	-136,61	-102,61	-67,93	-67,93
FNPV Bad	-15,00	-15,00	-15,00	-15,00	-309,78	-309,78	-296,63	-296,63	-283,48	-283,48	-271,55	-263,77	-256,25	-249,71	-243,46	-243,46

Fuzzy NPV (FNPV) is the pay-distribution from the project. In this case it is calculated by deducting the bad case cost from the good case revenue, the base case cost from the base case revenue, and the good case cost from the bad case revenue - this way the resulting distribution includes the extremes (this is in line with the standard operations with fuzzy numbers).

Figure 3: Detailed calculations used in the case.

the fuzzy net present value (FNPV). The value of the R&D is directly included in the cost cash-flow table and the resulting ROV is what the work in [7] calls total project value. This is a minor issue, as the [7] project option value is the total project value + the R&D Cost.

4. Discussion and Conclusions

There is a reason to expect that the simplicity of the presented method is an advantage over more complex methods. Using triangular and trapezoidal fuzzy numbers makes very easy implementations possible with the most commonly used spreadsheet software; this opens avenues for real option valuation to find its way to more practitioners. The method is flexible as it can be used when the fuzzy NPV is generated from scenarios or as fuzzy numbers from the beginning of the analysis. Fuzzy NPV is a distribution of the possible values that can take place for NPV; this means that it is by definition perceived as impossible at the time of the assessment that values outside of the number can happen. This is in line with the situation that real option value is zero when all the values of the fuzzy NPV are lower than zero. If we compare this to the presented case, we can see that in practice it is often that managers are not interested to use the full distribution of possible outcomes, but rather want to limit their assessment to the most possible alternatives (and leaving out the tails of the distribution). We think that the tails should be included in the real option analysis, because even remote possibilities should be taken into consideration.

The method brings forth an issue that has not gotten very much attention in academia, the dynamic nature of the assessment of investment profitability, that is, the assessment

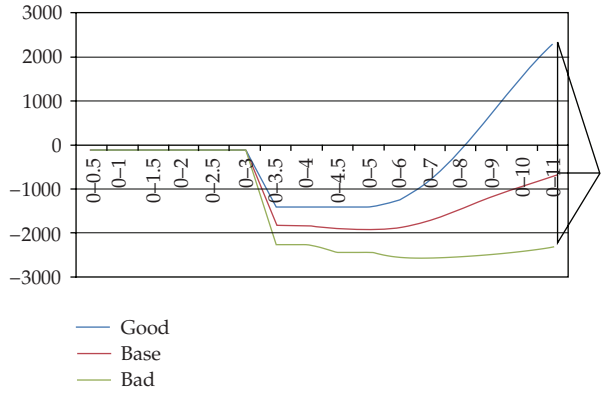


Figure 4: Three NPV scenarios for the duration of the synergies that are used to generate (triangular) fuzzy NPV.

changes when information changes. As cash flows taking place in the future come closer, information changes, and uncertainty is reduced this should be reflected in the fuzzy NPV, the more there is uncertainty the wider the distribution should be, and when uncertainty is reduced, the width of the distribution should decrease. Only under full certainty should the distribution be represented by a single number, as the method uses fuzzy NPV there is a possibility to have the size of the distribution decrease with a lesser degree of uncertainty, this is an advantage vis-à-vis probability-based methods.

The common decision rules for ROV analysis are applicable with the ROV derived with the presented method. We suggest that the single number NPV needed for comparison purposes is derived from the (same) fuzzy NPV by calculating the fuzzy mean value. This means that in cases when all the values of the fuzzy NPV are greater than zero, the single number NPV equals ROV, which indicates immediate investment.

We feel that the presented new method opens possibilities for making simpler generic and modular real option valuation tools that will help construct real options analyses for systems of real options that are present in many types of investments.

References

[1] R. Pindyck, "Irreversibility, uncertainty, and investment," *Journal of Economic Literature*, vol. 29, no. 3, pp. 1110–1148, 1991.

[2] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *Journal of Political Economy*, vol. 81, no. 3, pp. 637–659, 1973.

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

[4] P. P. Boyle, "Options: a Monte Carlo approach," *Journal of Financial Economics*, vol. 4, no. 3, pp. 323–338, 1977.

[5] V. Datar and S. Mathews, "A practical method for valuing real options: the boeing approach," *Journal of Applied Corporate Finance*, vol. 19, no. 2, pp. 95–104, 2007.

[6] S. Mathews and J. Salmon, "Business engineering: a practical approach to valuing high-risk, high-return projects using real options," in *Tutorials in Operations Research*, P. Gray, Ed., Informs, Hanover, Md, USA, 2007.

[7] V. Datar and S. Mathews, "European real options: an intuitive algorithm for the black scholes formula," *Journal of Applied Finance*, vol. 14, no. 1, pp. 45–51, 2004.

[8] F. Knight, *Risk, Uncertainty and Profit*, Hart, Schaffner & Marx, Boston, Mass, USA, 1921.

[9] M. Tarrazo, "A methodology and model for qualitative business planning," *International Journal of Business Research*, vol. 3, no. 1, pp. 41–62, 1997.
[10] C. Ponsard, "Fuzzy mathematical models in economics," *Fuzzy Sets and Systems*, vol. 28, no. 3, pp. 273–283, 1988.
[11] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.
[12] R. E. Bellman and L. A. Zadeh, "Decision-making in a fuzzy environment," *Management Science*, vol. 17, pp. B141–B164, 1970.
[13] C. Carlsson and R. Fullér, "On possibilistic mean value and variance of fuzzy numbers," *Fuzzy Sets and Systems*, vol. 122, no. 2, pp. 315–326, 2001.
[14] J. J. Buckley, "The fuzzy mathematics of finance," *Fuzzy Sets and Systems*, vol. 21, no. 3, pp. 257–273, 1987.
[15] S. Muzzioli and C. Torricelli, "A model for pricing an option with a fuzzy payoff," *Fuzzy Economics Review*, vol. 6, no. 1, pp. 49–62, 2000.
[16] Y. Yoshida, "The valuation of European options in uncertain environment," *European Journal of Operational Research*, vol. 145, no. 1, pp. 221–229, 2003.
[17] Z. Zmeškal, "Application of the fuzzy-stochastic methodology to appraising the firm value as a European call option," *European Journal of Operational Research*, vol. 135, no. 2, pp. 303–310, 2001.
[18] C. Carlsson and R. Fullér, "A fuzzy approach to real option valuation," *Fuzzy Sets and Systems*, vol. 139, no. 2, pp. 297–312, 2003.
[19] M. Collan, C. Carlsson, and P. Majlender, "Fuzzy black and scholes real options pricing," *Journal of Decision Systems*, vol. 12, no. 3-4, pp. 391–416, 2003.
[20] C. Carlsson and P. Majlender, "On fuzzy real option valuation," in *Proceedings of the 9th Annual International Conference on Real Options*, Paris, France, June 2005.
[21] T. Chen, J. Zhang, S. Lin, and B. Yu, "Fuzzy real option analysis for IT investment in nuclear power station," in *Proceedings of the 7th International Conference on Computational Science (ICCS '07)*, pp. 953–959, Beijing, China, May 2007.
[22] C. Tolga and C. Kahraman, "Fuzzy multiattribute evaluation of R&D projects using a real options valuation model," *International Journal of Intelligent Systems*, vol. 23, no. 11, pp. 1153–1176, 2008.
[23] M. Collan, *Giga-investments: modelling the valuation of very large industrial real investments*, Ph.D. thesis, Turku Centre for Computer Science, Turku, Finland, 2004.

Hindawi Publishing Corporation
Journal of Applied Mathematics and Decision Sciences
Volume 2009, Article ID 215163, 11 pages
doi:10.1155/2009/215163

Research Article

Valuation for an American Continuous-Installment Put Option on Bond under Vasicek Interest Rate Model

Guoan Huang,[1] Guohe Deng,[2] and Lihong Huang[3]

[1] *Department of Computer Science, Guilin College of Aerospace Technology, Guilin 541004, China*
[2] *School of Mathematics Science, Guangxi Normal University, Guilin 541004, China*
[3] *College of Mathematics and Econometrics, Hunan University, Changsha 410082, China*

Correspondence should be addressed to Guohe Deng, dengguohe@mailbox.gxnu.edu.cn

Received 2 December 2008; Accepted 6 March 2009

Recommended by Lean Yu

The valuation for an American continuous-installment put option on zero-coupon bond is considered by Kim's equations under a single factor model of the short-term interest rate, which follows the famous Vasicek model. In term of the price of this option, integral representations of both the optimal stopping and exercise boundaries are derived. A numerical method is used to approximate the optimal stopping and exercise boundaries by quadrature formulas. Numerical results and discussions are provided.

1. Introduction

Although there has been a large literature dealing with numerical methods for American options on stocks [1] and references cited therein, [2], there are not many papers for American options on default-free bonds, see, for example, [3–7], and so on. Numerical methods such as finite differences, binomial tree methods and Least-Square Monte Carlo simulations are still widely used. However, these methods have several shortcomings including time consuming, unbounded domain and discontinuous derivative with respect to the variate of payoff function. The most recent papers, like [8–11] provide different types of methods.

In this paper we consider an alternative form of American option in which the buyer pays a smaller up-front premium and then a constant stream of installments at a certain rate per unit time. So the buyer can choose at any time to stop making installment payments by either exercising the option or stopping the option contract. This option is called American continuous-installment (CI) option. Installment options are a recent financial innovation that

helps the buyer to reduce the cost of entering into a hedging strategy and the liquidity risk. Nowadays, the installment options are the most actively traded warrant throughout the financial world, such as the installment warrants on Australian stock and a 10-year warrant with 9 annual payments offered by Deutsche bank, and so on. There is very little literature on pricing the installment option, in particular, for pricing the American CI options. Ciurlia and Roko [12], and Ben-Ameur et al. [13] provide numerical procedures for valuing American CI options on stock under the geometric Brownian motion framework. However, in practice the option on bond is more useful than option on stock, and pricing the former is more complicated, because it is dependent on interest rates variable which is modelled by many economical models.

The aim of this paper is to present an approximation method for pricing American CI put option written on default-free, zero-coupon bond under Vasicek interest rate model. This method is based on Kim integral equations using quadrature formula approximations, such as the trapezoidal rule and the Simpson rule. The layout of this paper is as follows. Section 2 introduces the model and provides some preliminary results. In Section 3 we formulate the valuation problem for the American CI put option on bond describe as a free boundary problem and describe the Kim integral equations. Numerical method and results are presented in Section 4. Section 5 concludes.

2. The Model and Preliminary Results

In the one-factor Vasicek model [14], the short-term interest rate r_t is modeled as a mean-reverting Gaussian stochastic process on a probability space $(\Omega, \mathcal{F}, P)$ equipped with a filtration $(\mathcal{F}_t)_{t\geq 0}$. Under the the risk-neutral probability measure Q, it satisfies the linear stochastic differential equation (SDE)

$$dr_t = \kappa(r_\infty - r_t)dt + \sigma dW_t, \tag{2.1}$$

where $(W_t)_{t\geq 0}$ is a standard Q-Brownian motion, $\kappa > 0$ is the speed of mean reversion, $r_\infty > 0$ is the long-term value of interest rate, and σ is a constant volatility.

Consider a frictionless and no-arbitrage financial market which consists of a bank account A_t with its price process given by $dA_t = r_t A_t dt$ and a T_1-maturity default-free, zero-coupon bond $B(t, r, T_1) = B_t$ with its no-arbitrage price at time t given by

$$B(t, r, T_1) = E_Q\left\{e^{-\int_t^{T_1} r_u du}\mathcal{F}_t\right\} \widehat{=} E_Q^t\left[e^{-\int_t^{T_1} r_s ds}\right], \tag{2.2}$$

where E_Q is the expectation under the risk-neutral probability measure Q. Vasicek [14] provides the explicit form of the zero-bond as follows:

$$B(t, r, T_1) = a(T_1 - t)e^{-b(T_1-t)r_t} \tag{2.3}$$

with

$$a(u) = \exp\left\{-\left[R_\infty u - R_\infty b(u) + \frac{\sigma^2}{4\kappa}b^2(u)\right]\right\},$$
$$b(u) = \frac{1-e^{-\kappa u}}{\kappa}, \qquad R_\infty = r_\infty - \frac{\sigma^2}{2\kappa^2}. \tag{2.4}$$

From (2.3), we are easy to obtain the following partial differential equation (P.D.E.):

$$\frac{\partial B_t}{\partial t} + \kappa(r_\infty - r)\frac{\partial B_t}{\partial r} + \frac{1}{2}\sigma^2\frac{\partial^2 B_t}{\partial r^2} - rB_t = 0 \tag{2.5}$$

with terminal condition $B(T_1, r, T_1) = 1$.

The payoff of a European-style put option without paying any dividends written on the zero-coupon bond $B(t, r, T_1)$ with maturity T ($T < T_1$), and strike price K is $h(T, r) = \max\{K - B(T, r, T_1), 0\}$. The no-arbitrage price at time t ($0 \le t \le T$) of this option is denoted by $p_e(t, r, K; T)$. Following Jamshidian [15], the price of this option can generally be expressed as follows:

$$p_e(t, r, K; T) = E_Q^t\left[e^{-\int_t^T r_s ds}h(T, r)\right] = KB(t, r, T)N(-d_2) - B(t, r, T_1)N(-d_1), \tag{2.6}$$

where $N(\cdot)$ is the 1-dimensional standard cumulative normal distribution, and

$$d_{1,2} = \frac{1}{\sigma_0}\ln\frac{B(t, r, T_1)}{KB(t, r, T)} \pm \frac{1}{2}\sigma_0,$$
$$\sigma_0 = \sigma b(T_1 - T)\sqrt{\frac{1 - e^{-2\kappa(T-t)}}{2\kappa}}. \tag{2.7}$$

Now we consider a CI option written on the zero-coupon bond $B(t, r, T_1)$. Denote the initial premium of this option to be $V_t = V(t, r; q)$, which depends on the interest rate, time t, and the continuous-installment rate q. Applying Ito's Lemma to V_t, the dynamics for the initial value of this option is obtained as follows:

$$dV_t = \left[\frac{\partial V_t}{\partial t} + \frac{1}{2}\sigma^2\frac{\partial^2 V_t}{\partial r^2} + \kappa(r_\infty - r_t)\frac{\partial V_t}{\partial r} - q\right]dt + \sigma\frac{\partial V_t}{\partial r}dW_t. \tag{2.8}$$

Theorem 2.1. *In the Vasicek interest rates term structure model* (2.1)*. The contingent claim* $V(t, r; q)$ *satisfies the inhomogeneous partial differential equation*

$$\frac{\partial V_t}{\partial t} + \frac{1}{2}\sigma^2\frac{\partial^2 V_t}{\partial r^2} + \kappa(r_\infty - r_t)\frac{\partial V_t}{\partial r} - rV_t = q. \tag{2.9}$$

Proof. We now consider a self-financing trading strategy $\psi = (\psi_1, \psi_2)$, where ψ_1 and ψ_2 represent positions in bank account and T_1-maturity zero-coupon bonds, respectively. It is apparent that the wealth process π_t satisfies

$$\pi_t = \psi_1 A_t + \psi_2 B_t = V_t, \tag{2.10}$$

where the second equality is a consequence of the assumption that the trading strategy ψ replicate the option. Furthermore, since ψ is self-financing, its wealth process π_t also satisfies

$$d\pi_t = \psi_1 dA_t + \psi_2 dB_t, \tag{2.11}$$

so that

$$d\pi_t = \psi_1 r_t \ A_t dt + \psi_2 \left[\frac{\partial B_t}{\partial t} + \kappa(r_\infty - r_t) \frac{\partial B_t}{\partial r} + \frac{1}{2}\sigma^2 \frac{\partial^2 B_t}{\partial r^2} \right] dt + \sigma \psi_2 \frac{\partial B_t}{\partial r} dW_t. \tag{2.12}$$

From (2.8) and (2.10), we get

$$\begin{aligned} &\left[\frac{\partial V_t}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V_t}{\partial r^2} + \kappa(r_\infty - r_t) \frac{\partial V_t}{\partial r} - q - r V_t \right] dt + \sigma \frac{\partial V_t}{\partial r} dW_t \\ &\quad = \psi_2 \left[\frac{\partial B_t}{\partial t} + \kappa(r_\infty - r_t) \frac{\partial B_t}{\partial r} + \frac{1}{2}\sigma^2 \frac{\partial^2 B_t}{\partial r^2} - r B_t \right] dt + \psi_2 \sigma \frac{\partial B_t}{\partial r} dW_t. \end{aligned} \tag{2.13}$$

Setting $\psi_2 = (\partial B_t / \partial r)/(\partial V_t / \partial r)$ the coefficient of dW_t vanishes. It follows from (2.5) that, V_t satisfies (2.9). □

3. Kim Equations for the Price of American CI Put Option

Consider an American CI put option written on the zero-coupon bond B_t with the same strike price K and maturity time T ($T < T_1$). Although the underlying asset is the bond, the independent variable is the interest rate. Similar to American continuous-installment option on stock [12], there is an upper critical interest rate r_t^u above which it is optimal to stop the installment payments by exercising the option early, as well as a lower critical interest rate r_t^l below which it is advantageous to terminate payments by stopping the option contract. We may call r_t^u to be exercising boundary and r_t^l to be stopping boundary. Denote the initial premium of this put option at time t by $P(t, r; q) = P_t$, defined on the domain

$\mathfrak{D} = \{(r_t, t) \in [0, +\infty) \times [0, T]\}$. It is known that $P(t, r; q)$, r_t^u and r_t^l are the solution of the following free boundary problem [4]:

$$\begin{aligned}
&\frac{\partial P_t}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 P_t}{\partial r^2} + \kappa(r_\infty - r_t)\frac{\partial P_t}{\partial r} - rP_t = q, \quad \forall (r,t) \in \mathcal{C},\\
&P(t, r; q) = 0, \quad (r,t) \in \mathcal{S},\\
&P(t, r; q) = K - B(t, r, T_1), \quad (r,t) \in \mathcal{E},\\
&P(T, r; q) = h(T, r), \quad r \geq 0,\\
&P(t, r_t^u; q) = K - B(t, r_t^u, T_1),\\
&P\left(t, r_t^l; q\right) = 0, \quad t \in [0, T],
\end{aligned} \tag{3.1}$$

where $\mathcal{C} = \{(r_t, t) \in (r_t^l, r_t^u) \times [0, T)\}$ is a continuation region, $\mathcal{S} = \{((r_t, t) \in [0, r_t^l] \times [0, T]\}$ is a stopping region, and $\mathcal{E} = \{((r_t, t) \in [r_t^u, +\infty) \times [0, T]\}$ is a exercise region.

Remark 3.1. Due to the decreasing property of the price $B(t, r, T_1)$ on the state variable r, the strike price K should be strictly less than $B(T, 0, T_1)$. Otherwise, exercise would never be optimal.

It should be noted that although the value of the American CI put option has been expressed through the use of PDEs and their boundary conditions, there is still no explicit solution for the P.D.E. in (3.1). Numerical methods must be applied to value the price of the American CI option on bond. In the following we will solve this problem (3.1) with the integral equation method discussed in [8–12]. This method expresses the price of the American option as the sum of the price of the corresponding European option and the early exercise gains depending on the optimal exercise boundary. Jamshidian [3] uses this method to value the American bond option in Vasicek model.

Theorem 3.2. *Let the short interest rate r_t satisfy model* (2.1)*. Then the initial premium of the American CI put option , $P(t, r; q)$, can be written as*

$$\begin{aligned}
P(t, r; q) = {} & p_e(t, r, K; T) + q\int_t^T B(t, r, s) N\left(e\left(r, r_s^l\right)\right) ds\\
& + \int_t^T B(t, r, s)\left\{-qN(e(r, r_s^u)) + K[r_s^u - \sigma_1 e(r, r_s^u)]N(-e(r, r_s^u))\right.\\
& \left. + \frac{K\sigma_1}{\sqrt{2\pi}} \exp\left\{-\frac{e^2(r, r_s^u)}{2}\right\}\right\} ds.
\end{aligned} \tag{3.2}$$

Moreover, the optimal stopping and exercise boundaries, r^u and r^l, are solutions to the following system of recursive integral equations:

$$\begin{aligned}
K - B(t,r^u,T_1) &= p_e(t,r^u,K;T) + q\int_t^T B(t,r^u,s)N\left(c\left(r^u,r_s^l\right)\right)ds \\
&\quad + \int_t^T B(t,r^u,s)\left\{-qN(e(r^u,r_s^u)) + K[r_s^u - \sigma_1 e(r^u,r_s^u)]N(-e(r^u,r_s^u))\right. \\
&\qquad \left. + \frac{K\sigma_1}{\sqrt{2\pi}}\exp\left\{-\frac{e^2(r^u,r_s^u)}{2}\right\}\right\}ds, \\
0 &= p_e\left(t,r^l,K;T\right) + q\int_t^T B\left(t,r^l,s\right)N\left(c\left(r^l,r_s^l\right)\right)ds \\
&\quad + \int_t^T B\left(t,r^l,s\right)\left\{-qN\left(e\left(r^l,r_s^u\right)\right) + K\left[r_s^u - \sigma_1 e\left(r^l,r_s^u\right)\right]N\left(-e\left(r^l,r_s^u\right)\right)\right. \\
&\qquad \left. + \frac{K\sigma_1}{\sqrt{2\pi}}\exp\left\{-\frac{e^2(r^l,r_s^u)}{2}\right\}\right\}ds,
\end{aligned} \tag{3.3}$$

subject to the boundary conditions

$$B(T,r^u,T_1) = K, \qquad B\left(T,r^l,T_1\right) = K, \tag{3.4}$$

where $e(r,r_s^*) = ((r_s^* - r_t) - \kappa(r_t - r_\infty)b(s-t) + (1/2)\sigma^2 b^2(s-t))/\sigma_1$ *and* $\sigma_1^2 = (\sigma^2/2\kappa)(1-e^{-2\kappa(s-t)})$.

Proof. Let $Z(s,r) = e^{-\int_0^s r_u\,du}P(s,r;q)$ be the discounted initial premium function of the American CI put option in the domain $\mathfrak{D}$. It is known that the function $Z(s,r) \in C^{1,2}(\mathfrak{D})$. We can apply Ito Lemma to $Z(s,r)$ and write

$$Z(T,r) = Z(t,r) + \int_t^T\left[\frac{\partial Z(s,r)}{\partial s}ds + \frac{\partial Z(s,r)}{\partial r}dr + \frac{1}{2}\sigma^2\frac{\partial^2 Z(s,r)}{\partial r^2}ds\right]. \tag{3.5}$$

In terms of $P(t,r;q)$ this means

$$\begin{aligned}
e^{-\int_t^T r_s ds}P(T,r;q) &= P(t,r;q) + \int_t^T e^{-\int_t^s r_u du}\left[\frac{\partial P_s}{\partial s} + \frac{1}{2}\sigma^2\frac{\partial^2 P_s}{\partial r^2} + \kappa(r_\infty - r)\frac{\partial P_s}{\partial r} - rP_s\right]ds \\
&\quad + \int_t^T e^{-\int_t^s r_u du}\sigma\frac{\partial P_s}{\partial r}dW_s.
\end{aligned} \tag{3.6}$$

From (3.1) we know that $P(T,r;q) = h(T,r)$ and $P(s,r;q) = P(s,r;q)\mathbf{1}_{(r,s)\in\mathcal{C}} + P(s,r;q)\mathbf{1}_{(r,s)\in\mathcal{S}} + P(s,r;q)\cdot\mathbf{1}_{(r,s)\in\mathcal{E}} = P(s,r;q)\mathbf{1}_{(r,s)\in\mathcal{C}} + [K - B(s,r,T_1)]\mathbf{1}_{(r,s)\in\mathcal{E}}$. Substituting and taking expectation under Q on both sides of (3.6) give

$$\begin{aligned} p_e(t,r,K;T) &= E_Q^t\left[e^{-\int_t^T r_s ds} g(T,r)\right] \\ &= P(t,r;q) + \int_t^T E_Q^t\left\{e^{-\int_t^s r_u du}\left[\frac{\partial P_s}{\partial s} + \frac{1}{2}\sigma^2\frac{\partial^2 P_s}{\partial r^2} + \kappa(r_\infty - r)\frac{\partial P_s}{\partial r} - rP_s\right]\right\} ds \\ &= P(t,r;q) + q\int_t^T E_Q^t\left[e^{-\int_t^s r_u du} 1_{(r_s^l < r_s < r_s^u)}\right] ds - K\int_t^T E_Q^t\left\{e^{-\int_t^s r_u du} r_s 1_{(r_s \geq r_s^u)}\right\} ds. \end{aligned} \tag{3.7}$$

From (2.1), it is easy to obtain that the state variable r_s follows

$$r_s = r_t e^{-\kappa(s-t)} + r_\infty\left(1 - e^{-\kappa(s-t)}\right) + \sigma\int_t^s e^{-\kappa(s-u)} dW_u \tag{3.8}$$

for every $s > t$. Then the state variable r_s follows the normal distribution. Furthermore, using s-forward measures discussed in [16] and the normal distribution produces the representation (3.2). The recursive equations (3.3) for the optimal stopping and exercise boundaries are obtained by imposing the boundary conditions $P(t, r_t^u; q) = K - B(t, r_t^u, T_1)$ and $P(t, r_t^l; q) = 0$. The boundary conditions (3.4) hold since the limitation for (3.3) as $t \uparrow T$. □

Remark 3.3. From (3.2), when r_t^l and r_t^u are obtained by (3.3), the value of American CI put option is also derived. However, (3.3) are Volterra integral equations and can be solved numerically. Notice that the stopping and exercise boundary functions, r_t^l and r_t^u, cannot be proved to be monotone function of time t. So we use trapezoidal rule method to deal with them.

4. Numerical Method and Results

In this section we provide our method for pricing American CI put option by solving the Kim equations and present numerical results. This method consists of the following three steps. The first is to approximate the quadrature representations in (3.3) by using the trapezoidal rule. The second step is needed to find the numerical values of both the stopping and exercise boundaries, r_t^l and r_t^u from the equations approximated above with the Newton-Raphson (NR) iteration approach. When the values of r_t^l and r_t^u are obtained, the third step, numerical integration of (3.2), yields the value of a given American CI put option. This method is widely used to value American option by several authors, for example, [8, 11].

We now divide the time interval $[0,T]$ into N subintervals: $t_i = i\Delta t$, $i = 0,1,2,\ldots,N$, $\Delta t = T/N$. Denote $r_{t_i}^l = r_i^l$ and $r_{t_i}^u = r_i^u$ for $i = 0,1,2,\ldots,N$. Since $T_N = T$, we get by (2.3) and (3.4)

$$r_N^l = r_N^u = \frac{1}{b(T_1 - T)}\ln\frac{a(T_1 - T)}{K}. \tag{4.1}$$

We define the integrand of (3.3) as the following functions:

$$f(t,r,s,r_s^*) = B(t,r,s)\left\{-qN(e(r,r_s^*)) + K[r_s^* - \sigma_1 e(r,r_s^*)]N(-e(r,r_s^*)) + \frac{K\sigma_1}{\sqrt{2\pi}}\exp\left\{-\frac{e^2(r,r_s^*)}{2}\right\}\right\},$$
$$g(t,r,s,r_s^*) = qB(t,r,s)N(c(r,r_s^*)). \tag{4.2}$$

We use the trapezoidal rule to represent the system of recursive integral equations (3.3) as follows:

$$\begin{aligned}
&p(t_i,r_i^u,K;T) + \Delta t\left[\frac{1}{2}g\left(t_i,r_i^u,t_i,r_i^l\right) + \sum_{j=i+1}^{N-1} g\left(t_i,r_i^u,t_j,r_j^l\right) + \frac{1}{2}g\left(t_i,r_i^u,t_N,r_N^l\right)\right] \\
&\quad + \Delta t\left[\frac{1}{2}f(t_i,r_i^u,t_i,r_i^u) + \sum_{j=i+1}^{N-1} f\left(t_i,r_i^u,t_j,r_j^u\right) + \frac{1}{2}f(t_i,r_i^u,t_N,r_N^u)\right] \\
&\quad + B(t_i,r_i^u,T_1) - K = 0, \\
&p\left(t_i,r_i^l,K;T\right) + \Delta t\left[\frac{1}{2}g\left(t_i,r_i^l,t_i,r_i^l\right) + \sum_{j=i+1}^{N-1} g\left(t_i,r_i^l,t_j,r_j^l\right) + \frac{1}{2}g\left(t_i,r_i^l,t_N,r_N^l\right)\right] \\
&\quad + \Delta t\left[\frac{1}{2}f\left(t_i,r_i^l,t_i,r_i^u\right) + \sum_{j=i+1}^{N-1} f\left(t_i,r_i^l,t_j,r_j^u\right) + \frac{1}{2}f\left(t_i,r_i^l,t_N,r_N^u\right)\right] \\
&\quad = 0, \quad i = 0,\ldots,N-1.
\end{aligned} \tag{4.3}$$

Since there are nonlinear system equations, one can solve it using the NR iteration. In a similar way, numerical values of both r_i^l and r_i^u, $i = N-1, N-2, \ldots, 0$ can be obtained recursively from (4.3). We denote the representation of left side in (4.3) by $F_1(r_i^l, r_i^u)$ and $F_2(r_i^l, r_i^u)$, respectively. Then, by the NR iteration the values (r_i^l, r_i^u) have approximations $(r_i^l(k), r_i^u(k))$ of order k, where $k = 0, 1, 2, \ldots$

$$\begin{pmatrix} r_i^l(k+1) \\ r_i^u(k+1) \end{pmatrix} = \begin{pmatrix} r_i^l(k) \\ r_i^u(k) \end{pmatrix} - \begin{pmatrix} \dfrac{\partial F_1}{\partial x} & \dfrac{\partial F_1}{\partial y} \\ \dfrac{\partial F_2}{\partial x} & \dfrac{\partial F_2}{\partial y} \end{pmatrix}^{-1} \cdot \begin{pmatrix} F_1(x,y) \\ F_2(x,y) \end{pmatrix}\Big|_{(x,y)=(r_i^l(k),r_i^u(k))}, \tag{4.4}$$

where $\partial F_j/\partial x$ and $\partial F_j/\partial y$, $j = 1, 2$ are, respectively, partial derivatives of functions $F_j(x,y)$ with respect to x and y. When the values of all (r_i^l, r_i^u) for $i = N, \ldots, 0$ are obtained, using

Table 1: Value of parameters.

κ	r_∞	σ	K	T	T_1
0.05	0.083	0.015	0.95	1	5

Table 2: Initial premium of option on bond.

r_0	European put option	American CI put option		
		$q = 1$	$q = 10$	$q = 30$
0.04	0.0782	0.4409	0.3766	0.0893
0.10	0.2193	0.5928	0.4432	0.2296
0.15	0.3057	0.6556	0.4953	0.3258

Simpson's rule for (3.2) we get the approximation, $\widehat{P}_0(r,q)$, of the value at time $t = 0$ for the American CI put bond option in the following way: assuming N is an even number we have

$$\begin{aligned}\widehat{P}_0(r,q) = p(0,r,K;T) + \frac{\Delta t}{3}\Big[&g\left(0,r,0,r_0^l\right) + 4g\left(0,r,t_1,r_1^l\right) + 2g\left(0,r,t_2,r_2^l\right) \\ &+ 4g\left(0,r,t_3,r_3^l\right) + \cdots + 2g\left(0,r,t_{N-2},r_{N-2}^l\right) \\ &+4g\left(0,r,t_{N-1},r_{N-1}^l\right) + g\left(0,r,T,r_T^l\right)\Big] \\ + \frac{\Delta t}{3}\big[&f\left(0,r,0,r_0^u\right) + 4f\left(0,r,t_1,r_1^u\right) + 2f\left(0,r,t_2,r_2^u\right) + 4f\left(0,r,t_3,r_3^u\right) + \cdots \\ &+2f\left(0,r,t_{N-2},r_{N-2}^u\right) + 4f\left(0,r,t_{N-1},r_{N-1}^u\right) + f\left(0,r,T,r_T^u\right)\big].\end{aligned} \tag{4.5}$$

In Table 1, we describe the parameters in this section. In our example, we take $N = 6$. Table 2 provides the initial premium of this put option on bond for different installment rate $q = 1, 10$, and 30 with different initial interest rate $r_0 = 0.04, 0.10$, and 0.15.

Table 2 shows that the larger the initial interest rate is, the higher the price of American CI put option on bond is. However, the larger the installment rate is, the lower the price of this option is.

Figure 1 displays the curves of both the optimal stopping and exercise boundaries versus different installment rates q. We find out that the two boundaries decrease when the installment rate is arising. That shows that the larger the installment rate is, the higher probability the exercising of the option is.

5. Conclusions

A simple approximated method for pricing the American CI option written on the zero-bond under Vasicek model is proposed. Numerical example is provided to analyze the effects of the installment rate q on the price of this option and the optimal stopping and exercise boundaries. However, the Vasicek model allows for negative values of interest rate. This property is manifestly incompatible with reality. For this reason, work is ongoing to extend them to other models.

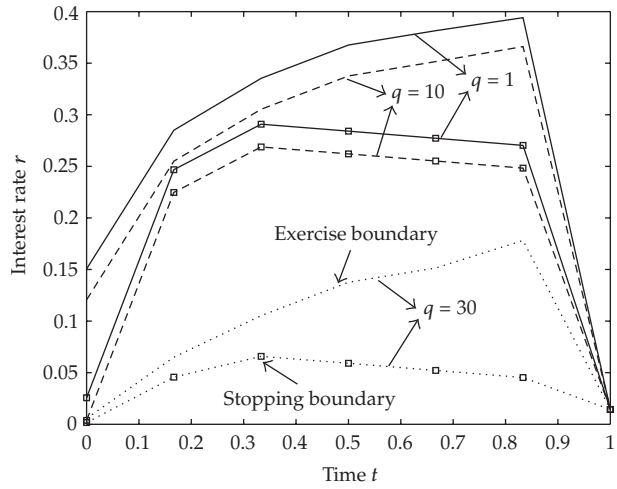


Figure 1: Optimal stopping and exercise boundaries for different installment rates.

Acknowledgment

The work has been partially supported by the NNSF of China with no. 40675023 and the Guangxi Natural Science Foundation with no. 0991091.

References

[1] W. Allegretto, Y. Lin, and H. Yang, "Finite element error estimates for a nonlocal problem in American option valuation," *SIAM Journal on Numerical Analysis*, vol. 39, no. 3, pp. 834–857, 2001.

[2] M. Broadie and J. B. Detemple, "Option pricing: valuation models and applications," *Management Science*, vol. 50, no. 9, pp. 1145–1177, 2004.

[3] F. Jamshidian, "An analysis of American options," *Review of Futures Markets*, vol. 11, no. 1, pp. 72–80, 1992.

[4] M. Chesney, R. Elliott, and R. Gibson, "Analytical solutions for the pricing of American bond and yield options," *Mathematical Finance*, vol. 3, no. 3, pp. 277–294, 1993.

[5] T. S. Ho, R. C. Stapleton, and M. G. Subrahmanyam, "The valuation of American options on bonds," *Journal of Banking & Finance*, vol. 21, no. 11-12, pp. 1487–1513, 1997.

[6] W. Allegretto, Y. Lin, and H. Yang, "Numerical pricing of American put options on zero-coupon bonds," *Applied Numerical Mathematics*, vol. 46, no. 2, pp. 113–134, 2003.

[7] L. ShuJin and L. ShengHong, "Pricing American interest rate option on zero-coupon bond numerically," *Applied Mathematics and Computation*, vol. 175, no. 1, pp. 834–850, 2006.

[8] I. J. Kim, "The analytic valuation of American options," *Review of Financial Studies*, vol. 3, no. 4, pp. 547–572, 1990.

[9] S. D. Jacka, "Optimal stopping and the American put," *Mathematical Finance*, vol. 1, no. 1, pp. 1–14, 1991.

[10] P. Carr, R. Jarrow, and R. Myneni, "Alternative characterizations of American put options," *Mathematical Finance*, vol. 2, no. 1, pp. 87–106, 1992.

[11] S. Kallast and A. Kivinukk, "Pricing and hedging American options using approximations by Kim integral equations," *European Finance Review*, vol. 7, no. 3, pp. 361–383, 2003.

[12] P. Ciurlia and I. Roko, "Valuation of American continuous-installment options," *Computational Economics*, vol. 25, no. 1-2, pp. 143–165, 2005.

[13] H. Ben-Ameur, M. Breton, and P. François, "A dynamic programming approach to price installment options," *European Journal of Operational Research*, vol. 169, no. 2, pp. 667–676, 2006.

[14] O. Vasicek, "An equilibrium characterization of the term structure," *Journal of Financial Economics*, vol. 5, no. 2, pp. 177–188, 1977.

[15] F. Jamshidian, "An exact bond option formula," *The Journal of Finance*, vol. 44, no. 1, pp. 205–209, 1989.

[16] H. Geman, N. El Karoui, and J.-C. Rochet, "Changes of numéraire, changes of probability measure and option pricing," *Journal of Applied Probability*, vol. 32, no. 2, pp. 443–458, 1995.

Hindawi Publishing Corporation
Journal of Applied Mathematics and Decision Sciences
Volume 2009, Article ID 593986, 13 pages
doi:10.1155/2009/593986

Research Article

Callable Russian Options and Their Optimal Boundaries

Atsuo Suzuki[1] and Katsushige Sawaki[2]

[1] *Faculty of Urban Science, Meijo University, 4-3-3 Nijigaoka, Kani, Gifu 509-0261, Japan*
[2] *Nanzan Business School, Nanzan University, 18 Yamazato-cho, Showa-ku, Nagoya 466-8673, Japan*

Correspondence should be addressed to Atsuo Suzuki, atsuo@urban.meijo-u.ac.jp

Received 28 November 2008; Accepted 10 February 2009

Recommended by Lean Yu

We deal with the pricing of callable Russian options. A callable Russian option is a contract in which both of the seller and the buyer have the rights to cancel and to exercise at any time, respectively. The pricing of such an option can be formulated as an optimal stopping problem between the seller and the buyer, and is analyzed as Dynkin game. We derive the value function of callable Russian options and their optimal boundaries.

1. Introduction

For the last two decades there have been numerous papers (see [1]) on valuing American-style options with finite lived maturity. The valuation of such American-style options may often be able to be formulated as optimal stopping or free boundary problems which provide us partial differential equations with specific conditions. One of the difficult problems with pricing such options is finding a closed form solution of the option price. However, there are shortcuts that make it easy to calculate the closed form solution to that option (see [2–4]). Perpetuities can provide us such a shortcut because free boundaries of optimal exercise policies no longer depend on the time.

In this paper, we consider the pricing of Russian options with call provision where the issuer (seller) has the right to call back the option as well as the investor (buyer) has the right to exercise it. The incorporation of call provision provides the issuer with option to retire the obligation whenever the investor exercises his/her option. In their pioneering theoretical studies on Russian options, Shepp and Shiryaev [5, 6] gave an analytical formula for pricing the noncallable Russian option which is one of perpetual American lookback options. The result of this paper is to provide the closed formed solution and optimal boundaries of

the callable Russian option with continuous dividend, which is different from the pioneering theoretical paper Kyprianou [2] in the sense that our model has dividend payment.

The paper is organized as follows. In Section 2, we introduce a pricing model of callable Russian options by means of a coupled optimal stopping problem given by Kifer [7]. Section 3 represents the value function of callable Russian options with dividend. Section 4 presents numerical examples to verify analytical results. We end the paper with some concluding remarks and future work.

2. Model

We consider the Black-Scholes economy consisting of two securities, that is, the riskless bond and the stock. Let B_t be the bond price at time t which is given by

$$dB_t = rB_t dt, \quad B_0 > 0, \quad r > 0, \tag{2.1}$$

where r is the riskless interest rate. Let S_t be the stock price at time t which satisfies the stochastic differential equation

$$dS_t = (r-d)S_t dt + \kappa S_t d\widetilde{W}_t, \quad S_0 = x, \tag{2.2}$$

where d and $\kappa > 0$ are constants, d is dividend rate, and $\widetilde{W}_t$ is a standard Brownian motion on a probability space $(\Omega, \mathcal{F}, \widetilde{P})$. Solving (2.2) with the initial condition $S_0 = x$ gives

$$S_t = x \exp\left\{\left(r - d - \frac{1}{2}\kappa^2\right)t + \kappa\widetilde{W}_t\right\}. \tag{2.3}$$

Define another probability measure $\widehat{P}$ by

$$\frac{d\widehat{P}}{d\widetilde{P}} = \exp\left(\kappa\widetilde{W}_t - \frac{1}{2}\kappa^2 t\right). \tag{2.4}$$

Let

$$\widehat{W}_t = \widetilde{W}_t - \kappa t, \tag{2.5}$$

where $\widehat{W}_t$ is a standard Brownian motion with respect to $\widehat{P}$. Substituting (2.5) into (2.2), we get

$$dS_t = \left(r - d + \kappa^2\right)S_t dt + \kappa S_t d\widehat{W}_t. \tag{2.6}$$

Solving the above equation, we obtain

$$S_t(x) = x \exp\left\{\left(r - d + \frac{1}{2}\kappa^2\right)t + \kappa\widehat{W}_t\right\}. \tag{2.7}$$

Russian option was introduced by Shepp and Shiryaev [5, 6] and is the contract that only the buyer has the right to exercise it. On the other hand, a callable Russian option is the contract that the seller and the buyer have both the rights to cancel and to exercise it at any time, respectively. Let σ be a cancel time for the seller and τ be an exercise time for the buyer. We set

$$\Psi_t(\psi) \equiv \frac{\max\left(\psi x, \sup_{0\le u\le t} S_u\right)}{S_t}, \quad \psi \ge 1. \tag{2.8}$$

When the buyer exercises the contract, the seller pay $\Psi_\tau(\psi)$ to the buyer. When the seller cancels it, the buyer receives $\Psi_\sigma(\psi) + \delta$. We assume that seller's right precedes buyer's one when $\sigma = \tau$. The payoff function of the callable Russian option is given by

$$R(\sigma,\tau) = \left(\Psi_\sigma(\psi) + \delta\right)1_{\{\sigma<\tau\}} + \Psi_\tau(\psi)1_{\{\tau\le\sigma\}}, \tag{2.9}$$

where δ is the penalty cost for the cancel and a positive constant.

Let $\mathcal{T}_{0,\infty}$ be the set of stopping times with respect to filtration $\mathcal{F}$ defined on the nonnegative interval. Letting α and ψ be some given parameters satisfying $\alpha > 0$ and $\psi \ge 1$, the value function of the callable Russian option $V(\psi)$ is defined by

$$V(\psi) = \inf_{\sigma\in\mathcal{T}_{0,\infty}} \sup_{\tau\in\mathcal{T}_{0,\infty}} \widehat{E}\left[e^{-\alpha(\sigma\wedge\tau)}R(\sigma,\tau)\right], \quad \alpha > 0. \tag{2.10}$$

The infimum and supremum are taken over all stopping times σ and τ, respectively.

We define two sets A and B as

$$\begin{aligned} A &= \left\{\psi \in \mathbf{R}^+ \mid V(\psi) = \psi + \delta\right\}, \\ B &= \left\{\psi \in \mathbf{R}^+ \mid V(\psi) = \psi\right\}. \end{aligned} \tag{2.11}$$

A and B are called the seller's cancel region and the buyer's exercise region, respectively. Let σ_A^ψ and τ_B^ψ be the first hitting times that the process $\Psi_t(\psi)$ is in the region A and B, that is,

$$\begin{aligned} \sigma_A^\psi &= \inf\left\{t > 0 \mid \Psi_t(\psi) \in A\right\}, \\ \tau_B^\psi &= \inf\left\{t > 0 \mid \Psi_t(\psi) \in B\right\}. \end{aligned} \tag{2.12}$$

Lemma 2.1. *Assume that* $d - (1/2)\kappa^2 - 2r < 0$. *Then, one has*

$$\lim_{t\to\infty} e^{-rt}\Psi_t(\psi) = 0. \tag{2.13}$$

Proof. First, suppose that $\max(\psi x, \sup S_u) = \psi x$. Then, it holds

$$\lim_{t\to\infty} e^{-rt}S_t^{-1} = \lim_{t\to\infty} \exp\left\{-\kappa\widetilde{W}_t + \left(d + \frac{1}{2}\kappa^2 - 2r\right)t\right\} = 0. \tag{2.14}$$

Next, suppose that $\max(\psi x, \sup S_u) = \sup S_u$. By the same argument as Karatzas and Shreve [1, page 65], we obtain

$$\begin{aligned}\lim_{t\to\infty} \sup S_u &= x \exp\left\{\kappa \cdot \sup_{0<u<\infty}\left(\widetilde{W}_u + \frac{r-d}{\kappa} - \frac{1}{2}\kappa^2\right)\right\} \\ &= x \exp\left\{\kappa W^*\right\},\end{aligned} \tag{2.15}$$

where W^* is the standard Brownian motion which attains the supremum in (2.15). Therefore, it follows that

$$\lim_{t\to\infty} e^{-rt}\frac{\sup S_u}{S_t} = 0. \tag{2.16}$$

The proof is complete. □

By this lemma, we may apply Proposition 3.3 in Kifer [7]. Therefore, we can see that the stopping times $\widehat{\sigma}^{\psi} = \sigma_A^{\psi}$ and $\widehat{\tau}^{\psi} = \tau_B^{\psi}$ attain the infimum and the supremum in (2.10). Then, we have

$$V(\psi) = \widehat{E}\left[e^{-\alpha(\widehat{\sigma}^{\psi}\wedge\widehat{\tau}^{\psi})}R\left(\widehat{\sigma}^{\psi}, \widehat{\tau}^{\psi}\right)\right]. \tag{2.17}$$

And $V(\psi)$ satisfies the inequalities

$$\psi \le V(\psi) \le \psi + \delta, \tag{2.18}$$

which provides the lower and the upper bounds for the value function of the callable Russian option. Let $V_R(\psi)$ be the value function of Russian option. And we know $V(\psi) \le V_R(\psi)$ because the seller as a minimizer has the right to cancel the option. Moreover, it is clear that $V(\psi)$ is increasing in ψ and x.

Should the penalty cost δ be large enough, it is optimal for the seller not to cancel the option. This raises a question how large such a penalty cost should be. The following lemma is to answer the question.

Lemma 2.2. *Set $\delta^* = V(1) - 1$. If $\delta \ge \delta^*$, the seller never cancels. Therefore, callable Russian options are reduced to Russian options.*

Proof. We set $h(\psi) = V(\psi) - \psi - \delta$. $h'(\psi) = V'(\psi) - 1 < 0$. Because we know $h(1) = V(1) - 1 - \delta = \delta^* - \delta < 0$ by the condition $\delta \ge \delta^*$, we have $h(\psi) < 0$, that is, $V(\psi) < \psi + \delta$ holds. By using the relation $V(\psi) \le V_R(\psi)$, we obtain $V(\psi) < \psi + \delta$, that is, it is optimal for the seller not to cancel. Therefore, the seller never cancels the contract for $\delta \ge \delta^*$. □

Lemma 2.3. *Suppose $r > d$. Then, the function $V(\psi)$ is Lipschitz continuous in ψ. And it holds*

$$0 \le \frac{dV(\psi)}{d\psi} \le 1. \tag{2.19}$$

Proof. Set

$$J^{\psi}\left(\widehat{\sigma}^{\psi},\widehat{\tau}^{\phi}\right)=\widehat{E}\left[e^{-\alpha\left(\widehat{\sigma}^{\psi}\wedge\widehat{\tau}^{\psi}\right)}R\left(\widehat{\sigma}^{\psi},\widehat{\tau}^{\psi}\right)\right]. \tag{2.20}$$

Replacing the optimal stopping times $\widehat{\sigma}^{\phi}$ and $\widehat{\tau}^{\psi}$ from the nonoptimal stopping times $\widehat{\sigma}^{\psi}$ and $\widehat{\tau}^{\phi}$, we have

$$\begin{aligned} V(\psi)&\geq J^{\psi}\left(\widehat{\sigma}^{\psi},\widehat{\tau}^{\phi}\right),\\ V(\phi)&\leq J^{\phi}\left(\widehat{\sigma}^{\psi},\widehat{\tau}^{\phi}\right), \end{aligned} \tag{2.21}$$

respectively. Note that $z_1^+-z_2^+\leq(z_1-z_2)^+$. For any $\phi>\psi$, we have

$$\begin{aligned} 0&\leq V(\phi)-V(\psi)\\ &\leq J^{\phi}\left(\widehat{\sigma}^{\psi},\widehat{\tau}^{\phi}\right)-J^{\psi}\left(\widehat{\sigma}^{\psi},\widehat{\tau}^{\phi}\right)\\ &=\widehat{E}\left[e^{-\alpha\left(\widehat{\sigma}^{\psi}\wedge\widehat{\tau}^{\phi}\right)}\left(\Psi_{\widehat{\sigma}^{\psi}\wedge\widehat{\tau}^{\phi}}(\phi)-\Psi_{\widehat{\sigma}^{\psi}\wedge\widehat{\tau}^{\phi}}(\psi)\right)\right]\\ &=\widehat{E}\left[e^{-\alpha\left(\widehat{\sigma}^{\psi}\wedge\widehat{\tau}^{\phi}\right)}H^{-1}_{\widehat{\sigma}^{\psi}\wedge\widehat{\tau}^{\phi}}\left(\left(\phi-\sup H_u\right)^+-\left(\psi-\sup H_u\right)^+\right)\right]\\ &\leq(\phi-\psi)\widehat{E}\left[e^{-\alpha\left(\widehat{\sigma}^{\psi}\wedge\widehat{\tau}^{\phi}\right)}H^{-1}_{\widehat{\sigma}^{\psi}\wedge\widehat{\tau}^{\phi}}\right], \end{aligned} \tag{2.22}$$

where $H_t=\exp\{(r-d+(1/2)\kappa^2)t+\kappa\widehat{W}_t\}$. Since the above expectation is less than 1, we have

$$0\leq V(\phi)-V(\psi)\leq\phi-\psi. \tag{2.23}$$

This means that V is Lipschitz continuous in ψ, and (2.19) holds. □

By regarding callable Russian options as a perpetual double barrier option, the optimal stopping problem can be transformed into a constant boundary problem with lower and upper boundaries. Let $\widetilde{B}=\{\psi\in\mathbf{R}^+\mid V_R(\psi)=\psi\}$ be the exercise region of Russian option. By the inequality $V(\psi)\leq V_R(\psi)$, it holds $B\supset\widetilde{B}\neq\varnothing$. Consequently, we can see that the exercise region B is the interval $[l_*,\infty)$. On the other hand, the seller minimizes $R(\sigma,\tau)$ and it holds $\Psi_t(\psi)\geq\Psi_0(\psi)=\psi\geq1$. From this, it follows that the seller's optimal boundary A is a point $\{1\}$. The function $V(\psi)$ is represented by

$$V(\psi)=\begin{cases}V_{\psi}(l_*), & 1\leq\psi\leq l_*,\\ \psi, & \psi\geq l_*,\end{cases} \tag{2.24}$$

where

$$V_{\psi}(l_*)=(1+\delta)\widehat{E}\left[e^{-\alpha\sigma_1^{\psi}}1_{\left\{\sigma_1^{\psi}<\tau^{\psi}_{[l_*,\infty)}\right\}}\right]+l\widehat{E}\left[e^{-\alpha\tau^{\psi}_{[l_*,\infty)}}1_{\left\{\tau^{\psi}_{[l_*,\infty)}\leq\sigma_1^{\psi}\right\}}\right]. \tag{2.25}$$

In order to calculate (2.25), we prepare the following lemma.

Lemma 2.4. *Let σ_a^x and τ_b^x be the first hitting times of the process $S_t(x)$ to the points $\{a\}$ and $\{b\}$. Set $\nu = (r-d)/\kappa - (1/2)\kappa$, $\eta_1 = (1/\kappa)(\sqrt{\nu^2+2\alpha}+\nu)$, and $\eta_2 = (1/\kappa)(\sqrt{\nu^2+2\alpha}-\nu)$. Then for $a<x<b$, one has*

$$\widetilde{E}\left[e^{-\alpha\sigma_a^x}1_{\{\sigma_a^x<\tau_b^x\}}\right] = \frac{(b/x)^{\eta_1}-(x/b)^{\eta_2}}{(b/a)^{\eta_1}-(a/b)^{\eta_2}}, \tag{2.26}$$

$$\widetilde{E}\left[e^{-\alpha\tau_b^x}1_{\{\tau_b^x<\sigma_a^x\}}\right] = \frac{(a/x)^{\eta_1}-(x/a)^{\eta_2}}{(a/b)^{\eta_1}-(b/a)^{\eta_2}}. \tag{2.27}$$

Proof. First, we prove (2.26). Define

$$L_t = \exp\left(-\frac{1}{2}\nu^2 t - \nu\widetilde{W}_t\right). \tag{2.28}$$

We define $\widehat{P}$ as $d\widehat{P} = L_T d\widetilde{P}$. By Girsanov's theorem, $\widehat{W}_t \equiv \widetilde{W}_t + \nu t$ is a standard Brownian motion under the probability measure $\widehat{P}$. Let T_{ρ_1} and T_{ρ_2} be the first time that the process $\widehat{W}_t$ hits ρ_1 or ρ_2, respectively, that is,

$$\begin{aligned} T_{\rho_1} &= \inf\left\{t>0 \mid \widehat{W}_t = \rho_1\right\}, \\ T_{\rho_2} &= \inf\left\{t>0 \mid \widehat{W}_t = \rho_2\right\}. \end{aligned} \tag{2.29}$$

Since we obtain $\log S_t(x) = \log x + \kappa\widehat{W}_t$ from $S_t(x) = x\exp(\kappa\widehat{W}_t)$, we have

$$\begin{aligned} \sigma_a^x &= T_{\rho_1}, \quad \text{a.s., } \rho_1 = \frac{1}{\kappa}\log\frac{a}{x}, \\ \tau_b^x &= T_{\rho_2}, \quad \text{a.s., } \rho_2 = \frac{1}{\kappa}\log\frac{b}{x}, \\ L_{T_{\rho_1}}^{-1} &= \exp\left(\frac{1}{2}\nu^2 T_{\rho_1} + \nu\widetilde{W}_{T_{\rho_1}}\right) \\ &= \exp\left(-\frac{1}{2}\nu^2 T_{\rho_1} + \nu\widehat{W}_{T_{\rho_1}}\right) \\ &= \exp\left(-\frac{1}{2}\nu^2 T_{\rho_1} + \nu\rho_1\right). \end{aligned} \tag{2.30}$$

Therefore, we have

$$\begin{aligned} \widetilde{E}\left[e^{-\alpha\sigma_a^x}1_{\{\sigma_a^x<\tau_b^x\}}\right] &= \widetilde{E}\left[e^{-\alpha T_{\rho_1}}1_{\{T_{\rho_1}<T_{\rho_2}\}}\right] \\ &= \widehat{E}\left[\exp\left(-\frac{1}{2}\nu^2 T_{\rho_1} + \nu\rho_1\right)e^{-\alpha T_{\rho_1}}1_{\{T_{\rho_1}<T_{\rho_2}\}}\right] \\ &= e^{\nu\rho_1}\widehat{E}\left[\exp\left\{-\left(\alpha+\frac{1}{2}\nu^2\right)T_{\rho_1}\right\}1_{\{T_{\rho_1}<T_{\rho_2}\}}\right]. \end{aligned} \tag{2.31}$$

From Karatzas and Shreve [8, Exercise 8.11, page 100], we can see that

$$\widehat{E}\left[\exp\left\{-\left(\alpha+\frac{1}{2}\nu^2\right)T_{\rho_1}\right\}1_{\{T_{\rho_1}<T_{\rho_2}\}}\right]=\frac{\sinh\rho_2\sqrt{\nu^2+2\alpha}}{\sinh(\rho_2-\rho_1)\sqrt{\nu^2+2\alpha}}. \tag{2.32}$$

Therefore, we obtain

$$\begin{aligned}&\widetilde{E}\left[e^{-\alpha\sigma_a^x}1_{\{\sigma_a^x<\tau_b^x\}}\right]\\&=\frac{\sinh\rho_2\sqrt{\nu^2+2\alpha}}{\sinh(\rho_2-\rho_1)\sqrt{\nu^2+2\alpha}}\frac{e^{\nu\rho_2}}{e^{\nu(\rho_2-\rho_1)}}=\frac{e^{\kappa\rho_2\gamma_1}-e^{-\kappa\rho_2\gamma_2}}{e^{\kappa(\rho_2-\rho_1)\gamma_1}-e^{-\kappa(\rho_2-\rho_1)\gamma_2}}=\frac{(b/x)^{\gamma_1}-(x/b)^{\gamma_2}}{(b/a)^{\gamma_1}-(a/b)^{\gamma_2}}.\end{aligned} \tag{2.33}$$

We omit the proof of (2.27) since it is similar to that of (2.26). □

We study the boundary point l_* of the exercise region for the buyer. For $1<\psi<l<\infty$, we consider the function $V(\psi,l)$. It is represented by

$$V(\psi,l)=\begin{cases}V_\psi(l), & 1\le\psi\le l,\\ \psi, & \psi\ge l.\end{cases} \tag{2.34}$$

The family of the functions $\{V(\psi,l),1<\psi<l\}$ satisfies

$$V(\psi)=V(\psi,l_*)=\sup_{1<\psi<l}V(\psi,l). \tag{2.35}$$

To get an optimal boundary point l_*, we compute the partial derivative of $V(\psi,l)$ with respect to l, which is given by the following lemma.

Lemma 2.5. *For any $1<\psi<l$, one has*

$$\frac{\partial V}{\partial l}(\psi,l)=\frac{\psi^{\eta_2}-\psi^{-\eta_1}}{l(l^{\eta_1}-l^{-\eta_2})^2}l^{\eta_1}l^{-\eta_2}\left\{(1-\eta_2)l^{\eta_1+1}-(1+\eta_1)l^{-\eta_2+1}+(1+\delta)(\eta_1+\eta_2)\right\}. \tag{2.36}$$

Proof. First, the derivative of the first term is

$$\begin{aligned}\frac{\partial}{\partial l}\left(\frac{(l/\psi)^{\eta_1}-(\psi/l)^{\eta_2}}{l^{\eta_1}-l^{-\eta_2}}\right)&=\frac{1}{(l^{\eta_1}-l^{-\eta_2})^2}\left\{\left(\eta_1\left(\frac{l}{\psi}\right)^{\eta_1-1}\frac{1}{\psi}+\eta_2\left(\frac{\psi}{l}\right)^{\eta_2}\frac{1}{l}\right)(l^{\eta_1}-l^{-\eta_2})\right.\\&\qquad\left.-\left(\left(\frac{l}{\psi}\right)^{\eta_1}-\left(\frac{\psi}{l}\right)^{\eta_2}\right)(\eta_1l^{\eta_1-1}+\eta_2l^{-\eta_2-1})\right\}\\&=\frac{1}{l(l^{\eta_1}-l^{-\eta_2})^2}(\eta_1+\eta_2)\left\{l^{\eta_1}\left(\frac{\psi}{l}\right)^{\eta_2}-l^{-\eta_2}\left(\frac{l}{\psi}\right)^{\eta_1}\right\}\\&=\frac{1}{l(l^{\eta_1}-l^{-\eta_2})^2}(\eta_1+\eta_2)l^{\eta_1}l^{-\eta_2}(\psi^{\eta_2}-\psi^{-\eta_1}).\end{aligned} \tag{2.37}$$

Next, the derivative of the second term is

$$\begin{aligned}\frac{\partial}{\partial l}\left(\frac{l}{l^{\eta_2}-l^{-\eta_1}}\right) &= \frac{(1-\eta_2)l^{\eta_2}-(1+\eta_1)l^{-\eta_1}}{\left(l^{\eta_2}-l^{-\eta_1}\right)^2} \\ &= \frac{(1-\eta_2)l^{\eta_1}-(1+\eta_1)l^{-\eta_2}}{\left(l^{\eta_1}-l^{-\eta_2}\right)^2}l^{\eta_1-\eta_2},\end{aligned} \tag{2.38}$$

where the last equality follows from the relation

$$\left(l^{\eta_2}-l^{-\eta_1}\right)l^{\eta_1-1}l^{-\eta_2+1} = l^{\eta_1}-l^{-\eta_2}. \tag{2.39}$$

After multiplying (2.37) by $(1+\delta)$ and (2.38) by $\varphi^{\eta_2}-\varphi^{-\eta_1}$, we obtain (2.36). □

We set

$$f(l) = (1-\eta_2)l^{\eta_1+1} - (1+\eta_1)l^{-\eta_2+1} + (1+\delta)(\eta_1+\eta_2). \tag{2.40}$$

Since $f(1) = \delta(\eta_1+\eta_2) > 0$ and $f(\infty) = -\infty$, the equation $f(l) = 0$ has at least one solution in the interval $(1,\infty)$. We label all real solutions as $1 < l_n < l_{n-1} < \cdots < l_1 < \infty$. Then, we have

$$\frac{\partial V}{\partial l}(\varphi, l)\Big|_{l=l_i} = 0, \quad i = 1,\ldots,n \ \forall\varphi. \tag{2.41}$$

Then $l_* = l_1$ attains the supremum of $V(\varphi, l)$. In the following, we will show that the function $V(\varphi)$ is convex and satisfies smooth-pasting condition.

Lemma 2.6. *$V(\varphi)$ is a convex function in φ.*

Proof. From (2.50), V satisfies

$$\frac{1}{2}\kappa^2\varphi^2\frac{d^2V}{d\varphi^2} = -(r-d)\varphi\frac{dV}{d\varphi} + \alpha V(\varphi). \tag{2.42}$$

If $r \le d$, we get $d^2V/d\varphi^2 > 0$. Next assume that $r > d$. We consider function $\widetilde{V}(\varphi) = V(-\varphi)$ for $\varphi < 0$. Then,

$$\frac{1}{2}\kappa^2\varphi^2\frac{d^2\widetilde{V}}{d\varphi^2} - (r-d)\varphi\frac{d\widetilde{V}}{d\varphi} - r\widetilde{V} = \frac{1}{2}\kappa^2\varphi^2\frac{d^2V}{d\varphi^2} + (r-d)\varphi\frac{dV}{d\varphi} - rV = 0. \tag{2.43}$$

Since we find that $d^2\widetilde{V}/d\varphi^2 > 0$ from the above equation, $\widetilde{V}$ is a convex function. It follows from this the fact that V is a convex function. □

Lemma 2.7. *$V(\psi)$ satisfies*

$$\frac{dV}{d\psi}(l_*^-) = \frac{dV}{d\psi}(l_*^+) = 1. \tag{2.44}$$

Proof. Since $V(\psi) = \psi$ for $\psi > l_*$, it holds $(dV/d\psi)(l_*^+) = 1$. For $1 \le \psi < l_*$, we derivative (2.47):

$$\begin{aligned}
\frac{dV}{d\psi} &= \frac{l}{l^{\eta_2} - l^{-\eta_1}}\left(\eta_2\psi^{\eta_2-1} + \eta_1 l^{-\eta_1-1}\right) + \frac{1+\delta}{l^{\eta_1} - l^{-\eta_2}}\left(-\eta_1\left(\frac{l}{\psi}\right)^{\eta_1}\frac{1}{\psi} - \eta_2\left(\frac{\psi}{l}\right)^{\eta_2}\frac{1}{\psi}\right) \\
&= \frac{1}{\psi(l^{\eta_1} - l^{-\eta_2})}\left\{l^{\eta_1-\eta_2+1}\left(\eta_2\psi^{\eta_2} + \eta_1\psi^{-\eta_1}\right) - (1+\delta)\left(\eta_1\left(\frac{l}{\psi}\right)^{\eta_1} + \eta_2\left(\frac{\psi}{l}\right)^{\eta_2}\right)\right\} \\
&= \frac{1}{\psi(l^{\eta_1} - l^{-\eta_2})}\left\{\eta_2\left(\frac{\psi}{l}\right)^{\eta_2} l^{\eta_1+1} + \eta_1\left(\frac{l}{\psi}\right)^{\eta_1} l^{-\eta_2+1} - (1+\delta)\left(\eta_1\left(\frac{l}{\psi}\right)^{\eta_1} + \eta_2\left(\frac{\psi}{l}\right)^{\eta_2}\right)\right\}.
\end{aligned} \tag{2.45}$$

Therefore, we get

$$\begin{aligned}
\frac{dV}{d\psi}(l_*) - 1 &= \frac{1}{\left(l_*^{\eta_1+1} - l_*^{1-\eta_2}\right)}\left\{\eta_2 l_*^{\eta_1+1} + \eta_1 l_*^{-\eta_2+1} - (1+\delta)(\eta_1+\eta_2) - \left(l_*^{\eta_1+1} - l_*^{-\eta_2+1}\right)\right\} \\
&= \frac{1}{\left(l_*^{\eta_1} - l_*^{-\eta_2}\right)}\left\{(\eta_2 - 1)l_*^{\eta_1+1} + (\eta_1+1)l_*^{-\eta_2+1} - (1+\delta)(\eta_1+\eta_2)\right\} \\
&= \frac{1}{\left(l_*^{\eta_1} - l_*^{-\eta_2}\right)} f(l_*) \\
&= 0.
\end{aligned} \tag{2.46}$$

This completes the proof. □

Therefore, we obtain the following theorem.

Theorem 2.8. *The value function of callable Russian option $V(\psi)$ is given by*

$$V(\psi) = \begin{cases} (1+\delta)\dfrac{(l_*/\psi)^{\eta_1} - (\psi/l_*)^{\eta_2}}{l_*^{\eta_1} - l_*^{-\eta_2}} + l_*\dfrac{\psi^{\eta_2} - \psi^{-\eta_1}}{l_*^{\eta_2} - l_*^{-\eta_1}}, & 1 \le \psi \le l_*, \\ \psi, & \psi \ge l_*. \end{cases} \tag{2.47}$$

And the optimal stopping times are

$$\begin{aligned}
\widehat{\sigma}^{\psi} &= \inf\left\{t > 0 \mid \Psi_t(\psi) = 1\right\}, \\
\widehat{\tau}^{\psi} &= \inf\left\{t > 0 \mid \Psi_t(\psi) \ge l_*\right\}.
\end{aligned} \tag{2.48}$$

The optimal boundary for the buyer l_* is the solution in $(1,\infty)$ to $f(l)=0$, where

$$f(l)=(1-\eta_2)l^{\eta_1+1}-(1+\eta_1)l^{-\eta_2+1}+(1+\delta)(\eta_1+\eta_2). \tag{2.49}$$

We can get (2.47) by another method. For $1<\psi<l$, the function $V(\psi)$ satisfies the differential equation

$$\frac{1}{2}\kappa^2\psi^2\frac{d^2V}{d\psi^2}+(r-d)\psi\frac{dV}{d\psi}-\alpha V(\psi)=0. \tag{2.50}$$

Also, we have the boundary conditions as follows:

$$V(1)=C_1+C_2=1+\delta, \tag{2.51}$$

$$V(l)=C_1l^{\lambda_1}+C_2l^{\lambda_2}=l, \tag{2.52}$$

$$V'(l)=C_1\lambda_1l^{\lambda_1-1}+C_2\lambda_2l^{\lambda_2-1}=1. \tag{2.53}$$

The general solution to (2.50) is represented by

$$V(\psi)=C_1\psi^{\lambda_1}+C_2\psi^{\lambda_2}, \tag{2.54}$$

where C_1 and C_2 are constants. Here, λ_1 and λ_2 are the roots of

$$\frac{1}{2}\kappa^2\lambda^2+\left(r-d-\frac{1}{2}\kappa^2\right)\lambda-\alpha=0. \tag{2.55}$$

Therefore, λ_1,λ_2 are

$$\lambda_{1,2}=\frac{\pm\sqrt{\nu^2+2\alpha}-\nu}{\kappa}. \tag{2.56}$$

From conditions (2.51) and (2.52), we get

$$C_1=\frac{l-(\delta+1)l^{\lambda_2}}{l^{\lambda_1}-l^{\lambda_2}},\qquad C_2=\frac{(\delta+1)l^{\lambda_1}-l}{l^{\lambda_1}-l^{\lambda_2}}. \tag{2.57}$$

And from (2.57) and (2.53), we have

$$(1-\eta_2)l^{\eta_1+1}-(1+\eta_1)l^{-\eta_2+1}+(1+\delta)(\eta_1+\eta_2)=0. \tag{2.58}$$

Substituting (2.57) into (2.54), we can obtain (2.47).

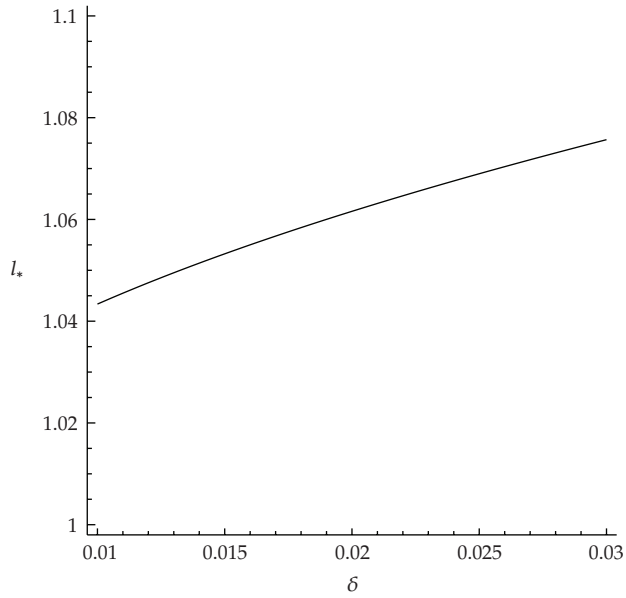


Figure 1: Optimal boundary for the buyer.

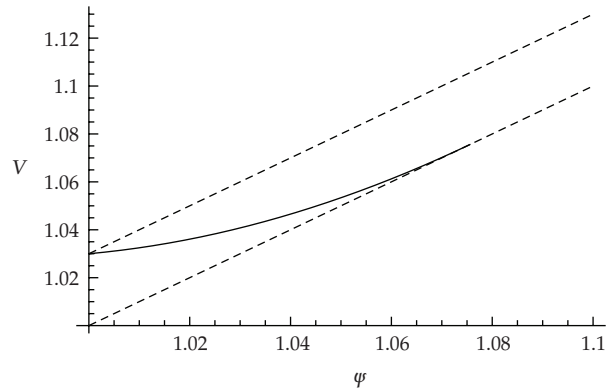


Figure 2: The value function $V(\psi)$ ($\delta = 0.03$).

3. Numerical Examples

In this section, we present some numerical examples which show that theoretical results are varied and some effects of the parameters on the price of the callable Russian option. We use the values of the parameters as follows: $\alpha = 0.5$, $r = 0.1$, $d = 0.09$, $\kappa = 0.3$, $\delta = 0.03$.

Figure 1 shows an optimal boundary for the buyer as a function of penalty costs δ, which is increasing in δ. Figures 2 and 3 show that the price of the callable Russian option has the low and upper bounds and is increasing and convex in ψ. Furthermore, we know that $V(\psi)$ is increasing in δ. Figure 4 demonstrates that the price of the callable Russian option with dividend is equal to or less than the one without dividend. Table 1 presents the values of the optimal boundaries for several combinations of the parameters.

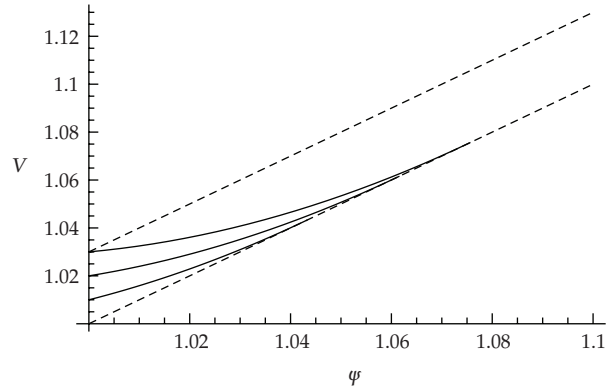


Figure 3: The value function $V(\psi)$ ($\delta = 0.01, 0.02, 0.03$).

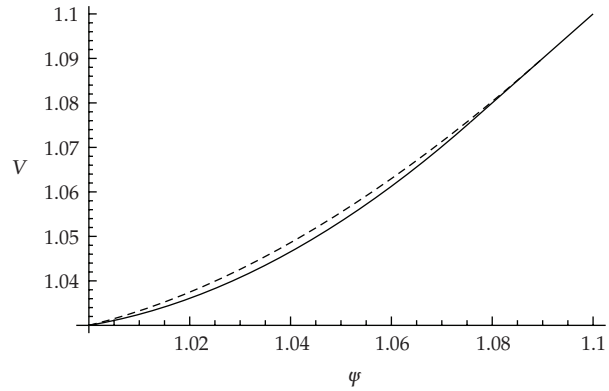


Figure 4: Real line with dividend; dash line without dividend.

4. Concluding Remarks

In this paper, we considered the pricing model of callable Russian options, where the stock pays continuously dividend. We derived the closed-form solution of such a Russian option as well as the optimal boundaries for the seller and the buyer, respectively. It is of interest to note that the price of the callable Russian option with dividend is not equal to the one as dividend value d goes to zero. This implicitly insist that the price of the callable Russian option without dividend is not merely the limit value of the one as if dividend vanishes as d goes to zero. We leave the rigorous proof for this question to future research. Further research is left for future work. For example, can the price of callable Russian options be decomposed into the sum of the prices of the noncallable Russian option and the callable discount? If the callable Russian option is finite lived, it is an interesting problem to evaluate the price of callable Russian option as the difference between the existing price formula and the premium value of the call provision.

Table 1: Penalty δ, interest rate r, dividend rate d, volatility κ, discount factor α, and the optimal boundary for the buyer l_*.

δ	r	d	κ	α	l_*
0.01	0.1	0.09	0.3	0.5	1.04337
0.02	0.1	0.09	0.3	0.5	1.0616
0.03	0.1	0.09	0.3	0.5	1.07568
0.03	0.2	0.09	0.3	0.5	1.08246
0.03	0.3	0.09	0.3	0.5	1.09228
0.03	0.4	0.09	0.3	0.5	1.10842
0.03	0.5	0.09	0.3	0.5	1.14367
0.03	0.1	0.01	0.3	0.5	1.08092
0.03	0.1	0.05	0.3	0.5	1.07813
0.03	0.1	0.1	0.3	0.5	1.07511
0.03	0.1	0.3	0.3	0.5	1.06633
0.03	0.1	0.5	0.3	0.5	1.06061
0.03	0.1	0.09	0.1	0.5	1.02468
0.03	0.1	0.09	0.2	0.5	1.04997
0.03	0.1	0.09	0.4	0.5	1.1018
0.03	0.1	0.09	0.5	0.5	1.12833
0.03	0.1	0.09	0.3	0.1	1.18166
0.03	0.1	0.09	0.3	0.2	1.12312
0.03	0.1	0.09	0.3	0.3	1.09901
0.03	0.1	0.09	0.3	0.4	1.08505

Acknowledgment

This work was supported by Grant-in-Aid for Scientific Research (A) 20241037, (B) 20330068, and a Grant-in-Aid for Young Scientists (Start-up) 20810038.

References

[1] I. Karatzas and S. E. Shreve, *Methods of Mathematical Finance*, Springer, New York, NY, USA, 1998.
[2] A. E. Kyprianou, "Some calculations for Israeli options," *Finance and Stochastics*, vol. 8, no. 1, pp. 73–86, 2004.
[3] A. Suzuki and K. Sawaki, "The pricing of callable perpetual American options," *Transactions of the Operations Research Society of Japan*, vol. 49, pp. 19–31, 2006 (Japanese).
[4] A. Suzuki and K. Sawaki, "The pricing of perpetual game put options and optimal boundaries," in *Recent Advances in Stochastic Operations Research*, pp. 175–188, World Scientific, River Edge, NJ, USA, 2007.
[5] L. A. Shepp and A. N. Shiryaev, "The Russian option: reduced regret," *The Annals of Applied Probability*, vol. 3, no. 3, pp. 631–640, 1993.
[6] L. A. Shepp and A. N. Shiryaev, "A new look at pricing of the "Russian option"," *Theory of Probability and Its Applications*, vol. 39, no. 1, pp. 103–119, 1994.
[7] Y. Kifer, "Game options," *Finance and Stochastics*, vol. 4, no. 4, pp. 443–463, 2000.
[8] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer, New York, NY, USA, 2nd edition, 1991.

Hindawi Publishing Corporation
Journal of Applied Mathematics and Decision Sciences
Volume 2009, Article ID 945923, 17 pages
doi:10.1155/2009/945923

Research Article

Valuation of Game Options in Jump-Diffusion Model and with Applications to Convertible Bonds

Lei Wang and Zhiming Jin

College of Science, National University of Defense Technology, ChangSha 410073, China

Correspondence should be addressed to Lei Wang, wlnudt@163.com

Received 17 November 2008; Accepted 6 March 2009

Recommended by Lean Yu

Game option is an American-type option with added feature that the writer can exercise the option at any time before maturity. In this paper, we consider some type of game options and obtain explicit expressions through solving Stefan(free boundary) problems under condition that the stock price is driven by some jump-diffusion process. Finally, we give a simple application about convertible bonds.

1. Introduction

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space hosting a Brownian motion $W = \{W_t : t \geq 0\}$ and an independent Poisson process $N = \{N_t : t \geq 0\}$ with the constant arrival rate λ, both adapted to some filtration $\mathbf{F} = \{\mathcal{F}_t\}_{t\geq 0}$ satisfying usual conditions. Consider the Black-Scholes market. That is, there is only one riskless bond B and a risky asset S. They satisfy, respectively,

$$\begin{aligned} dB_t &= rB_t dt, \quad t \geq 0, \\ dS_t &= S_{t-}\left[\mu dt + \sigma dW_t - y_0\left(dN_t - \lambda dt\right)\right] \end{aligned} \tag{1.1}$$

for some constants $\mu \in \mathbf{R}$, $r, \sigma > 0$ and $y_0 \in (0,1)$. Note that the absolute value of relative jump sizes is equal to y_0, and jumps are downwards. It can be comprehended as a downward tendency of the risky asset price brought by bad news or default and so on. From Itô formula we can obtain

$$S_t = S_0 \exp\left\{\left(\mu - \frac{1}{2}\sigma^2 + \lambda y_0\right)t + \sigma W_t\right\}\left(1 - y_0\right)^{N_t}. \tag{1.2}$$

Suppose that $X = \{X_t : t \le T\}$ and $Y = \{Y_t : t \le T\}$ be two continuous stochastic processes defined on $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$ such that for all $0 \le t \le T$, $X_t \le Y_t$a.s.. The game option is a contract between a holder and writer at time $t = 0$. It is a general American-type option with the added property that the writer has the right to terminate the contract at any time before expiry time T. If the holder exercises first, then he/she may obtain the value of X at the exercise time and if the writer exercise first, then he/she is obliged to pay to the holder the value of Y at the time of exercise. If neither has exercised at time T and $T < \infty$, then the writer pays the holder the value X_T. If both decide to claim at the same time then the lesser of the two claims is paid. In short, if the holder will exercise with strategy τ and the writer with strategy γ, we can conclude that at any moment during the life of the contract, the holder can expect to receive $Z(\tau, \gamma) \triangleq X_\tau 1_{(\tau \le \gamma)} + Y_\gamma 1_{(\gamma < \tau)}$. For a detailed description and the valuation of game options, we refer the reader to Kifer [1], Kyprianou [2], Ekström [3], Baurdoux and Kyprianou [4], Kühn et al. [5], and so on.

It is well known that in the no-arbitrage pricing framework, the value of a contract contingent on the asset S is the maximum of the expectation of the total discounted payoff of the contract under some equivalent martingale measure. Since the market is incomplete, there are more than one equivalent martingale measure. Following Dayanik and Egami [6], let the restriction to $\mathcal{F}_t$ of every equivalent martingale measure $\mathbf{P}^\alpha$ in a large class admit a Radon-Nikodym derivative in the form of

$$\begin{aligned} \left.\frac{d\mathbf{P}^\alpha}{d\mathbf{P}}\right|_{\mathcal{F}_t} &\triangleq \eta_t, \\ d\eta_t &= \eta_{t-}\left[\beta dW_t + (\alpha - 1)\left(dN_t - \lambda dt\right)\right], \quad t \ge 0,\ \eta_0 = 1 \end{aligned} \tag{1.3}$$

for some constants $\beta \in \mathbf{R}$ and $\alpha > 0$. The constants β and α are known as the market price of the diffusion risk and the market price of the jump risk, respectively, and satisfy the drift condition

$$\mu - r + \sigma\beta - \lambda y_0(\alpha - 1) = 0. \tag{1.4}$$

Then the discounted value process $\{e^{-rt}S_t : t \ge 0\}$ is a $(\mathbf{P}^\alpha, \mathbf{F})$-martingale. By the Girsanov theorem, the process $\{W_t^\alpha \triangleq W_t - \beta t : t \ge 0\}$ is a Brownian motion under the measure $\mathbf{P}^\alpha$, and $\{N_t : t \ge 0\}$ is a homogeneous Poisson process with the intensity $\lambda^\alpha \triangleq \lambda\alpha$ independent of the Brownian motion W^α under the same measure. The infinitesimal generator of the process S under the probability measure $\mathbf{P}^\alpha$ is given by

$$\mathcal{A}^\alpha f(x) \triangleq \left(r + \lambda^\alpha y_0\right)x\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 f}{\partial x^2} + \lambda^\alpha\left[f\left(x(1 - y_0)\right) - f(x)\right], \tag{1.5}$$

on the collection of twice-continuously differentiable functions $f(\cdot)$. It is easily checked that $(\mathcal{A}^\alpha - r)f(x) = 0$ admits two solutions $f(x) = x^{k_1}$ and $f(x) = x^{k_2}$, where $k_1 < 0 < 1 = k_2$ satisfy

$$\frac{1}{2}\sigma^2 k(k-1) + \left(r + \lambda^\alpha y_0\right)k - \left(r + \lambda^\alpha\right) + \lambda^\alpha\left(1 - y_0\right)^k = 0. \tag{1.6}$$

Suppose that $\mathbf{P}_x^\alpha$ is the equivalent martingale measure for S under the assumption that $S_0 = x$ for a specified market price $\alpha(\cdot)$ of the jump risk, and denote $\mathbf{E}_x^\alpha$ to be expectation under $\mathbf{P}_x^\alpha$. The following theorem is the Kifer pricing result.

Theorem 1.1. *Suppose that for all $x > 0$*

$$\mathbf{E}_x^\alpha\left(\sup_{0\le t\le T} e^{-rt}Y_t\right) < \infty \tag{1.7}$$

and if $T = \infty$ that $\mathbf{P}_x^\alpha(\lim_{t\uparrow\infty} e^{-rt}Y_t = 0) = 1$. Let $\mathcal{S}_{t,T}$ be the class of $\mathbf{F}$-stopping times valued in $[t,T]$, and $\mathcal{S} \equiv \mathcal{S}_{0,\infty}$, then the price of the game option is given by

$$V(x) = \inf_{\gamma\in\mathcal{S}_{0,T}} \sup_{\tau\in\mathcal{S}_{0,T}} \mathbf{E}_x^\alpha\left(e^{-r(\tau\wedge\gamma)}Z_{\tau,\gamma}\right) = \sup_{\tau\in\mathcal{S}_{0,T}} \inf_{\gamma\in\mathcal{S}_{0,T}} \mathbf{E}_x^\alpha\left(e^{-r(\tau\wedge\gamma)}Z_{\tau,\gamma}\right). \tag{1.8}$$

Further the optimal stopping strategies for the holder and writer, respectively, are

$$\tau^* = \inf\left\{t \ge 0 : V(S_t) = X_t\right\} \wedge T, \qquad \gamma^* = \inf\left\{t \ge 0 : V(S_t) = Y_t\right\} \wedge T. \tag{1.9}$$

2. A Game Version of the American Put Option (Perpetual Israeli δ-Penalty Put Option)

In this case, continuous stochastic processes are, respectively, given by

$$X_t = \left(K - S_t\right)^+, \qquad Y_t = \left(K - S_t\right)^+ + \delta, \tag{2.1}$$

where $K > 0$ is the strike-price of the option, $\delta > 0$ is a constant and can be considered as penalty for terminating contract by the writer. For the computation of the following, let us first consider the case of the perpetual American put option with the same parameter K. From Jin [7] we know that the price of the option is

$$V^A(x) = \sup_{\tau\in\mathcal{S}} \mathbf{E}_x^\alpha\left(e^{-r\tau}\left(K - S_\tau\right)^+\right) \tag{2.2}$$

with the superscript A representing American. Through martingale method we have the following.

Theorem 2.1. *The price of the perpetual American option is given by*

$$V^A(x) = \begin{cases} K - x & x \in (0, x^*], \\ \left(K - x^*\right)\left(\dfrac{x}{x^*}\right)^{k_1} & x \in (x^*, \infty), \end{cases} \tag{2.3}$$

where $x^* = k_1K/(k_1 - 1)$, the optimal stopping strategy is

$$\tau_* = \inf\left\{t \ge 0 : S_t \le x^*\right\}. \tag{2.4}$$

Proposition 2.2. *$V^A(x)$ is decreasing and convex on $(0,\infty)$, and under equivalent martingale measure $\mathbf{P}_x^\alpha$, one has that $\{e^{-rt}V^A(S_t) : t \geq 0\}$ and $\{e^{-r(t\wedge\tau_{x^*})}V^A(S_{t\wedge\tau_{x^*}}) : t \geq 0\}$ are supermartingale and martingale, respectively.*

Now, let us consider this game option. It is obvious that for the holder, in order to obtain the most profit, he will exercise when S becomes as small as possible. Meanwhile, he must not wait too long for this to happen, otherwise he will be punished by the exponential discounting. Then the compromise is to stop when S is smaller than a given constant. While for the writer, a reasonable strategy is to terminate the contract when the value of the asset S equals to K. Then only the burden of a payment of the form $\delta e^{-r\tau}$ is left. For this case, if the initial value of the risky asset is below K then it would seem rational to terminate the contract as soon as S hits K. On the other hand, if the initial value of the risky asset is above K, it is not optimal to exercise at once although the burden of the payment at this time is only δ. A rational strategy is to wait until the last moment that $S_t \geq K$ in order to prolong the payment. However, it should be noted that the value of the δ must not be too large, otherwise it will be never optimal for the writer to terminate the contract in advance.

Theorem 2.3. *Let $\delta^* \triangleq V^A(K) = (K - x^*)(K/x^*)^{k_1}$, one has the following.*

(1) *If $\delta \geq \delta^*$, then the price of this game option is equal to the price of the perpetual American put option, that is, it is not optimal for the writer to terminate the contract in advance.*

(2) *If $\delta < \delta^*$, then the price of the game option is*

$$V(x) = \begin{cases} K - x & x \in (0, k_*], \\ Ax + Bx^{k_1} & x \in (k_*, K), \\ \delta\left(\dfrac{x}{K}\right)^{k_1} & x \in [K, \infty) \end{cases} \tag{2.5}$$

with

$$A = \frac{\delta k_*^{k_1} - (K - k_*)K^{k_1}}{Kk_*^{k_1} - k_*K^{k_1}}, \qquad B = \frac{K(K - k_*) - \delta k_*}{Kk_*^{k_1} - k_*K^{k_1}}, \tag{2.6}$$

and the optimal stopping strategies for the holder and writer, respectively, are

$$\tau^* = \inf\{t \geq 0 : S_t \leq k_*\}, \qquad \gamma^* = \inf\{t \geq 0 : S_t = K\}, \tag{2.7}$$

where k_* is the (unique) solution in $(0, K)$ to the equation

$$(\delta + K)(1 - k_1)x^{k_1} + K^2k_1x^{k_1-1} - K^{1+k_1} = 0. \tag{2.8}$$

Before the proof, we will first give two propositions.

Proposition 2.4. *Equation* (2.8) *has and only has one root in* $(0, K)$.

Remark 2.5. If we denote the root of (2.8) in $(0, K)$ by k_*, then from Proposition 2.4 we know that $K(K-k_*)-\delta k_* > 0$, thus $B > 0$.

Proposition 2.6. *$V(x)$ defined by the right-hand sides of* (2.5) *is convex and decreasing on* $(0, \infty)$.

Proof. From the expression of $V(x)$ and Remark 2.5 we know that $V(x)$ is convex on $(0, K)$ and (K, ∞). Thus, we only need to prove the convexity of $V(x)$ at the point K, that is, $V'(K+) \geq V'(K-)$. Through elementary calculations we obtain

$$\begin{aligned} V'(K-) &= \frac{1}{Kk_*^{k_1} - k_* K^{k_1}} \left[\delta k_*^{k_1} - (K-k_*)K^{k_1} + \left(K(K-k_*) - \delta k_*\right) k_1 K^{k_1-1}\right], \\ V'(K+) &= \frac{\delta k_1}{K}. \end{aligned} \tag{2.9}$$

Then if we can prove that

$$\delta k_*^{k_1} - (K-k_*)K^{k_1} \leq 0, \tag{2.10}$$

$V'(K+) \geq V'(K-)$ will hold. From (2.8) we can easily find that when $\delta = \delta^*$, $k_* = x^*$. Further, as δ decreases the solution k_* increases. Especially, when $\delta = 0$, $k_* = K$. So if $0 < \delta < \delta^*$, we have $x^* < k_* < K$.

Now let us verify the correctness of (2.10). If not, that is, $\delta > (K-k_*)(K/k_*)^{k_1}$, then from (2.8) we obtain

$$K^{1+k_1} - K^2 k_1 k_*^{k_1-1} - K(1-k_1)k_*^{k_1} = \delta(1-k_1)k_*^{k_1} > (K-k_*)(1-k_1)K^{k_1}, \tag{2.11}$$

rearranging it we have

$$\left(k_* K^{k_1} - K k_*^{k_1}\right)\left(1 - k_1 + \frac{k_1 K}{k_*}\right) > 0. \tag{2.12}$$

Since $k_* > x^*$, so $1 - k_1 + k_1 K/k_* > 0$, whereas $k_* K^{k_1} - K k_*^{k_1} < 0$, which contradicts with (2.12). So the hypothesis is not true, that is, (2.10) holds, which also implies that $A \leq 0$. So $V(x)$ is decreasing on $(0, \infty)$. □

Proof of Theorem 2.3. (1) Suppose that $\delta \geq \delta^*$. From the expression of $V^A(x)$ we can easily find that

$$(K-x)^+ \leq V^A(x) \leq (K-x)^+ + \delta. \tag{2.13}$$

By means of Proposition 2.2 and the Doob Optional Stopping Theorem, we have

$$\begin{aligned} V^A(x) &= \inf_{\gamma\in\mathcal{S}} \mathbf{E}_x^\alpha\left[e^{-r(\tau_*\wedge\gamma)}V^A\left(S_{\tau_*\wedge\gamma}\right)\right] \\ &\le \inf_{\gamma\in\mathcal{S}} \mathbf{E}_x^\alpha\left\{e^{-r\tau_*}\left(K-S_{\tau_*}\right)^+1_{(\tau_*\le\gamma)} + e^{-r\gamma}\left[\left(K-S_\gamma\right)^+ + \delta\right]1_{(\gamma<\tau_*)}\right\} \\ &\le \inf_{\gamma\in\mathcal{S}}\sup_{\tau\in\mathcal{S}} \mathbf{E}_x^\alpha\left\{e^{-r\tau}\left(K-S_\tau\right)^+1_{(\tau\le\gamma)} + e^{-r\gamma}\left[\left(K-S_\gamma\right)^+ + \delta\right]1_{(\gamma<\tau)}\right\} \\ &= \sup_{\tau\in\mathcal{S}}\inf_{\gamma\in\mathcal{S}} \mathbf{E}_x^\alpha\left\{e^{-r\tau}\left(K-S_\tau\right)^+1_{(\tau\le\gamma)} + e^{-r\gamma}\left[\left(K-S_\gamma\right)^+ + \delta\right]1_{(\gamma<\tau)}\right\} \\ &\le \sup_{\tau\in\mathcal{S}} \mathbf{E}_x^\alpha\left[e^{-r\tau}\left(K-S_\tau\right)^+\right] \\ &= V^A(x). \end{aligned} \tag{2.14}$$

That is, the price of the game option is equal to the price of the perpetual American put option.

(2) If $\delta < \delta^*$, according to the foregoing discussion and Theorem 1.1, there exists a number k such that the continuation region is

$$C = \left\{x : g_1(x) < V(x) < g_2(x)\right\} = \{x : k < x < \infty,\ x \ne K\} \tag{2.15}$$

with $g_1(x) = (K-x)^+$, $g_2(x) = (K-x)^+ + \delta$, $k \in (0, K)$ a constant to be confirmed, while the stopping area is

$$D = D_1 \cup D_2, \tag{2.16}$$

where $D_1 = \{x : V(x) = g_1(x)\} = \{x : x \le k\}$ is the stopping area of the holder, $D_2 = \{x : V(x) = g_2(x)\} = \{x : x = K\}$ is the stopping area of the writer. For search of the optimal k_* and the value of $V(x)$, we consider the following Stefan(free boundary) problem with unknown number k and $V = V(x)$:

$$\begin{aligned} V(x) &= K - x, \quad x \in (0, k], \\ \left(\mathcal{A}^\alpha - r\right)V(x) &= 0, \quad x \in (k, K) \cup (K, \infty), \end{aligned} \tag{2.17}$$

and additional conditions on the boundary k and K are given by

$$\lim_{x\downarrow k} V(x) = K - k, \qquad \lim_{x\to K} V(x) = \delta, \qquad \lim_{x\downarrow k}\frac{\partial V(x)}{\partial x} = -1, \qquad \lim_{x\uparrow\infty} V(x) = 0. \tag{2.18}$$

By computing Stefan problem we can easily obtain the expression of $V(x)$ (denote it by $\widetilde{V}(x)$) defined by the right-hand sides of (2.5), while from (2.18) we can obtain (2.8). Proposition 2.4 implies that this equation has and only has one root in $(0, K)$, denote it by k_*. Accordingly, we can obtain the expression (2.6) of A and B and optimal stopping strategy τ^* for the holder. Now we must prove that the solution of the Stefan problem gives, in fact,

the solution to the optimal stopping problem, that is, $V(x) = \widetilde{V}(x)$. For that it is sufficient to prove that

(a) $\forall \tau \in \mathcal{S}$, $\mathbf{E}_x^{\alpha} e^{-r(\tau \wedge \gamma^*)} Z_{\tau,\gamma^*} \leq \widetilde{V}(x)$;

(b) $\forall \gamma \in \mathcal{S}$, $\mathbf{E}_x^{\alpha} e^{-r(\tau^* \wedge \gamma)} Z_{\tau^*,\gamma} \geq \widetilde{V}(x)$;

(c) $\mathbf{E}_x^{\alpha} e^{-r(\tau^* \wedge \gamma^*)} Z_{\tau^*,\gamma^*} = \widetilde{V}(x)$.

First, from Proposition 2.6 we know that $\widetilde{V}(x)$ is a convex function on $(0, \infty)$ such that

$$(K - x)^+ \leq \widetilde{V}(x) \leq (K - x)^+ + \delta. \tag{2.19}$$

Since $\widetilde{V}(x) \in C^1(0,K) \cap C^2(0,K) \setminus \{k_*\}$, for $x \in (0,K)$, we can apply Itô formula to the process $\{e^{-r(t\wedge\gamma^*)}\widetilde{V}(S_{t\wedge\gamma^*}) : t \geq 0\}$ and have

$$\begin{aligned} e^{-r(t\wedge\gamma^*)}\widetilde{V}\left(S_{t\wedge\gamma^*}\right) = \widetilde{V}(x) &+ \int_0^{t\wedge\gamma^*} e^{-ru}\left(\mathcal{A}^{\alpha} - r\right)\widetilde{V}\left(S_u\right)du + \int_0^{t\wedge\gamma^*} e^{-ru}\sigma S_u \widetilde{V}'\left(S_u\right)dW_u^{\alpha} \\ &+ \int_0^{t\wedge\gamma^*} e^{-ru}\left[\widetilde{V}\left(S_{u-}(1-y_0)\right) - \widetilde{V}\left(S_{u-}\right)\right]\left(dN_u - \lambda^{\alpha}du\right). \end{aligned} \tag{2.20}$$

Note that in $(0,K)$, $\mathcal{A}^{\alpha}\widetilde{V}(x) - r\widetilde{V}(x) \leq 0$, while the last two integrals of (2.20) are local martingales, then by choosing localizing sequence and apply the Fatou lemma, we obtain

$$\mathbf{E}_x^{\alpha} e^{-r(\tau\wedge\gamma^*)}\widetilde{V}\left(S_{\tau\wedge\gamma^*}\right) \leq \widetilde{V}(x), \tag{2.21}$$

whereas

$$\begin{aligned} Z_{\tau,\gamma^*} &= \left(K - S_\tau\right)^+ 1_{\tau\leq\gamma^*} + \left[\left(K - S_{\gamma^*}\right)^+ + \delta\right]1_{\gamma^*<\tau} \\ &= \left(K - S_\tau\right)^+ 1_{\tau\leq\gamma^*} + \delta 1_{\gamma^*<\tau} \\ &\leq \widetilde{V}\left(S_{\tau\wedge\gamma^*}\right). \end{aligned} \tag{2.22}$$

For the inequality we have used (2.19), hence from (2.21) we have

$$\mathbf{E}_x^{\alpha} e^{-r(\tau\wedge\gamma^*)} Z_{\tau,\gamma^*} \leq \widetilde{V}(x). \tag{2.23}$$

It is simple for the case that $x \in (K, \infty)$ and the method is the same as before. Thus, we obtain (a).

The proof of (b): apply Itô formula to the process $\{e^{-r(\tau^*\wedge t)}\widetilde{V}(S_{\tau^*\wedge t}) : t \geq 0\}$ and note that $\widetilde{V}$ is only continuous at K, we have

$$\begin{aligned} e^{-r(\tau^*\wedge t)}\widetilde{V}(S_{\tau^*\wedge t}) &= \widetilde{V}(x) + \int_0^{\tau^*\wedge t} e^{-ru}(\mathcal{A}^\alpha - r)\widetilde{V}(S_u)du + \int_0^{\tau^*\wedge t} e^{-ru}\sigma S_u \widetilde{V}'(S_u)dW_u^\alpha \\ &\quad + \int_0^{\tau^*\wedge t} e^{-ru}\left[\widetilde{V}(S_{u-}(1-y_0)) - \widetilde{V}(S_{u-})\right](dN_u - \lambda^\alpha du) \\ &\quad + e^{-r(\tau^*\wedge t)}\left[\widetilde{V}'(K+) - \widetilde{V}'(K-)\right]L^K_{\tau^*\wedge t}, \end{aligned} \tag{2.24}$$

where L^K is the local time at K of S. Since $\widetilde{V}(x)$ is convex on $(0,\infty)$, hence $\widetilde{V}'(K+) - \widetilde{V}'(K-) \geq 0$. While in $(k_*,\infty)\setminus\{K\}$, $(\mathcal{A}^\alpha - r)\widetilde{V}(x) = 0$, then using the same method as before we have

$$\mathbf{E}_x^\alpha e^{-r(\tau^*\wedge\gamma)}\widetilde{V}(S_{\tau^*\wedge\gamma}) \geq \widetilde{V}(x). \tag{2.25}$$

Moreover, since

$$\begin{aligned} \widetilde{V}(S_{\tau^*\wedge\gamma}) &= \widetilde{V}(S_{\tau^*})1_{(\tau^*\leq\gamma)} + \widetilde{V}(S_\gamma)1_{(\gamma<\tau^*)} \\ &\leq (K - S_{\tau^*})^+ 1_{(\tau^*\leq\gamma)} + \left[(K - S_\gamma)^+ + \delta\right]1_{(\gamma<\tau^*)} \\ &= Z_{\tau^*,\gamma}, \end{aligned} \tag{2.26}$$

we can obtain

$$\mathbf{E}_x^\alpha e^{-r(\tau^*\wedge\gamma)}Z_{\tau^*,\gamma} \geq \widetilde{V}(x), \quad \forall\gamma\in\mathcal{M}. \tag{2.27}$$

The proof of (c): taking $\tau = \tau^*$, $\gamma = \gamma^*$, it is sufficient to note that in (k_*, K), we have $\mathcal{A}^\alpha V(x) - rV(x) = 0$ and

$$\begin{aligned} \widetilde{V}(S_{\tau^*\wedge\gamma^*}) &= \widetilde{V}(S_{\tau^*})1_{(\tau^*\leq\gamma^*)} + \widetilde{V}(S_{\gamma^*})1_{(\gamma^*<\tau^*)} \\ &= (K - S_{\tau^*})^+ 1_{(\tau^*\leq\gamma^*)} + \delta 1_{(\gamma^*<\tau^*)} \\ &= Z_{\tau^*,\gamma^*}. \end{aligned} \tag{2.28}$$

The same result is true for the case that $x \in (K,\infty)$. □

3. Game Option with Barrier

Karatzas and Wang [8] obtain closed-form expressions for the prices and optimal hedging strategies of American put options in the presence of an up-and-out barrier by reducing this problem to a variational inequality. Now we will consider the game option connected with this barrier option. Following Karatzas and Wang, the holder may exercise to take the claim of this barrier option

$$X_t = (K - S_t)^+ 1_{(t<\tau_h)}, \quad 0 \leq t < \infty. \tag{3.1}$$

Here $h > 0$ is the barrier, whereas

$$\tau_h = \inf\{t \geq 0 : S_t > h\} \tag{3.2}$$

is the time when the option becomes "knocked-out". The writer is punished by an amount δ for terminating the contract early

$$Y_t = \left[(K - S_t)^+ + \delta\right] 1_{(t<\tau_h)}. \tag{3.3}$$

First, let us consider this type of barrier option. The price is given by

$$V^B(x) = \sup_{\tau \in \mathcal{S}} \mathbf{E}_x e^{-r\tau} (K - S_\tau)^+ 1_{(\tau<\tau_h)} \tag{3.4}$$

with the superscript B representing barrier. Similarly to Karatzas and Wang we can obtain the following.

Theorem 3.1. *The price of American put-option in the presence of an up-and-out barrier is*

$$V^B(x) = \begin{cases} K - x & x \in (0, p_*], \\ Ax + Bx^{k_1} & x \in (p_*, h), \\ 0 & x \in [h, \infty), \end{cases} \tag{3.5}$$

where $A = (p_* - K)h^{k_1}/(hp_*^{k_1} - p_* h^{k_1})$, $B = (K - p_*)h/(hp_*^{k_1} - p_* h^{k_1})$, *and the optimal stopping strategy is*

$$\tau_* = \inf\{t \geq 0 : S_t \leq p_*\}, \tag{3.6}$$

where p_* *is the (unique) solution in* $(0, K)$ *to the equation*

$$h(1 - k_1)x^{k_1} + Kh\,k_1 x^{k_1 - 1} - Kh^{k_1} = 0. \tag{3.7}$$

The proof of the theorem mainly depends on the following propositions and the process will be omitted.

Proposition 3.2. *The expression of* $V^B(x)$ *defined by* (3.5) *is convex and decreasing on* $(0, \infty)$, *and under risk-neutral measure* $\mathbf{P}_x^\alpha$, *one has that* $\{e^{-rt} V^B(S_t) : t \geq 0\}$ *and* $\{e^{-r(t\wedge\tau_*)} V^B(S_{t\wedge\tau_*}) : t \geq 0\}$ *are supermartingale and martingale, respectively.*

Proposition 3.3. *Equation* (3.7) *has and only has one root in* $(0, K)$.

Now let us consider the game option with barrier h. The price is given by

$$V(x) = \sup_{\tau \in \mathcal{S}} \inf_{\gamma \in \mathcal{S}} \mathbf{E}_x \left(e^{-r\tau} (K - S_\tau)^+ 1_{(\tau \leq \gamma)} \cdot 1_{(\tau<\tau_h)} + e^{-r\gamma} \left[(K - S_\gamma)^+ + \delta\right] 1_{(\gamma<\tau)} \cdot 1_{(\gamma<\tau_h)} \right). \tag{3.8}$$

For this game option, the logic of its solution is similar to the former, and based on this consideration, we have the following theorem.

Theorem 3.4. *Let* $\delta^* \triangleq V^B(K) = (K - p_*)(hK^{k_1} - Kh^{k_1})/(hp_*^{k_1} - p_*h^{k_1})$, *one has the following.*

(1) *If* $\delta \geq \delta^*$, *then the price of this game option is equal to the price of American put options in the presence of an up-and-out barrier, that is, it is not optimal for the writer to exercise early.*

(2) *If* $\delta < \delta^*$, *then the price of the game option is given by*

$$V(x) = \begin{cases} K - x & x \in (0, b_*], \\ C_1 x + C_2 x^{k_1} & x \in (b_*, K), \\ D_1 x + D_2 x^{k_1} & x \in [K, h), \\ 0 & x \in [h, \infty), \end{cases} \tag{3.9}$$

where

$$\begin{aligned} C_1 &= \frac{\delta b_*^{k_1} - (K - b_*)K^{k_1}}{Kb_*^{k_1} - b_*K^{k_1}}, \qquad C_2 = \frac{K(K - b_*) - \delta b_*}{Kb_*^{k_1} - b_*K^{k_1}}, \\ D_1 &= \frac{-\delta h^{k_1}}{hK^{k_1} - Kh^{k_1}}, \qquad D_2 = \frac{\delta h}{hK^{k_1} - Kh^{k_1}}, \end{aligned} \tag{3.10}$$

and b_* *is the (unique) solution in* $(0, K)$ *to the equation*

$$(\delta + K)(1 - k_1)x^{k_1} + K^2 k_1 x^{k_1 - 1} - K^{1 + k_1} = 0, \tag{3.11}$$

and the optimal stopping strategies for the holder and writer, respectively, are

$$\tau^* = \inf\{t \geq 0 : S_t \leq b_*\}, \qquad \gamma^* = \inf\{t \geq 0 : S_t = K\}. \tag{3.12}$$

Proposition 3.5. *The function* $V(x)$ *defined by* (3.9) *is convex and decreasing on* $(0, \infty)$.

Proof. Similar to Proposition 2.6, we only need to prove the convexity of $V(x)$ at the point K, that is,

$$V'(K+) - V'(K-) = (D_2 - C_2)k_1 K^{k_1 - 1} + (D_1 - C_1) \geq 0. \tag{3.13}$$

Through lengthy calculations we know that it is sufficient to show that

$$\delta(hb_*^{k_1} - b_*h^{k_1}) \leq (K - b_*)(hK^{k_1} - Kh^{k_1}). \tag{3.14}$$

Suppose that (3.14) does not hold, that is, $\delta > (K-b_*)(hK^{k_1}-Kh^{k_1})/(hb_*^{k_1}-b_*h^{k_1})$, then from (3.11) we find that

$$\begin{aligned} K^{1+k_1}-K^2k_1x^{k_1-1}-K(1-k_1)b_*^{k_1} &= \delta(1-k_1)b_*^{k_1} \\ &> (1-k_1)b_*^{k_1}\frac{(K-b_*)(hK^{k_1}-Kh^{k_1})}{(hb_*^{k_1}-b_*h^{k_1})}, \end{aligned} \tag{3.15}$$

rearranging it we have

$$h(1-k_1)b_*^{k_1}+Kh\,k_1b_*^{k_1-1}-Kh^{k_1}<0. \tag{3.16}$$

From (3.11), through complex verification we get that when $\delta=\delta^*$, $b_*=p_*$. Furthermore, as δ decreases the solution b_* increases, especially when $\delta=0$, $b_*=K$. So if $0<\delta<\delta^*$, we have $p_*<b_*<K$. Thus from the property of (3.7) we know that $h(1-k_1)b_*^{k_1}+Kh\,k_1b_*^{k_1-1}-Kh^{k_1}>0$, which contradicts with (3.16). So the hypothesis is not true, that is, (3.14) holds. It is evident that $V(x)$ is decreasing. □

Remark 3.6. It is obvious that (2.8) is the same as (3.11), however, their roots not always be equal to each other. Because of these two cases, the scope of δ is different. Penalty with barrier is usually smaller than the other, that is, $V^B(K)<V^A(K)$.

Proof of Theorem 3.4. (1) Suppose that $\delta\geq\delta^*$. From Proposition 3.2 we know that

$$(K-x)^+\leq V^B(x)\leq(K-x)^++\delta. \tag{3.17}$$

By means of the Doob optional stopping theorem and (3.17), we have

$$\begin{aligned} V^B(x) &= \inf_{\gamma\in\mathcal{S}}\mathbf{E}_x^{\alpha}\left[e^{-r(\tau^*\wedge\gamma\wedge\tau_h)}V^B\left(S_{\tau^*\wedge\gamma\wedge\tau_h}\right)\right] \\ &\leq \inf_{\gamma\in\mathcal{S}}\mathbf{E}_x^{\alpha}\left\{e^{-r\tau^*}\left(K-S_{\tau^*}\right)^+1_{(\tau^*\leq\gamma)}1_{(\tau^*<\tau_h)}+e^{-r\gamma}\left[\left(K-S_\gamma\right)^++\delta\right]1_{(\gamma<\tau^*)}1_{(\gamma<\tau_h)}\right\} \\ &\leq \inf_{\gamma\in\mathcal{S}}\sup_{\tau\in\mathcal{S}}\mathbf{E}_x^{\alpha}\left\{e^{-r\tau}\left(K-S_\tau\right)^+1_{(\tau\leq\gamma)}1_{(\tau<\tau_h)}+e^{-r\gamma}\left[\left(K-S_\gamma\right)^++\delta\right]1_{(\gamma<\tau)}1_{(\gamma<\tau_h)}\right\} \\ &= \sup_{\tau\in\mathcal{S}}\inf_{\gamma\in\mathcal{S}}\mathbf{E}_x^{\alpha}\left\{e^{-r\tau}\left(K-S_\tau\right)^+1_{(\tau\leq\gamma)}1_{(\tau<\tau_h)}+e^{-r\gamma}\left[\left(K-S_\gamma\right)^++\delta\right]1_{(\gamma<\tau)}1_{(\gamma<\tau_h)}\right\} \\ &\leq \sup_{\tau\in\mathcal{S}}\mathbf{E}_x^{\alpha}\left[e^{-r\tau}\left(K-S_\tau\right)^+1_{(\tau<\tau_h)}\right] \\ &= V^B(x). \end{aligned} \tag{3.18}$$

That is, the price of the game option is equal to the price of American put-options in the presence of an up-and-out barrier.

(2) Suppose that $\delta<\delta^*$. Then we may conclude that the holder should search optimal stopping strategy in the class of the stopping times of the form $\tau_b=\inf\{t\geq0:S_t\leq b\}$ with

$b \in (0, K)$ to be confirmed. While the optimal stopping strategy for the writer is $\gamma^* = \inf\{t \geq 0 : S_t = K\}$. Considering the following Stefan problem:

$$V(x) = K - x, \quad x \in (0, b], \tag{3.19}$$

$$(\mathcal{A} - r)V(x) = 0, \quad x \in (b, K) \cup (K, h), \tag{3.20}$$

$$V(x) = 0, \quad x \in [h, \infty), \tag{3.21}$$

$$\lim_{x \downarrow b} V(x) = K - b, \qquad \lim_{x \to K} V(x) = \delta, \qquad \lim_{x \uparrow h} V(x) = 0, \qquad \lim_{x \downarrow b} \frac{\partial V(x)}{\partial x} = -1. \tag{3.22}$$

Through straightforward calculations we can obtain the expression of $V(x)$ (denote it by $\widetilde{V}(x)$) defined by the right-hand sides of (3.9). From condition (3.22) we can obtain (3.11). Proposition 2.4 implies that the root of this equation is unique in $(0, K)$, denote it by b_* and consequently τ_{b_*} by τ^*. Now we only need to prove that $V(x) = \widetilde{V}(x)$. For that it is sufficient to prove that

$$\text{(a)} \quad \forall \tau \in \mathcal{S}, \quad \mathbf{E}_x^{\alpha} e^{-r(\tau \wedge \gamma^*)} Z_{\tau, \gamma^*} 1_{(\tau \wedge \gamma^* < \tau_h)} \leq \widetilde{V}(x); \tag{3.23}$$

$$\text{(b)} \quad \forall \gamma \in \mathcal{S}, \quad \mathbf{E}_x^{\alpha} e^{-r(\tau^* \wedge \gamma)} Z_{\tau^*, \gamma} 1_{(\tau^* \wedge \gamma < \tau_h)} \geq \widetilde{V}(x). \tag{3.24}$$

(c) Taking stopping time $\tau = \tau^*$, $\gamma = \gamma^*$, we have

$$\mathbf{E}_x^{\alpha} e^{-r(\tau^* \wedge \gamma^*)} Z_{\tau^*, \gamma^*} 1_{(\tau^* \wedge \gamma^* < \tau_h)} = \widetilde{V}(x). \tag{3.25}$$

First, from Proposition 3.5 we know that $\widetilde{V}(x)$ is convex in $(0, \infty)$ and further

$$(K - x)^+ \leq \widetilde{V}(x) \leq (K - x)^+ + \delta. \tag{3.26}$$

Applying Itô formula to the process $\{e^{-r(t \wedge \gamma^* \wedge \tau_h)} \widetilde{V}(S_{t \wedge \gamma^* \wedge \tau_h}) : t \geq 0\}$, we have

$$\begin{aligned} e^{-r(t \wedge \gamma^* \wedge \tau_h)} \widetilde{V}\left(S_{t \wedge \gamma^* \wedge \tau_h}\right) &= \widetilde{V}(x) + \int_0^{t \wedge \gamma^* \wedge \tau_h} e^{-ru} \left(\mathcal{A}^{\alpha} - r\right) \widetilde{V}\left(S_u\right) du + \int_0^{t \wedge \gamma^* \wedge \tau_h} e^{-ru} \sigma S_u \widetilde{V}'\left(S_u\right) dW_u^{\alpha} \\ &\quad + \int_0^{t \wedge \gamma^* \wedge \tau_h} e^{-ru} \left[\widetilde{V}\left(S_{u-}(1 - y_0)\right) - \widetilde{V}\left(S_{u-}\right)\right] \left(dN_u - \lambda^{\alpha} du\right). \end{aligned} \tag{3.27}$$

It is obvious that when $x \in (0, K) \cup (K, h)$, we have $(\mathcal{A} - r)\widetilde{V}(x) \leq 0$. Since the second and the third integrals of the right-hand sides of (3.27) are local martingales, so

$$\mathbf{E}_x^{\alpha} e^{-r(\tau \wedge \gamma^* \wedge \tau_h)} \widetilde{V}\left(S_{\tau \wedge \gamma^* \wedge \tau_h}\right) \leq \widetilde{V}(x), \tag{3.28}$$

while

$$\mathbf{E}_x^{\alpha} e^{-r(\tau\wedge\gamma^*\wedge\tau_h)}\widetilde{V}\left(S_{\tau\wedge\gamma^*\wedge\tau_h}\right) = \mathbf{E}_x^{\alpha} e^{-r(\tau\wedge\gamma^*)}\widetilde{V}\left(S_{\tau\wedge\gamma^*}\right)1_{(\tau\wedge\gamma^*<\tau_h)} \geq \mathbf{E}_x^{\alpha} e^{-r(\tau\wedge\gamma^*)} Z_{\tau,\gamma^*} 1_{(\tau\wedge\gamma^*<\tau_h)}. \tag{3.29}$$

The inequality is obtained from (3.26), and combining (3.28) we obtain (3.23), that is, (a) holds.

Applying Itô formula to the process $\{e^{-r(\tau^*\wedge t\wedge\tau_h)}\widetilde{V}(S_{\tau^*\wedge t\wedge\tau_h}) : t \geq 0\}$, we have

$$\begin{aligned} e^{-r(\tau^*\wedge t\wedge\tau_h)}\widetilde{V}\left(S_{\tau^*\wedge t\wedge\tau_h}\right) &= \widetilde{V}(x) + \int_0^{\tau^*\wedge t\wedge\tau_h} e^{-ru}(\mathcal{A}-r)\widetilde{V}\left(S_u\right)du + \int_0^{\tau^*\wedge t\wedge\tau_h} e^{-ru}\sigma\left(S_u\right)\widetilde{V}'\left(S_u\right)dW_u^{\alpha} \\ &\quad + e^{-r(\tau^*\wedge t\wedge\tau_h)}\left[\widetilde{V}'(K+) - \widetilde{V}'(K-)\right]L^K_{\tau^*\wedge t\wedge\tau_h} \\ &\quad + \int_0^{\tau^*\wedge t\wedge\tau_h} e^{-ru}\left[\widetilde{V}\left(S_{u-}(1-y_0)\right) - \widetilde{V}\left(S_{u-}\right)\right]\left(dN_u - \lambda^{\alpha}du\right). \end{aligned} \tag{3.30}$$

The definition of L^K is the same as Theorem 2.3. From the convexity of $\widetilde{V}(x)$ we know that $\widetilde{V}'(K+) - \widetilde{V}'(K-) \geq 0$. Since when $x \in (b_*, K)\cup(K,h)$, $(\mathcal{A}-r)\widetilde{V}(x) = 0$, so from above expression we have

$$\mathbf{E}_x^{\alpha} e^{-r(\tau^*\wedge\gamma\wedge\tau_h)}\widetilde{V}\left(S_{\tau^*\wedge\gamma\wedge\tau_h}\right) \geq \widetilde{V}(x). \tag{3.31}$$

Similarly we have

$$\mathbf{E}_x^{\alpha} e^{-r(\tau^*\wedge\gamma\wedge\tau_h)}\widetilde{V}\left(S_{\tau^*\wedge\gamma\wedge\tau_h}\right) = \mathbf{E}_x^{\alpha} e^{-r(\tau^*\wedge\gamma)}\widetilde{V}\left(S_{\tau^*\wedge\gamma}\right)1_{(\tau^*\wedge\gamma<\tau_h)} \leq \mathbf{E}_x^{\alpha} e^{-r(\tau^*\wedge\gamma)} Z_{\tau^*,\gamma} 1_{(\tau^*\wedge\gamma<\tau_h)}. \tag{3.32}$$

From (3.31) and (3.32) we know that (b) holds. Combining (a) and (b) we can easily obtain (c). □

4. A Simple Example: Application to Convertible Bonds

To raise capital on financial markets, companies may choose among three major asset classes: equity, bonds, and hybrid instruments, such as convertible bonds. As hybrid instruments, convertible bonds has been investigated rather extensively during the recent years. It entitles its owner to receive coupons plus the return of the principle at maturity. However, the holder can convert it into a preset number of shares of stock prior to maturity. Then the price of the bond is dependent on the price of the firm stock. Finally, prior to maturity, the firm may call the bond, forcing the bondholder to either surrender it to the firm for a previously agreed price or else convert it for stock as above. Therefore, the pricing problem has also a game-theoretic aspect. For more detailed information and research about convertible bonds, one is referred to Gapeev and Kühn [9], Sîrbu et al. [10, 11], and so on.

Now, we will give a simple example of pricing convertible bonds, as the application of pricing game options. Consider the stock process which pays dividends at a certain fixed rate $d \in (0, r)$, that is,

$$dS_t = S_{t-}\left[(\mu - d)dt + \sigma dW_t - y_0\left(dN_t - \lambda dt\right)\right]. \tag{4.1}$$

Then the infinitesimal generator of S becomes

$$\mathcal{A}^{\alpha} f(x) \triangleq \left(r - d + \lambda^{\alpha} y_0\right) x \frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 f}{\partial x^2} + \lambda^{\alpha}\left[f\left(x(1 - y_0)\right) - f(x)\right], \tag{4.2}$$

and $(\mathcal{A}^{\alpha} - r) f(x) = 0$ admits two solutions $f(x) = x^{k_1}$ and $f(x) = x^{k_2}$ with $k_1 < 0 < 1 < k_2$ satisfying

$$\frac{1}{2}\sigma^2 k(k-1) + \left(r - d + \lambda^{\alpha} y_0\right) k - \left(r + \lambda^{\alpha}\right) + \lambda^{\alpha}\left(1 - y_0\right)^k = 0. \tag{4.3}$$

At any time, the bondholder can convert it into a predetermined number $\eta > 0$ of stocks, or continue to hold the bond and collecting coupons at the fixed rate $c > 0$. On the other hand, at any time the firm can call the bond, which requires the bondholder to either immediately surrender it for the fixed conversion price $K > 0$ or else immediately convert it as described above. In short, the firm can terminate the contract by paying the amount $\max\{K, \eta S\}$ to the holder. Then, if the holder terminates the contract first by converting the bond into η stocks, he/she can expect to (discounted) receive

$$L_t = \int_0^t c \cdot e^{-ru} du + e^{-rt} \eta S_t, \tag{4.4}$$

while if the firm terminate the contract first, he/she will pay the holder

$$U_t = \int_0^t c \cdot e^{-ru} du + e^{-rt}\left(K \vee \eta S_t\right). \tag{4.5}$$

Then, according to Theorem 1.1, the price of the convertible bonds is given by

$$V^{\mathrm{CB}}(x) = \inf_{\gamma \in \mathcal{S}} \sup_{\tau \in \mathcal{S}} \mathbf{E}_x^{\alpha}\left(L_\tau 1_{(\tau \le \gamma)} + U_\gamma 1_{(\gamma < \tau)}\right) = \sup_{\tau \in \mathcal{S}} \inf_{\gamma \in \mathcal{S}} \mathbf{E}_x^{\alpha}\left(L_\tau 1_{(\tau \le \gamma)} + U_\gamma 1_{(\gamma < \tau)}\right). \tag{4.6}$$

Note that when $c \ge rK$, the solution of (4.6) is trivial and the firm should call the bond immediately. This implies that the bigger the coupon rate c, the more the payoff of the issuer, then they will choose to terminate the contract immediately. So we will assume that $c < rK$ in the following.

Now, let us first consider the logic of solving this problem. It is obvious that $\eta x \le V^{\mathrm{CB}}(x) \le K \vee \eta x$ for all $x > 0$ (choose $\tau = 0$ and $\gamma = 0$, resp.). Note when $S_t \ge K/\eta$, $L_t = U_t$,

then $V^{\mathrm{CB}}(x) = \eta x$ for all $x \geq K/\eta$. Hence the issuer and the holder should search optimal stopping in the class of stopping times of the form

$$\gamma_a = \inf\{t \geq 0 : S_t \geq a\}, \qquad \tau_b = \inf\{t \geq 0 : S_t \geq b\}, \tag{4.7}$$

respectively, with numbers $0 < a, b \leq K/\eta$ to be determined. Note when the process S fluctuates in the interval $(0, K/\eta)$, it is not optimal to terminate the contract simultaneously by both issuer and holder. For example, if the issuer chooses to terminate the contract at the first time that S exceeds some point $a \in (0, K/\eta)$, then $\eta a < K$, and the holder will choose the payoff of coupon rather than converting the bond into the stock, which is a contradiction. Similarly, one can explain another case. Then only the following situation can occur: either $a < b = K/\eta$, $b < a = K/\eta$, or $b = a = K/\eta$.

For search of the optimal a_*, b_* and the value of $V^{\mathrm{CB}}(x)$, we consider an auxiliary Stefan problem with unknown numbers a, b, and $V(x)$

$$\begin{aligned} &\left(\mathcal{A}^{\alpha} - r\right)V(x) = -c, \quad 0 < x < a \wedge b, \\ &\eta x < V(x) < \eta x \vee K, \quad 0 < x < a \wedge b \end{aligned} \tag{4.8}$$

with continuous fit boundary conditions

$$V(b-) = \eta b, \qquad V(x) = \eta x \tag{4.9}$$

for all $x > b$, $b \leq a = K/\eta$, and

$$V(a-) = K, \qquad V(x) = \eta x \vee K \tag{4.10}$$

for all $x > a$, $a \leq b = K/\eta$, and smooth fit boundary conditions

$$V'(b-) = \eta \quad \text{if } b < a = \frac{K}{\eta}, \qquad V'(a-) = 0 \quad \text{if } a < b = \frac{K}{\eta}. \tag{4.11}$$

By computing the Stefan problem we can obtain that if

$$K > \frac{k_2}{k_2 - 1}\frac{c}{r}, \tag{4.12}$$

then $b_* < a_* = K/\eta$, and the expression of $V(x)$ is given by

$$V(x) = \frac{\eta b_*}{k_2}\left(\frac{x}{b_*}\right)^{k_2} + \frac{c}{r} \tag{4.13}$$

for all $0 < x < b_*$, with

$$b_* = \frac{k_2}{\eta(k_2 - 1)} \frac{c}{r}, \tag{4.14}$$

and if

$$\frac{c}{r} < K \leq \frac{k_2}{k_2 - 1} \frac{c}{r}, \tag{4.15}$$

then $a_* = b_* = K/\eta$, and the value of $V(x)$ is

$$V(x) = \left(K - \frac{c}{r}\right)\left(\frac{\eta x}{K}\right)^{k_2} + \frac{c}{r} \tag{4.16}$$

for all $0 < x < K/\eta$.

From the result we can observe that there are only two regions for K, and the situation $a_* < b_* = K/\eta$ fails to hold. This implies that in this case, when S fluctuates in the interval $(0, K/\eta)$, the issuer will never recall the bond. Now, we only need to prove that $V(x) = V^{\mathrm{CB}}(x)$, and the stopping times γ^* and τ^* defined by (4.7) with boundaries a_* and b_* are optimal.

Applying Itô formula to the process $\{e^{-rt}V(S_t) : t \geq 0\}$, we have

$$\begin{aligned} e^{-rt}V(S_t) = V(x) &+ \int_0^t e^{-ru}(\mathcal{A}^\alpha - r)V(S_u)1_{(S_u \neq a_*, S_u \neq b_*, S_u \neq K/\eta)}du \\ &+ \int_0^t e^{-ru}\sigma S_u V'(S_u)1_{(S_u \neq a_*, S_u \neq b_*, S_u \neq K/\eta)}dW_u^\alpha \\ &+ \int_0^t e^{-ru}\left[V(S_{u-}(1 - y_0)) - V(S_{u-})\right](dN_u - \lambda^\alpha du) \\ &+ e^{-rt}\left[V'\left(\frac{K}{\eta} + \right) - V'\left(\frac{K}{\eta} - \right)\right]L_t^{K/\eta}. \end{aligned} \tag{4.17}$$

Let

$$\begin{aligned} M_t = &\int_0^t e^{-ru}\sigma S_u V'(S_u)1_{(S_u \neq a_*, S_u \neq b_*, S_u \neq K/\eta)}dW_u^\alpha \\ &+ \int_0^t e^{-ru}\left[V(S_{u-}(1 - y_0)) - V(S_{u-})\right](dN_u - \lambda^\alpha du). \end{aligned} \tag{4.18}$$

Note for all $0 < x < a_*$, $(\mathcal{A} - r)V(x) \leq -c$, while for all $0 < x < b_*$, $(\mathcal{A} - r)V(x) = -c$. Since $\eta x \leq V(x) \leq \eta x \vee K$, so for $0 < a_* \leq K/\eta$, $0 < b_* \leq K/\eta$, we have

$$\begin{aligned} L_{\tau\wedge\gamma^*} &= \int_0^{\tau\wedge\gamma^*} c \cdot e^{-ru}du + e^{-r(\tau\wedge\gamma^*)}\eta S_{\tau\wedge\gamma^*} \\ &\leq \int_0^{\tau\wedge\gamma^*} c \cdot e^{-ru}du + e^{-r(\tau\wedge\gamma^*)}V\left(S_{\tau\wedge\gamma^*}\right) \\ &\leq V(x) + M_{\tau\wedge\gamma^*}, \\ U_{\tau^*\wedge\gamma} &= \int_0^{\tau^*\wedge\gamma} c \cdot e^{-ru}du + e^{-r(\tau^*\wedge\gamma)}\left(\eta S_{\tau^*\wedge\gamma} \vee K\right) \\ &\geq \int_0^{\tau^*\wedge\gamma} c \cdot e^{-ru}du + e^{-r(\tau^*\wedge\gamma)}V\left(S_{\tau^*\wedge\gamma}\right) \\ &= V(x) + M_{\tau^*\wedge\gamma}. \end{aligned} \tag{4.19}$$

Because $V(S_{\gamma^*}) = K \vee \eta S_{\gamma^*}$, $V(S_{\tau^*}) = \eta S_{\tau^*}$, then by choosing localizing sequence and apply the Fatou lemma, we obtain

$$\mathbf{E}_x^{\alpha}\left[L_{\tau}1_{(\tau\leq\gamma^*)} + U_{\gamma^*}1_{(\gamma^*<\tau)}\right] \leq V(x) \leq \mathbf{E}_x^{\alpha}\left[L_{\tau^*}1_{(\tau^*\leq\gamma)} + U_{\gamma}1_{(\gamma<\tau^*)}\right]. \tag{4.20}$$

Taking supremum and infimum for τ and γ of both sides, respectively, we can obtain the result. While for (4.20), taking $\tau = \tau^*$, $\gamma = \gamma^*$, we have

$$V(x) = \mathbf{E}_x^{\alpha}\left[L_{\tau^*}1_{(\tau^*\leq\gamma^*)} + U_{\gamma^*}1_{(\gamma^*<\tau^*)}\right]. \tag{4.21}$$

References

[1] Y. Kifer, "Game options," *Finance and Stochastics*, vol. 4, no. 4, pp. 443–463, 2000.

[2] A. E. Kyprianou, "Some calculations for Israeli options," *Finance and Stochastics*, vol. 8, no. 1, pp. 73–86, 2004.

[3] E. Ekström, "Properties of game options," *Mathematical Methods of Operations Research*, vol. 63, no. 2, pp. 221–238, 2006.

[4] E. J. Baurdoux and A. E. Kyprianou, "Further calculations for Israeli options," *Stochastics and Stochastics Reports*, vol. 76, no. 6, pp. 549–569, 2004.

[5] C. Kühn, A. E. Kyprianou, and K. van Schaik, "Pricing Israeli options: a pathwise approach," *Stochastics*, vol. 79, no. 1-2, pp. 117–137, 2007.

[6] S. Dayanik and M. Egami, "Optimal stopping problems for asset management," in preparation.

[7] Z. M. Jin, *Foundations of Mathematical Finance*, Science Press, Beijing, China, 2006.

[8] I. Karatzas and H. Wang, "A barrier option of American type," *Applied Mathematics and Optimization*, vol. 42, no. 3, pp. 259–279, 2000.

[9] P. V. Gapeev and C. Kühn, "Perpetual convertible bonds in jump-diffusion models," *Statistics & Decisions*, vol. 23, no. 1, pp. 15–31, 2005.

[10] M. Sîrbu, I. Pikovsky, and S. E. Shreve, "Perpetual convertible bonds," *SIAM Journal on Control and Optimization*, vol. 43, no. 1, pp. 58–85, 2004.

[11] M. Sîrbu and S. E. Shreve, "A two-person game for pricing convertible bonds," *SIAM Journal on Control and Optimization*, vol. 45, no. 4, pp. 1508–1539, 2006.

Hindawi Publishing Corporation
Journal of Applied Mathematics and Decision Sciences
Volume 2009, Article ID 817137, 16 pages
doi:10.1155/2009/817137

Research Article

Fuzzy Real Options in Brownfield Redevelopment Evaluation

Qian Wang,[1] Keith W. Hipel,[1] and D. Marc Kilgour[2]

[1] *Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1*
[2] *Department of Mathematics, Wilfrid Laurier University, Waterloo, ON, Canada N2L 3C5*

Correspondence should be addressed to Qian Wang, q4wang@uwaterloo.ca

Received 22 December 2008; Accepted 22 March 2009

Recommended by Lean Yu

Real options modeling, which extends the ability of option pricing models to evaluate real assets, can be used to evaluate risky projects because of its capacity to handle uncertainties. This research utilizes possibility theory to represent private risks of a project, which are not reflected in the market and hence are not fully evaluated by standard option pricing models. Using a transformation method, these private risks can be represented as fuzzy variables and then priced with a fuzzy real options model. This principle is demonstrated by valuing a brownfield redevelopment project using a prototype decision support system based on fuzzy real options. Because they generalize the original model and enable it to deal with additional uncertainties, fuzzy real options are entirely suitable for the evaluation of such projects.

1. Introduction

An option is the right, but not the obligation, to buy or sell a certain security at a specified price at some time in the future [1]. The option pricing model developed by Black and Scholes [2] and Merton [3] is widely used to price financial derivatives. Because option pricing quantifies the values of uncertainties, this technique has migrated to broader usage, such as strategy selection [1], risky project valuation [4, 5], and policy assessment [6]. The idea of employing an option pricing model to value real assets or investments with uncertainties is usually called the real options approach or real options modeling [1, 7]. In real options, risky projects are modeled as a portfolio of options that can be valued using option pricing equations [4].

As options become "real" rather than financial, the underlying uncertainties become harder to deal with. Some risks associated with real options are not priced in the market, violating a basic assumption of option pricing. Hence, volatilities in real options usually cannot be accurately estimated.

The methods proposed to overcome the problem of private risk include utility theory [1, 7], the Integrated Value Process [4], Monte-Carlo simulation [8, 9], and unique and private risks [10]. This paper uses fuzzy real options, developed by Carlsson and Fuller [11], to represent private risk. Representing private risks by fuzzy variables leaves room for information other than market prices, such as expert experience and subjective estimation, to be taken into account. In addition, the model of Carlsson and Fuller can be generalized to allow parameters other than present value and exercise price [11] to be fuzzy variables, utilizing the transformation method of Hanss [12]. The added flexibility that allows the fuzzy real options model to tackle private risks issue also makes it more suitable for risky project evaluations.

We build a fuzzy real options model for brownfield evaluation by extending Lentz and Tse's [13], and develop a prototype decision support system for brownfield redevelopment evaluation to demonstrate the effectiveness of the fuzzy real options approach.

In this paper, option pricing models are introduced first and then used to value real assets as real options. The main issue of real options, private risks, will be addressed systematically. After a summary and comparison of methods to evaluate private risks, this research will focus on fuzzy real options. After a theoretical introduction, brownfield redevelopment, a typical risky project, is discussed briefly. Then a DSS for brownfield valuation based on fuzzy real options is designed and implemented as a prototype.

2. Fuzzy Real Options and Private Risks

2.1. Real Options

Black, Scholes, and Merton proposed their frameworks for pricing basic call and put options [2, 3] in 1973, establishing the theoretical foundation for pricing all options. Because option pricing models acknowledge the value of uncertainty explicitly, they came to be used to evaluate uncertain real assets, called real options. For instance, the managerial flexibility to terminate a project if it proves unprofitable was recognized as a kind of American put option, with sunk cost as the exercise price. Hence, the value of this risky project would be the sum of its initial cash flows and the value of its derivative American put option [8].

As suggested by the above example, the value of a risky project includes not only its present value, but also the portfolio of options associated with it, reflecting the values of uncertainties and associated managerial flexibilities. The following options may exist in different kinds of projects and situations and can be evaluated using option formulas developed for the financial market [1, 5, 7–9].

(i) *The option to defer*. The option of waiting for the best time to start a project can be valued as an American call option or a Bermuda call option.

(ii) *The option to expand*. The option of expanding the scale of the project can be valued as an American call option or a barrier option.

(iii) *The option to contract*. The option of shrinking the scale of the project can be valued as an American put option.

(iv) *The option to abandon*. The ability to quit the project can be valued as an American put option or a European put option.

(v) *The option of staging*. The ability to divide projects into several serial stages, with the option of abandoning the project at the end of each stage ("option on option"), can be valued as a compound option, also called the learning option in some articles.

(vi) *The option to switch*. The flexibility to change the project to another use can be valued as a switch option.

The output of option pricing models usually includes valuations, critical values, and strategy spaces to help a DM to make decisions [1]. This information includes

(i) *Valuations*. The most important output, and the main reason for using a real options model, is the value of the risky project.

(ii) *Critical values*. Threshold distinguishing the best strategy is usually defined in terms of parameters. For example, some critical values determine whether it is optimal to undertake the project. Critical values play a similar role to Net Present Value (NPV) zero.

(iii) *Strategy space*. The critical values divide the multidimensional strategy space into regions, corresponding to which option is best to implement. Often, this output is optional.

2.2. Private Risks

Unlike the uncertainties reflected in stock or bond prices or exchange rates, market data gives very little information about uncertainties in real options. Moreover, inappropriate consideration of uncertainties may make the real options model invalid, an important issue because of the basic assumptions of option pricing models [4, 14].

(i) *Complete market*. All risks can be hedged by a portfolio of options [15]. In other words, all risks have been reflected in the market price and can be replicated as options. In some literature, this is also called the Market Asset Disclaimer (MAD) approach [5].

(ii) *Arbitrage-free market*. There is no profit opportunity unless a player in the market is willing to take some risk [4]. In other words, there is no risk-free way of making money.

(iii) *Frictionless market*. There are no barriers to trading, borrowing, or shorting contracts, and no transaction costs for doing so. Furthermore, underlying assets are infinitely divisible [14].

These assumptions are generally realistic in the financial market, but may not be the case for real options, in part because of the many distinct sources of uncertainty. But more importantly, many uncertainties cannot be matched by any basked of market good, violating the complete market assumption [14]. In fact, it is unusual for project-specific uncertainties to be replicated as a market portfolio. The options modeling process must be customized to make the valuation framework flexible enough to fit real options.

In summary, private risk refers to risks that cannot be valued in the market [1], a simple but difficult issue in applying any real options model. Private risk challenges the complete market assumption, making the output values unreliable. Volatility (σ), which reflects uncertainties, cannot be estimated objectively.

2.3. Fuzzy Real Options

Projects usually have private risks, which usually cannot be estimated without expert knowledge. Using soft-computing techniques, experts can use their experience to make subjective predictions of private risk. These predictions can be improved using machine learning algorithms. Accumulating additional cases adds to expert experience, so that learning improves the accuracy of private risk estimation.

Among soft-computing techniques, fuzzy systems theory is especially suitable for representation of expert knowledge. Here, fuzzy real options are intended to deal with private risks that are hard to estimate objectively. The plan is to base fuzzy real options on possibility theory [11, 16–18].

The fuzzy approach cannot only model preferences [17], but also take into account subjective uncertainty [11]. In addition, it often requires less data, making it easier to use and quicker to produce satisfactory outputs.

Carlsson and Fuller assume that the present value and exercise price in the option formula are fuzzy variables with trapezoidal fuzzy membership functions. Because inputs include fuzzy numbers, the value of a fuzzy real options is a fuzzy number as well. A final crisp Real Option Value (ROV) can be calculated as the expected value of the fuzzy ROV [11]. Using α-cuts ($a^{(\alpha)}$ and $b^{(\alpha)}$ denote, respectively, the minimum and maximum values at the α-level of membership of a fuzzy variable), the possibilistic mean of the fuzzy variable (denoted A) and the variance was calculated by [11] to be (2.1) and (2.2), respectively.

$$E(A) = \int_0^1 \alpha\left(a^{(\alpha)} + b^{(\alpha)}\right) d\alpha, \tag{2.1}$$

$$\sigma^2(A) = \frac{1}{2}\int_0^1 \alpha\left(b^{(\alpha)} - a^{(\alpha)}\right)^2 d\alpha. \tag{2.2}$$

The concise and effective fuzzy real options approach proposed by Carlsson and Fuller is widely applicable; see [19–21]. But this specific model restrict fuzzy variables to the exercise price and current cash flow. Here, that restriction is avoided by employing fuzzy arithmetic and the transformation method, which is introduced in the following section.

2.4. Fuzzy Arithmetics and the Transformation Method

While fuzzy logics and inferencing systems are well-established, fuzzy arithmetic lags behind. Fuzzy arithmetic is restricted to simple operators [22, 23]. In applications, fuzzy arithmetic is usually limited to several predefined membership functions in MATLAB and Mathematica. This is mainly because the final result may be different depending on the procedure for implementing standard fuzzy arithmetic [12].

Hanss established the transformation method, thereby solving the multiple outputs problem and making generalized fuzzy arithmetic possible [24]. In the transformation method, a fuzzy function is defined as in Definition 2.1).

Definition 2.1. A fuzzy function with fuzzy output $\widetilde{q}$, n fuzzy inputs $\widetilde{p}_j$, and k normal inputs x_k, can be expressed as

$$\widetilde{q} = F\left(x_1, x_2, \ldots, x_k; \widetilde{p_1}, \widetilde{p_2}, \ldots, \widetilde{p_n}\right). \tag{2.3}$$

The idea underlying the transformation method has three important points: decompose fuzzy numbers into discrete form, use an α-cut for calculation purposes as a traditional function, and search the coordinates of the points in the hypersurfaces of the cube. The algorithm of the transformation method is given next [12]:

2.4.1. Decomposition of the Input Fuzzy Numbers

Similar to (2.1), $d\alpha$ is discretized into a large number of m intervals of length $\Delta\alpha = 1/m$. Each fuzzy input $\tilde{p}_i$ is decomposite into a series of crisp values at different α-cut levels μ_j, where $\mu_j = (j/m)\ (j = 0, 1, \ldots, m)$

$$P_i = \left\{ X_i^{(\mu_0)}, X_i^{(\mu_1)}, \ldots, X_i^{(\mu_j)}, \ldots, X_i^{(\mu_m)} \right\}, \quad i = 1, 2, ..., n, \tag{2.4}$$

where every element $X_i^{(j)}$ is define as

$$X_i^{(\mu_j)} = \left[a_i^{(\mu_j)}, b_i^{(\mu_j)} \right], \tag{2.5}$$

where $a_i^{(\mu_j)}$ and $b_i^{(\mu_j)}$ denote the minimum and maximum values at the μ_j-level of a given fuzzy variable, respectively, as previous.

2.4.2. Transformation of the Input Intervals

The intervals $X_i^{(j)}$, $i = 1, 2, \ldots, n$ of each level of membership μ_j, $j = 0, 1, 2, \ldots, m$, are transformed into arrays $\widehat{X_i^j}$ of the form with the number of 2^{i-1} pairs $(\alpha_i^{(j)}, \beta_i^{(j)})$,

$$\widehat{X_i}^j = \left(\left(\alpha_i^{(j)}, \beta_i^{(j)} \right), \left(\alpha_i^{(j)}, \beta_i^{(j)} \right), \ldots, \left(\alpha_i^{(j)}, \beta_i^{(j)} \right) \right) \tag{2.6}$$

with 2^{n-i} elements in each set of

$$\alpha_i^{(j)} = \left(a_i^{(j)}, \ldots, a_i^{(j)} \right), \qquad \beta_i^j = \left(b_i^j, \ldots, b_i^j \right). \tag{2.7}$$

The above formula fits the case of the reduced transformation method when the fuzzy function is monotonic or has only one fuzzy input. The general transformation method is similar to these formulae [12]. The main difference is that more points are tested.

Denotion of $\alpha_i^{(j)}$ and $\beta_i^{(j)}$ as pairs of α-cut values allows repetitive elements. The order of the elements is critical. The ultimate goal is that the 2^{i-1} of the ith element forms endpoints on the hypersurfaces. As illustrated in Figure 1, in the case of 3 fuzzy inputs shown as 3-dimensional space, the above definition means every $X_i^{(j)}$ has $2^{n-i} * 2^{i-1} = 2^n = 2^3 = 8$ elements, which are located on the endpoints of the cubicle.

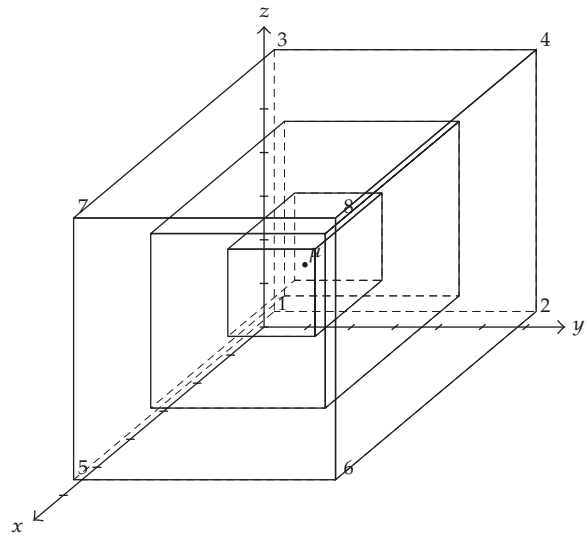


Figure 1: Fuzzy transformation diagram [12].

2.4.3. Evaluation of the Model

The function F is evaluated separately at each of the columns in the arrays using the classical arithmetic for crisp numbers. In other words, if the output $\tilde{q}$ can be decomposed to the arrays $\widehat{Z}^{(j)}$ $(j = 0, 1, 2, \ldots, m)$ using the algorithms mentioned above, the kth element can be obtained using the formula as

$$ {}^{k}\widehat{z}^{(j)} = F\left({}^{k}\widehat{x}_1^{(j)}, {}^{k}\widehat{x}_2^{(j)}, \ldots, {}^{k}\widehat{x}_n^{(j)}; y_1, y_2, \ldots, y_l \right), \tag{2.8} $$

where ${}^{k}\widehat{x}_i^{(j)}$ is the kth element of the array $\widehat{X_i}^{(j)}$ and y_i $(i = 1, 2, \ldots, k)$ are the other crisp inputs of the function.

2.4.4. Retransformation of the Output Array

Now the decomposition of the fuzzy output $\tilde{q}$ becomes the set $Q = \{Z^{(0)}, Z^{(1)}, Z^{(2)}, \ldots, Z^{(m)}\}$. Each element of this set should be the α-cut value at each level just like (2.5). Each value is obtained by retransforming the arrays $\widehat{Z}^{(j)}$ in a recursive manner using the following formula:

$$ \begin{aligned} a^{(j)} &= \min\left(a^{(j+1)}, {}^{k}\widehat{z}^{(j)} \right), \quad j = 0, 1, \ldots, m-1, \\ b^{(j)} &= \max\left(b^{(j+1)}, {}^{k}\widehat{z}^{(j)} \right), \quad j = 0, 1, \ldots, m-1, \\ a^{(m)} &= \min\left({}^{k}\widehat{z}^{(m)} \right), \qquad b^{(m)} = \max\left({}^{k}\widehat{z}^{(m)} \right). \end{aligned} \tag{2.9} $$

2.4.5. Recomposition of the Output Intervals

Recompositing the inverals $Z^{(j)}$, $j = 0, 1, 2, \ldots, m$ of the set Q based on their membership level μ_j, we can get the final fuzzy output $\tilde{q}$.

The transformation method is an ideal solution for implementing generalized arithmetic operations. Hence, it is employed in this paper to integrate the subjective uncertainties into the real options model. This algorithm is one of the key components in building the decision support system for the project evaluation of brownfield redevelopment using the fuzzy real options.

3. Brownfield Redevelopment

3.1. Uncertainties in Brownfield Redevelopment

A brownfield is an abandoned or underutilized property that is contaminated, or suspected to be contaminated, usually due to previous industrial usage [25]. Brownfields are common in cities transitioning from an industrial to a service-oriented economy, or when industrial enterprises have been relocated elsewhere or restructured themselves [26]. Brownfields are associated with an unsustainable development pattern, as they often arise when greenfields are developed while brownfields are abandoned.

Hence, brownfield redevelopment is helpful in enhancing regional sustainability. For example, municipal governments in Canada and elsewhere are encouraging brownfield redevelopment as part of a regional sustainable development plan. If brownfields were successfully redeveloped, local economic transitions would be more fluid; current infrastructure would be reused, and local public health would be more secure.

Brownfield redevelopment is a typical system-of-systems (SoSs) problem, as it involves various systems with complex interactions as illustrated in Figure 2. Brownfield redevelopment has the characteristics of an SoS; it possesses high uncertainty, nonlinear interactions within and among systems and is interdisciplinary in nature [27]. Due to the complex interactions of soil and groundwater systems with societal systems, as well as uncertainties in redevelopment costs, knowledge and technologies, and potentially high liabilities, brownfield redevelopment is difficult to initiate [28]. Uncertainties in brownfield redevelopment can be classified into the following categories.

(i) *Uncertainties due to limited knowledge of brownfields*. Currently, knowledge and data about brownfields are limited. Identifying appropriate models, characteristics, and parameters can be costly and time-consuming.

(ii) *Uncertainties originating from environmental systems*. Environmental systems have complex interactions in different systems, especially between groundwater and soil. Complex site-specific characteristics hinder remediation and redevelopment processes because they usually lead to highly uncertain remediation costs [29].

(iii) *Uncertainties originating from societal systems*. There are various kinds of stakeholders in brownfield redevelopment participating in complex conflicts and interactions, which create high levels of uncertainty in liabilities and cost sharing polices.

Because of the high uncertainties involved in brownfield redevelopment, traditional valuation methods are inoperable. It is very difficult to identify an appropriate discount rate for the Capital Asset Pricing Model (CAPM) [30]. Developers normally require high-risk premiums to compensate for the high uncertainties in brownfield redevelopment

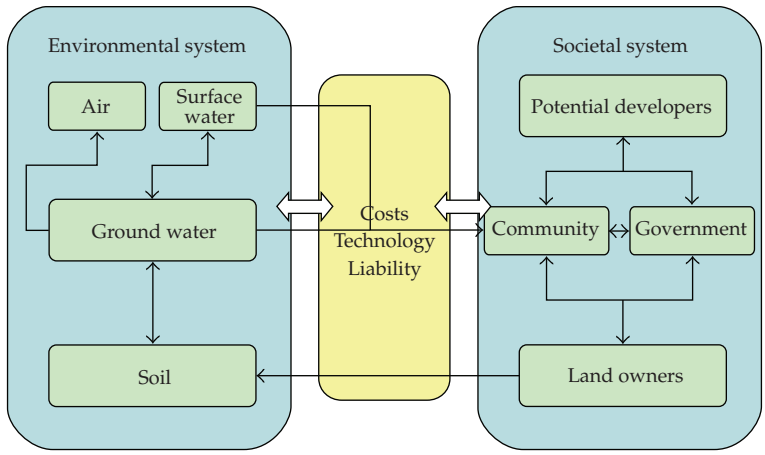


Figure 2: Systems diagram of brownfield redevelopment.

projects. Using Net Present Values (NPVs), developers usually calculated negative values for redevelopment projects and thus were reluctant to undertake them [31].

However, brownfield redevelopment can be profitable and even produces higher investment returns in some cases [31–33]. One explanation of the gap between predicted (conceptual) profit using NPV and the actual investment return is that the NPV method fails to map the value of opportunities created under a high uncertainty environment into project values.

These observations motivate the use of the real options model to evaluate redevelopment projects; it may provide more accurate valuations in the presence of high uncertainties. This research builds a prototype decision support system to implement a fuzzy real options model. The effectivenss of the model is tested using hypothetical data derived from actual brownfield redevelopment projects.

3.2. Fuzzy Real Options for Brownfield Redevelopment

Among a couple of available real options models for brownfield redevelopment, model proposed by Lentz and Tse is chosen to be extended with fuzzy variables, which includes an option to remove hazardous materials at the best time and an option to redevelop the brownfield, converting this site into other more profitable usage, at the best opportunity [13]. This model is more generic than others, such as Espinoza and Luccioni [32], in which only one American option is considered and Erzi-Akcelik [33], in which just applied Dixit and Pindyck's model [7]. Hence, an option to defer and option to learn are involved in the evaluation of contaminated properties.

The value of brownfield sites is regarded as two Wiener processes: the cash flow generated from this site without contamination (denoted as x) and the redevelopment cost for this site (denoted as R). To make private risks distinct from market ones, both of them are treated as two partially hedged portfolios, cash flow portfolio (denoted as P) and redevelopment cost (denoted as K), respectively.

In addition, four coefficient parameters with regard to x and R are involved. The parameters φ_1 and ϕ focus on cash flows. As cash flows from all states are being proportional

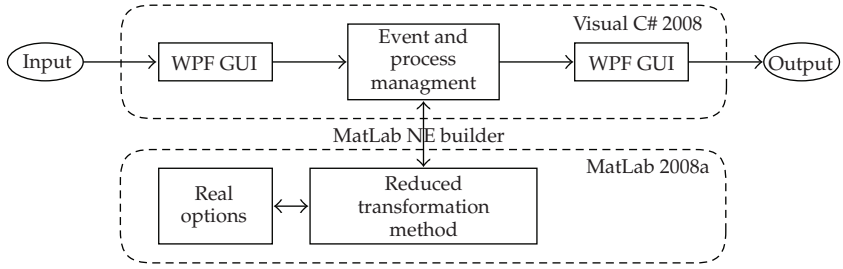


Figure 3: System architecture of the DSS.

to the clean one, the cash flow generated under contamination is $\varphi_1 x$; cash flow after removal is regarded as resumed to the clean one with x; and cash flow after redevelopment is ϕx. The coefficients α_1 and α_2 denote the removal and restoration costs as $\alpha_1 R$ and $\alpha_2 R$, which are assumed to be proportional to the total redevelopment cost R as well. Therefore, the cleanup cost C equals $(\alpha_1 + \alpha_2)R$.

Overall, three critical values are involved in deciding three kinds of strategies, which are denoted as Z^*, Y^*, and W^* in [13, Formula 14]: do nothing, remove pollutants and redevelop sequentially, or remove and redevelop simultaneously. Values to be compared with these critical values are, respectively, $Z = x/R$, the ratio of the clean cash flow (x) to the redevelopment cost (R), $Y = x/C$, the ratio of the clean cash flow (x) to the cleanup cost (C), and $W = x/(1+\alpha_1)R$, the ratio of the clean cash flow (x) to the combined cost of removal and redevelopment as a joint action. In addition, all supplementary formulae are shown as [13, Formula 12]:

$$\begin{aligned}
\sigma^2 &= \sigma_x^2 + \sigma_R^2 - 2\sigma_{xR}, \\
\gamma &= \omega_k - \mu_R, \\
\delta &= g - (\mu_R - \mu_K + r), \\
g &= \mu_x - (\mu_P - r)\,\beta_x, \\
\beta_x &= \rho_{xP}\frac{\sigma_x}{\sigma_P}, \\
\beta_R &= \rho_{RK}\frac{\sigma_R}{\sigma_K}, \\
\omega_K &= r + (\mu_K - r)\beta_R, \\
q &= 0.5\left(\frac{\sigma^2 - 2\delta}{\sigma^2} + \sqrt{\frac{(2\delta - \sigma^2)^2 + 8\gamma\sigma^2}{\sigma^4}}\right).
\end{aligned} \tag{3.1}$$

If the contaminated properties were to be cleaned and redeveloped sequentially, their values can be expressed as [13, Formula 13], depending on the critical value of Y^* as [13, Formula 14]. If $Y > Y^*$, the removal action should be taken right now. Otherwise, the optimal

executing time is in [13, Formula 16]. After the cleanup action, redevelopment is better to be conducted when $Z > Z^*$ in (3.5):

$$V1 = \begin{cases} \dfrac{\varphi_1 x}{r-g} + \left(\dfrac{(q-1)^{q-1}}{q^q}\right)\left(\left(\dfrac{1-\varphi_1}{r-g}\right)^q \left(\dfrac{x}{C}\right)^{q-1} + \left(\dfrac{\phi-1}{r-g}\right)^q \left(\dfrac{x}{R}\right)^{q-1}\right)x, & \text{if } Y \le Y^*; \\ \dfrac{x}{r-g} + \left(\dfrac{(q-1)^{q-1}}{q^q}\right)\left(\dfrac{\phi-1}{r-g}\right)^q \left(\dfrac{x}{R}\right)^{q-1} x - C, & \text{if } Y > Y^*; \end{cases} \tag{3.2}$$

$$\begin{aligned} Y^* &= \frac{r-g}{1-\varphi_1}\frac{q}{q-1}, \\ W^* &= \frac{r-g}{\phi-\varphi_1}\frac{q}{q-1}, \\ Z^* &= \frac{r-g}{\phi-1}\frac{q}{q-1}. \end{aligned} \tag{3.3}$$

If the brownfield sites were to be cleaned and redeveloped simultaneously, their values can be expressed in [13, Formula 15], depending on the critical value of W^* in (3.3). If $W > W^*$, the removal action should be taken right now. Otherwise, the optimal executing time is in (3.5):

$$V2 = \begin{cases} \dfrac{\varphi_1 x}{r-g} + \left(\dfrac{\phi-\varphi_1}{r-g}\right)^q \left(\dfrac{(q-1)^{q-1}}{q^q}\right)\dfrac{x^q}{(\alpha_1 R + R)}, & \text{if } W \le W^*; \\ \dfrac{\phi x}{r-g} - (\alpha_1 R + R), & \text{if } W > W^*; \end{cases} \tag{3.4}$$

$$\begin{aligned} \tau_Y &= \frac{\ln Y^* - \ln Y}{m_x - m_R}, \\ \tau_W &= \frac{\ln W^* - \ln W}{m_x - m_R}, \\ \tau_Z &= \frac{\ln Z^* - \ln Z}{m_x - m_R}, \end{aligned} \tag{3.5}$$

for $m_x > m_R$, where $m_x = \mu_x - 0.5\sigma_x^2$ and $m_R = \mu_R - 0.5\sigma_R^2$.

The final value of the brownfield site is the maximum of $V1$ and $V2$. An optimized redevelopment strategy can also be formed based on where it locates in the decision region, since all critical value can be converted into x/R.

4. Decision Support System Design and Case Study

4.1. System Architecture

The system architecture of the DSS is shown in Figure 3. Experts input parameters via the layer of the Windows Presentation Foundation (WPF) are given in Figure 4. After that, an event and process management module will control the work flow to convert all information

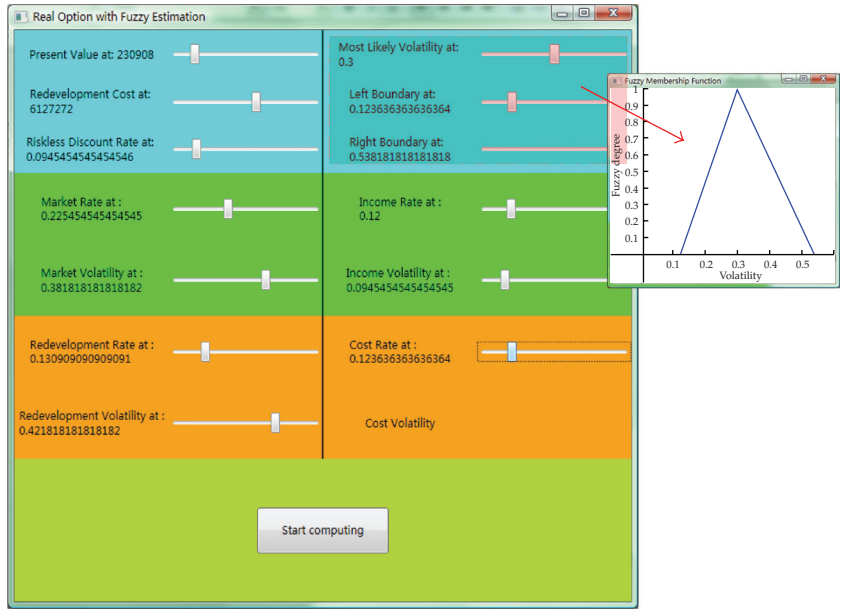


Figure 4: GUI for input of the DSS.

using the format of MatLab and feed them into MatLab for actual computation using the proposed fuzzy real options algorithm. Basically, fuzzy data are first converted (fuzzified and defuzzifed) into the crisp value needed for the real options model. Then, output can be obtained by calling the real options formula. Finally, the last output will be presented graphically via WPF to users.

Although prototype developed in this paper is primitive, the system architecture is quite generic and extendable. This DSS can be gradually expanded in scale and become more complex with more functions. The developed prototype satisfies the goal of a feasibility study on both the algorithm and technical approach in building this DSS.

4.2. An Illustrative Example

To demonstrate the fuzzy real options modeling, a brief example using hypothetical values is presented to illustrate an application of the associated DSS. These data mainly come from Lentz and Tse's paper [13], so that the DSS can be tested by comparing with their result. In addition, although inputs data are imaginary, they are modified according to real data in some articles on brownfield redevelopment, such as [33, 34] and added with some fuzzy parameters. The input data are shown in Table 1.

In addition to the data used in Table 1, all variables in this model are allowed to be fuzzy ones, in order to incorporate expert knowledge into parameter estimation. Given that the most difficult task in brownfield redevelopment is to estimate uncertainties regarding the redevelopment cost, which belongs to the private risk, the volatility rate of the redevelopment cost, σ_R, is deemed to be a fuzzy variable and studied intensively in this paper.

The volatility of the redevelopment cost is hard to estimate mainly because the dissemination of pollutants underground is highly complex. Soils, rocks, and materials

Table 1: Input data [13].

Variable name	Value
Current net cash flow of the clean property (x)	\$300000
Current redevelopment vost (R)	\$5000000
Riskless interest rate (r)	5%
Instantaneous return rate of the cash flow (μ_x)	10%
Volatility rate of the cash flow (σ_x)	20%
Instantaneous return rate of the portfolio hedging the cash flow (μ_P)	15%
Volatility rate of the portfolio hedging the cash flow (σ_P)	20%
Instantaneous return rate of the redevelopment cost (μ_R)	7%
Volatility rate of the redevelopment cost (σ_R)	20%
Instantaneous return rate of the portfolio hedging the redevelopment cost (μ_K)	15%
Volatility rate of the portfolio hedging the redevelopment cost (σ_K)	16%
Ratio of the contaminated cash flow to the clean one (φ_1)	0.4
Ratio of the restored cash flow to the clean one (φ_2)	1
Ratio of the redeveloped cash flow to the clean one (ϕ)	2
Ratio of the clean-up cost to the redevelopment cost (α_1)	0.3
Ratio of the restoration cost to the redevelopment cost (α_2)	0.2
Correlation between the hedging portfolio and underlying cash flow (ρ_{xP})	1
Correlation between the hedging portfolio and underlying cash flow (ρ_{RK})	1
Correlation between the cash flow and the redevelopment cost (ρ_{xR})	0

Note: Parameters, $\varphi_1, \varphi_2, \phi, \alpha_1, \alpha_2, \rho_{xP}, \rho_{RK}, \rho_{xR}$, are predefined in the DSS to simplify inputs. And since they are mainly coefficient parameters, there is no need to change them frequently.

are distributed ununiformly. Their hydraulic conductivities vary greatly according to the materials, elevation, and seasonal change. For instance, groundwater passes through peat (or cinders) at the velocity of 177 cm/d, which is hundreds of times the speed in the silt till (0.49 cm/d) at the site of the Ralgreen Community in Kitchener, ON, Canada [34]. Moreover, redevelopment cost also depends on the residual rate of pollutants and excavation cost, which are hard to estimate using market data neither.

To overcome this problem, the fuzzy redevelopment volatility is utilized as one with a triangle membership distribution based on parameters of minimum value, maximum value, and most likely value, because project managers and experts usually estimate uncertain parameters using the three-point estimation method [35]. Based on the hydraulic conductivity, volume of contaminated soil, and elevations, we found that the 20% volatility rate, used in Tse and Lentz's article [13], is roughly realistic. Nonetheless, since there are only two wells drilled for sampling, a relatively large interval should be added. As the result, the fuzzy redevelopment volatility is inputed as (0.15, 0.2, 0.25).

The main result from the DSS is shown in Figure 5, which includes the value of the brownfield site, a suggestion for a redevelopment strategy, and associated critical values that lead to this suggestion. In this case, the property value is a fuzzy variable with a mean of around 6.3 million and variance of around 1.7 million. Obviously, the private risk of redevelopment volatility has a great effect on the value of brownfield properties. And because the output indicators (Y and W) are less than their corresponding critical values (Y_* and W_*), this site is not worth of redeveloping now. This result partially explains why developers are reluctant to undertake this redevelopment task.

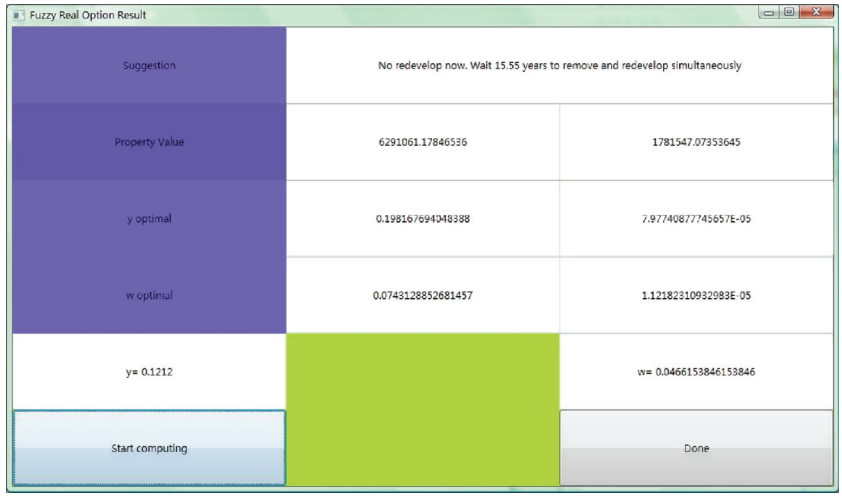


Figure 5: Output for the illustrative example.

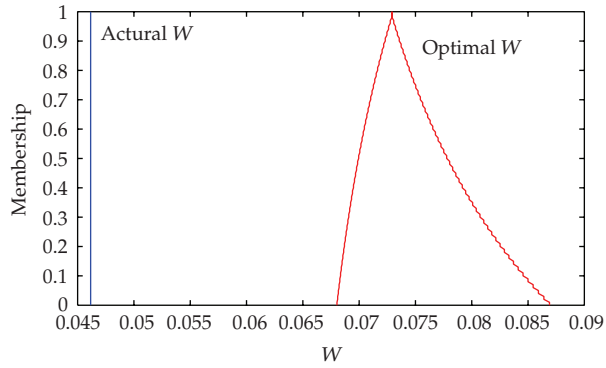


Figure 6: W (the red line in left) and W^* with fuzzy boundary (membership function).

Moreover, the critical values are fuzzy outputs as well. The fuzzy boundaries differentiating optimal strategies are illustrated in Figure 6. Also, these critical values can be converted into the ratio of x/R and shown in one figure as different decision regions in the strategy space (Figure 7). Fuzzy areas are calculated based on their fuzzy means and standard deviations. This DSS provides decision makers an intuitive decision suggestion with the aid of the decision region chart.

Finally, the effect of subjective estimation on property value is studied by changing the fuzzy intervals (minimum and maximum values of the fuzzy redevelopment volatility) while holding the most likely value unchanged, also as a kind of sensitivity analysis. Based on the illustrative example, we found that the value of the contaminated property increases as the fuzzy interval enlarges, as shown in Figure 8. This result implies that the subjective estimation of the private risk has an effect on the final evaluation result, although the change is not too much. Furthermore, experts can take advantage of their knowledge to make a slightly higher profit in the brownfield redevelopment projects than others. And the variation of the fuzzy output increases gradually, suggesting that there is no abrupt change and associated critical value in the fuzzy real options model.

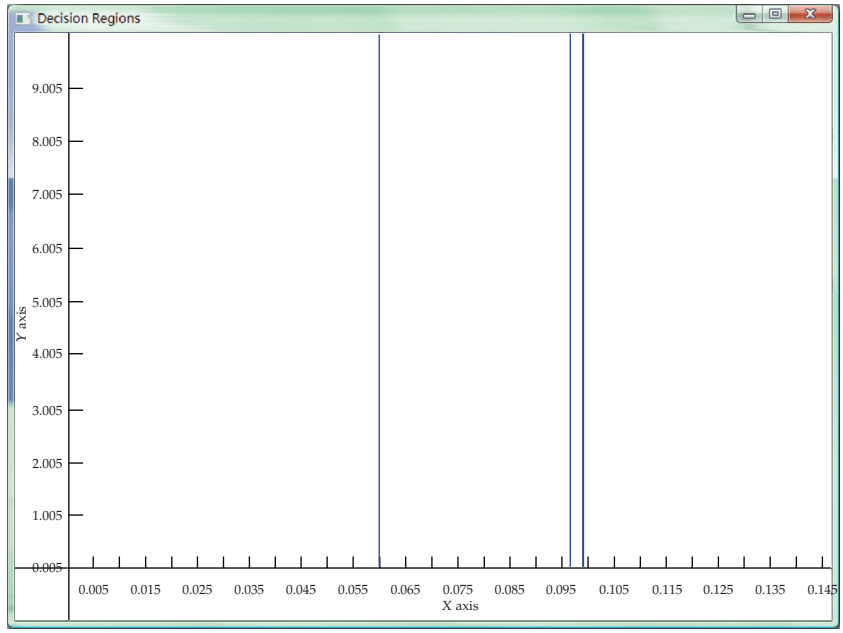


Figure 7: Decision regions divided by Y^* and W^*.

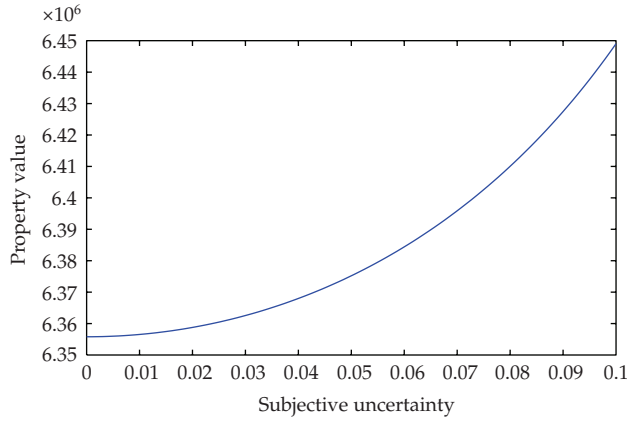


Figure 8: The effect of subjective uncertainty on property value.

5. Conclusions and Future Work

This paper employs a fuzzy real options approach to deal with the private risk problem. With help of the transformation method, any parameter in a real option model can be estimated as a fuzzy variable. Based on the results from the illustrative example using the prototype DSS, we found that this approach is effective in dealing with the private risk and generating satisfactory evaluations and useful suggestions. Hence, based on our limited test, the DSS based on the fuzzy real options is a useful tool in risky project evaluations. It potential for application should be further studied.

In addition, from the result of the illustrative example (Figure 8), we see that possibilities as well as probabilities can affect evaluations of risky projects. For the case of brownfield redevelopment, expertise can be utilized to make the contaminated property more valuable. This effect needs to be further analyzed.

But it should be recognized that the fuzzy real options model in this paper has several limitations, some of which may be removed in future work. Fuzzy arithmetic permits any membership function to be utilized in real options. This flexibility builds a foundation for future application of soft-computing techniques. For instance, the neural network could provide a nonparametric adaptive mechanism for private risk estimation. The DSS will be enhanced if it can be made to behave intelligently and adaptively.

A key feature of the fuzzy real options proposed here the mixture of fuzziness and randomness describing hybrid markets and private risks. But one limitation of our approach is that these two features are not well-integrated. Randomness is represented as a stochastic process, while fuzziness is represented using only fuzzy arithmetic. Incorporating a fuzzy process would clarify the structure, perhaps producing some important new insights.

A unifyied process including both fuzzy and stochastic features could strengthen the idea of fuzzy real options and extend enormous flexibility to soft-computing techniques and statistical models associated with these options. This goal may be achieved with the aid of the chance theory proposed by Liu [36].

Finally, since the fuzzy real options approach helps to improve developers' evaluation on brownfields, game-theoretic approaches could later be employed for negotiation of governmental assistance in the brownfield redevelopment to achieve optimal results.

References

[1] M. Amram and N. Kulatilaka, *Real Options: Managing Strategic Investment in an Uncertain World*, Harvard Business School Press, Boston, Mass, USA, 1999.

[2] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *Journal of Political Economy*, vol. 81, no. 3, pp. 637–659, 1973.

[3] R. C. Merton, "Theory of rational option pricing," *Bell Journal of Economics and Management Science*, vol. 4, no. 1, pp. 141–183, 1973.

[4] J. E. Smith and R. F. Nau, "Valuing risky projects: option pricing theory and decision analysis," *Management Science*, vol. 41, no. 5, pp. 795–816, 1995.

[5] M. Schneider, M. Tejeda, G. Dondi, F. Herzog, S. Keel, and H. Geering, "Making real options work for practitioners: a generic model for valuing R&D projects," *R&D Management*, vol. 38, no. 1, pp. 85–106, 2008.

[6] C. R. Cunningham, "Growth controls, real options, and land development," *The Review of Economics and Statistics*, vol. 89, no. 2, pp. 343–358, 2007.

[7] A. Dixit and R. Pindyck, *Investment under Uncertainty*, Princeton University Press, Princeton, NJ, USA, 1994.

[8] J. Mum, *Real Options Analysis: Tools and Techniques for Valuing Strategic Investments and Decisions*, John Wiley & Sons, New York, NY, USA, 2002.

[9] T. Copeland and V. Antikarov, *Real Options: A Practitioner's Guide*, Texere Press, New York, NY, USA, 2003.

[10] M. H. Mattar, *Private risk*, Ph.D. dissertation, MIT Press, Cambridge, Mass, USA, 2002.

[11] C. Carlsson and R. Fullér, "A fuzzy approach to real option valuation," *Fuzzy Sets and Systems*, vol. 139, no. 2, pp. 297–312, 2003.

[12] M. Hanss, *Applied Fuzzy Arithmetic: An Introduction with Engineering Applications*, Springer, Berlin, Germany, 2005.

[13] G. H. Lentz and K. S. M. Tse, "An option pricing approach to the valuation of real estate contaminated with hazardous materials," *The Journal of Real Estate Finance and Economics*, vol. 10, no. 2, pp. 121–144, 1995.

[14] M. H. Mattar and C. Y. J. Cheah, "Valuing large engineering projects under uncertainty: private risk effects and real options," *Construction Management and Economics*, vol. 24, no. 8, pp. 847–860, 2006.
[15] F. P. Ng and H. C. Björnsson, "Using real option and decision analysis to evaluate investments in the architecture, construction and engineering industry," *Construction Management and Economics*, vol. 22, no. 5, pp. 471–482, 2004.
[16] C. Carlsson and R. Fullér, "On possibilistic mean value and variance of fuzzy numbers," *Fuzzy Sets and Systems*, vol. 122, no. 2, pp. 315–326, 2001.
[17] D. Dubois and H. Prade, "Possibility theory as a basis for preference propagation in automated reasoning," in *Proceedings of the IEEE International Conference on Fuzzy Systems*, pp. 821–832, San Diego, Calif, USA, March 1992.
[18] D. Dubois and H. Prade, *Possibility Theory: An Approach to Computerized Processing of Uncertainty* , Plenum Press, New York, NY, USA, 1988.
[19] J. Wang and W.-L. Hwang, "A fuzzy set approach for R&D portfolio selection using a real options valuation model ," *Omega*, vol. 35, no. 3, pp. 247–257, 2007.
[20] D. Allenotor and R. K. Thulasiram, "G-FRoM: grid resources pricing a fuzzy real option model," in *Proceedings of the IEEE International Conference on e-Science and Grid Computing*, pp. 388–395, Bangalore, India, December 2007.
[21] J.-H. Cheng and C.-Y. Lee, "Product outsourcing under uncertainty: an application of fuzzy real option approach," in *Proceedings of the IEEE International Conference on Fuzzy Systems*, pp. 1–6, London, UK, July 2007.
[22] M. Ma, M. Friedman, and A. Kandel, "A new fuzzy arithmetic," *Fuzzy Sets and Systems*, vol. 108, no. 1, pp. 83–90, 1999.
[23] M. Mizumoto and K. Tanaka, "The four operations of arithmetic on fuzzy numbers," *Systems Computers and Controls*, vol. 7, no. 5, pp. 73–81, 1976.
[24] M. Hanss, "The transformation method for the simulation and analysis of systems with uncertain parameters," *Fuzzy Sets and Systems*, vol. 130, no. 3, pp. 277–289, 2002.
[25] K. W. Hipel, "Systems engineering approaches for brownfield redevelopment," October 2007, http://researchcommunity.iglooresearch.org/systemseng.
[26] United States Environmetnal Protection Agency (USEPA), "Characteristics of sustainable brownfields projects," Tech. Rep. EPA500-R-98-001G, USEPA, Washington, DC, USA, 1998.
[27] K. W. Hipel, M. M. Jamshidi, J. M. Tien, and C. C. White III, "The future of systems, man, and cybernetics: application domains and research methods," *IEEE Transactions on Systems, Man and Cybernetics, Part C*, vol. 37, no. 5, pp. 726–743, 2007.
[28] Q. Wang, K. W. Hipel, and D. M. Kilgour, "Conflict analysis in brownfield redevelopment: the erase program in Hamilton, Ontario," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC '08)*, Singapore, October 2008.
[29] D. El-Gamal, *Streamlined approach to vadose zone modeling: a contribution to brownfield redevelopment*, Ph.D. dissertation, Department of Civil and Environmental Engineering at Wayne State University, Detroit, Mich, USA, 2005.
[30] D. G. Luenberger, "A correlation pricing formula," *Journal of Economic Dynamics & Control*, vol. 26, no. 7-8, pp. 1113–1126, 2002.
[31] C. A. De Sousa, *The brownfield problem in urban canada: issues, approaches, and solutions*, Doctoral dissertation, University of Toronto, Toronto, Canada, 2000.
[32] City of Hamilton, "Hamilton ERASE (Environmental Remediation and Site Enhancement) Community Improvement Plan," Report, 2005.
[33] I. Erzi-Akcelik, *An analysis of uncertainty in brownfield redevelopment using real options*, Ph.D. dissertation, Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, Pa, USA, May 2002.
[34] Frontline Environmental Management Inc., "Restoration work plan for the ralgreen community," Internal Report 148-05, Frontline, Baltimore, Md, USA, November 2001.
[35] Project Management Institute, *A Guide to the Project Management Body of Knowledge*, Project Management Institute, Newton Square, Pa, USA, 2004.
[36] B. Liu, *Uncertainty Theory*, Springer, Berlin, Germany, 2nd edition, 2007.

Hindawi Publishing Corporation
Journal of Applied Mathematics and Decision Sciences
Volume 2009, Article ID 594793, 16 pages
doi:10.1155/2009/594793

Research Article

Discriminant Analysis of Zero Recovery for China's NPL

Yue Tang, Hao Chen, Bo Wang, Muzi Chen, Min Chen, and Xiaoguang Yang

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100090, China

Correspondence should be addressed to Xiaoguang Yang, xgyang@iss.ac.cn

Received 7 December 2008; Revised 20 February 2009; Accepted 23 March 2009

Recommended by Lean Yu

Classification of whether recovery of non-performing loans (NPL) is zero or positive is not only important in management of non-performing loans, but also is essential for estimating recovery rate and implementing the new Basel Capital Accord. Based on the largest database of NPL's recovering information in China, this paper tries to establish discriminant models to predict the loan with zero recovery. We first use Step-wise discrimination method to select variables; then give an in-depth analysis on why the selected variables are important factors influencing whether a loan is zero or positive recovery rate. Using the selected variables, we establish two-type discriminant models to classify the NPLs. Empirical results show that both models achieve high prediction accuracy, and the characteristics of obligors are the most important factors in determining whether a NPL is positively recovered or zero recovered.

1. Introduction

Credit risk is the major risk commercial banks are faced with, hence, measurement and management of credit risk is the core task of risk management. In China, commercial banks' credit risk mainly features the accumulation of a large number of nonperforming loans (NPLs), so it is critical to model various risk factors in NPLs in order to establish a sound credit risk management system.

Loss given default (LGD) is an equivalent concept with recovery rate. LGD is a critical parameter for measurement of credit risk, a basis for estimating expected loss (EL) and unexpected loss (UL). Modeling the LGD of NPLs effectively is very important for the management of NPLs and banking supervision.

The key issues of LGD (Schuermann [1]) are the following: definition and measurement; key drivers, modeling, and estimation approaches. The following studies explored the characteristics of Bond LGD.

Altman and Kishore [2], for the first time, documented the severity of bond defaults stratified by Standard Industrial Classification sector and by debt seniority. They found that the highest average recoveries came from public utilities (70%) and chemical, petroleum, and related products (63%). They concluded that the original rating of a bond issue as investment grade or below investment grade had virtually no effect on recoveries once seniority was accounted for. Neither the size of the issue nor the time to default from its original date of issuance had any association with the recovery rate. Acharya et al. [3] found that recoveries on individual bonds were affected not only by seniority and collateral but also by the industry conditions at the time of default. Altman et al. [4] analyzed the impact of various assumptions about the association between aggregate default probabilities and the loss given default on bank loans and corporate bonds and sought to empirically explain this critical relationship. They specified and empirically tested for a negative relationship between these two key inputs to credit loss estimates and found that the result was indeed significantly negative.

Currently, the bank loan LGD is not explored well by theoretical and empirical literature. In many cases, the subject of many studies on LGD is bond rather than loan. Little research has been done on the LGD of loans. There are almost no public loans' LGD models. Some of the most important research focusing on the bank loan markets is the following.

Carty and Lieberman [5] measured the recovery rate on a sample of 58 bank loans for the period 1989–1996 and reported skewness toward the high end of price scale with the average recovery of 71%. Gupton et al. [6] reported higher recovery rate of 70% for senior secured loans than for unsecured loans (52%) based on 1989–2000 data sample consisting of 181 observations.

The above studies focused on the U.S market. Hurt and Felsovalyi [7] who analyzed 1149 bank loan losses in Latin America over 1970–1996 found the average recovery rate was 68%. None of the above studies provided the information on predicting bank loan LGD.

Moody has established special LGD prediction models, which are known as LossCalc models. But most of the data sample used in the models is bond data; the sample includes only a small portion of the loan data. More details about LossCalc can be found in Gupton [8]. However, the NPLs in China's commercial banks have their own special characteristics. The factors affecting NPL recovery are also not the same with abroad. Therefore, it is essential to develop specific LGD models in accordance with the actual situation of China's commercial bank.

Now, domestic research on LGD is mainly concentrated in the qualitative discussion. Chen [9] discussed the importance of LGD. Moreover, he introduced several international modeling methods for LGD and discussed the major difficulty in modeling LGD in China.

Wu [10] analyzed the necessary conditions for banks to estimate LGD and explored the methods to estimate LGD under the Internal Ratings Based (IRB) framework. Liu and Yang [11] summarized and analyzed the international discussion about the performance of LGD and factors impacting LGD. Moreover, he thought that data was the basis for modeling LGD and suggested that in order to make up for deficiency in data, the central bank should take the lead to establish the joint NPL database. Shen [12] compared many new methods of modeling LGD, including the nonparametric and neural network models.

Studies mentioned above present valuable guidance for research on relevant domestic issues. But these studies are limited to qualitative discussions, so they cannot provide substantive deeper conclusions.

Ye and Liu [13] have carried out some empirical research with the relevant data from banks. With the data on defaulted loans from a bank in Wenzhou, a city in Zhejiang Province,

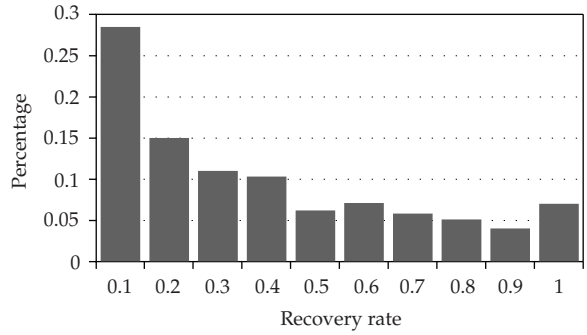


Figure 1: Distribution of recovery rate.

Ye and Liu [13] analyzed the LGD distribution characteristics of NPLs in terms of exposure at default, maturity structure, mortgage manner, industry, and so forth.

Empirical studies mentioned above laid a good foundation for domestic research on LGD, but they are limited by their very small data sample which does not reflect the national situation. Moreover, the analysis tools in the above studies are just simple descriptive statistics. It is hard to convince people from their results.

Data samples used in this paper come from the LossMetrics Database. LossMetrics is a large-scale database concerning NPLs, established by Doho Data Consulting Corporation. This database consists of rich and detailed information about NPLs coming from the four major commercial banks in China, namely, Industrial and Commercial bank of China (ICBC), China Construction Bank (CCB), Agricultural Bank of China (ABC), and Bank of China (BOC). Nearly tens of thousands of obligors' pieces of information are included in this database and the amount of NPLs comes to 600 billion Renminbi(RMB). These NPLs cover more than 20 industries and 10 provinces and can reflect the overall situation and distribution characteristics of China's NPLs.

Research both at home and abroad shows that the recovery rate appears as a bimodal distribution. According to Gupton [8], the recovery rate has peaks in the vicinity of 20% and 80%, respectively. But with China's data analysis, the peaks are located in the vicinity of 0 and 100%, respectively; the peak near zero is very high, while the peak near 100% is relatively low. According to the results of statistical data, the mean of NPLs' recovery rate in China's AMC is about 30%, which is much lower than the recovery rate in developed countries. More than 25% NPLs' recovery rate is 0, and nearly 10% NPLs' recovery rate is almost 100%. Figure 1 shows the distribution of recovery rate of NPLs in China.

Considering the bimodal features of recovery rate's probability distribution, estimating recovery rate directly will lead to a big bias.

It is effective and necessary to model recovery rate in two steps. The first step is to classify recovery into different categories: zero recovery and positive recovery. This paper devotes to the classification. This classification is not only a part of estimating LGD, more importantly if the recovery of an NPL can be determined to be zero or positive, the information can help commercial banks and AMCs to manage and dispose NPLs as they can allocate more resources to assets with positive recovery and avoid wasting money and time on loans with zero recovery, which help reduce financial cost and improve management efficiency.

In this paper, discriminant analysis method is used to classify recovery. Basic idea of discriminant analysis can be illustrated as follows: assuming that the research subject can

be classified into several categories, and there are observation data of given samples in each category. Based on this assumption, we can establish discriminant functions according to some criteria, and use the function to classify the samples with unknown category. There are plenty of discriminant analysis methods, such as Distance, Fisher, Sequential, Bayesian, and Stepwise discrimination. This paper applies Stepwise, Distance, and Bayesian discrimination to classify the recovery of NPLs and compares the results between Distance discrimination and Bayesian discrimination as well. More details about discriminant analysis can be found in Bruch [14].

Rest of the paper is organized as follows. Section 2 describes the data used to model and selects significant variables for classifying recovery by Stepwise discrimination. Section 3 presents an in-depth analysis why the selected variables have a significant influence on discrimination of zero recovery and positive recovery. Section 4 establishes several discrimination models and utilizes these models to predict. Section 5 concludes and discusses pros and cons of the current research and puts forward expectations and extensions for future research.

2. Data and Variable Selection

2.1. Data Samples

Generally speaking, an obligor can borrow one or several loans. Banks or AMCs collect NPL from obligors in a way they collect all the loans from an obligor as a whole. To give a clear classification, we first sift the obligors with only one loan from the database and start our analysis and modeling process with these obligors as samples; then we use the sample of obligors with several loans to test if the models still apply.

From LossMetrics database, we obtain 625 obligors with only one loan, among which, 425 obligors have positive recovery while 200 other obligors have zero recovery. The number of the obligors with several loans is 821, out of which, 592 have positive recovery, and 229 do not have recovery. The total number of loans mounts to be 4061.

2.2. Variable Selection

The variables associated to NPL can be divided into, broadly, two categories: debt information and obligor information.

Debt information refers to the five-category assets classification for bank loans, collateral, NPLs' transfer mechanism, and so forth. The five-category assets classification and for bank loans is a standard established by China's commercial banks, which can be referred to assess the quality of loans. It includes five categories: normal, concerned, subordination, doubt, and loss. Since all the loans studied are nonperforming loan, the samples don't cover the loans belonging to normal or concerned category. NPLs' transfer mechanism refers the ways to transfer NPLs from commercial banks to AMCs. One way is policy-off, that is, the central government buys NPLs from the commercial banks and gives them free to AMCs, and the other way is business acquisition, that is, AMCs buy NPLs directly from commercial banks.

Obligor information refers to obligor's business situation, the industry, and region obligor being in. Business situation can be divided into seven categories: normal, financial distress, ceasing-business, shutting-down, under-insolvency-process, bankruptcy, and unknown.

Table 1: The result of variable selection.

Significant variables	Partial-r square
Loss-category loan	0.4882
Bankruptcy	0.1611
Shutting-down	0.1229
Ceasing-business	0.061
Collateral loan	0.0437
Under-insolvency-process	0.0406
Unknown	0.0241
Business acquisition	0.0142
Tianjin area	0.0148
Fujian area	0.0132
Real estate industry	0.0128
Hebei area	0.0079
Retail industry	0.0054

As there are very few data in the database containing financial indices of the obligors, we can only use the indicator variables in the modeling process.

In order to avoid too many variables to interfere the effect of discrimination, we employ stepwise discrimination method to select the significant variables.

The result of variable selection is displayed in Table 1.

As Table 1 shows, the larger the Partial-r square of a variable is, the more important the variable is. In Table 1, the impact of five-category classification for loans, the business situation of companies, region, and industry is significant. while business situation of company and the quality of loans (five-category classification) play a leading role in discrimination with a proportion up to 94%, in contrast, the factors such as industry and region turn out to be less important.

3. Analysis of Selected Variables

In the above section, we use Stepwise discrimination to select some variables which are significant for distinguishing zero recovery and positive recovery from the view of discriminant function. In order to have a deeper understanding of these variables and hence to be more confident in predicting models in next section, we direct a further probe in this section to show that these variables have bigger impact on zero recovery of NPLs than other unselected variables.

In the linear discriminant function developed by Distance method, the sign of the coefficient of the variable indicates whether the variable acts positively or negatively; however, Bayesian discrimination can compute the probability of the recovery of a loan being zero or positive; therefore, it is more convenient and more accurate to use Bayesian method, since the changes in variables can be connected with the probability. Here, we use Bayesian method to establish linear discrimination function and see how these variables affect zero recovery.

The basic idea of Bayesian discriminant is as follows. Let the Bayesian linear discriminant function for positive recovery be $f_1 = a + a_1x_1 + a_2x_2 + \cdots + a_nx_n$, and the Bayesian linear discriminant function for zero recovery be $f_2 = b + b_1x_1 + b_2x_2 + \cdots + b_nx_n$. For

Table 2: Bayesian discriminant function.

Variables	Coefficient 1	Coefficient 12	Difference
Intercept	−2.389	−8.454	6.065
Loss-category loan	1.834	6.851	−5.018
Bankruptcy	1.656	10.521	−8.866
Shutting-down	1.785	7.706	−5.922
Ceasing-business	1.755	5.819	−4.064
Collateral loan	1.743	−0.110	1.853
Under-insolvency-process	0.196	8.673	−8.478
Unknown operation	2.700	5.593	−2.894
Policy-off	3.683	2.093	1.590
Tianjin area	3.339	0.633	2.706
Fujian area	1.694	−1.223	2.917
Real estate industry	1.966	−0.226	2.192
Hebei area	1.617	0.539	1.079
Retail industry	1.901	1.127	0.774

an NPL, the posterior probability of positive recovery will be $p_1 = (e^{f_1}/e^{f_1} + e^{f_2}$ as $e^{f_1})$, and the posterior probability for zero recovery will be $p_2 = (e^{f_1}/e^{f_1} + e^{f_2}$ as $e^{f_2})$. When $p_1 > p_2$, that is $p_1/p_2 = e^{c+c_1x_1+c_2x_2+\cdots+c_nx_n} > 1$, the NPL will be classified into the positive category, or else zero category. Here, $c_k = a_k - b_k$; if $c_k < 0$, this variable will have a positive influence on the NPL and is to be classified into positive category. That is, if this variable changes from 0 to 1 and the other variables keep constant, the probability for positive recovery will become larger; if $c_k > 0$, the probability will become smaller. Moreover, if a variable x_k changes from 0 to 1, the change of the probability for positive recovery will be e^{c_i} times of the probability for zero recovery. Therefore Bayesian linear discriminant function can be used in analyzing the classification efficiency of variables.

Using the data samples obtained from Section 2, we establish the Bayesian discriminant function as follows.

These variables can be classified into 6 classes: five-category classification, collateral properties, transfer mechanism of NPLs, operation situation of the obligors, industry, and region. Now, let us see how these variables affect the classification of NPLs.

3.1. Five-Category Loan Classification

The variable five-category classification plays a leading role in classification based on the results of stepwise discrimination. Only "loss category loan" has entered into the model, because the effects of "subcategory loan" and "doubt-category loan" for classification have been mainly reflected by "loss-category loan".

In the discriminant function, the difference of the coefficient of this variable equals −5.018, which shows that when a nonperforming loan belongs to loss category, the probability of its recovery being zero will be larger than the probability of being positive. Keep the other variables constants; if a loan enter a loss category from other categories, the changes of the probability for zero recovery are $e^{5.018}$ times of the changes of the probability. Combined with Figure 2 it will help understand why five-category classification has played such an

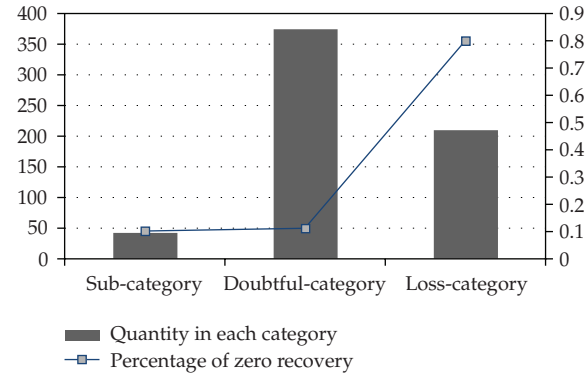


Figure 2

important role. In Figure 2 , the blue point represents the ratio of the NPLs not having recovery in each category.

In Figure 2 , the ratio of zero recovery has decreased along with the grade of loans. The ratio for subcategory and doubt-category is very low, less than 10%, while the ratio for loss-category loans is up to 80%. It shows that, first, whether a loan has recovery is mainly determined by the quality of loan. The worse its quality is, the larger the probability of zero recovery will be. Second, the standard of five-category classification built by China's commercial bank can differentiate the quality of loans effectively. So we should continue to pay much attention to the change of the grade of loans and improve the ability to distinguish the qualities of loans, which will help manage credit risk more effectively.

3.2. Business Situation of Obligors

The significant variables of this type are bankruptcy, under-insolvency-process, shutting-down, and ceasing-business. The variables "normal" and "financial distress" do not enter into the model due to multicollinearity between these variables. And obviously the ability to repay in both cases will be much greater than the other cases.

The differences of the coefficients for these significant variables between the two functions are negative, which means that the recovery rate of NPLs in these cases will be inclined to be classified into zero. Let us see Figure 3.

In Figure 3, we can see that there are significant differences among the ratios of zero recovery in different cases. When the obligors run well or in financial distress, the ratio of their NPLs with positive recovery is very high. Especially when the obligors operate well, the recovery rates of NPLs in this category are all positive.

When the business situation of an obligor is in a state of bankruptcy, under-insolvency-process, shutting-down, or ceasing-business, the ratio of their NPLs with zero recovery is very high; the highest is the obligors in the state of bankruptcy.

We can draw some conclusions from these results: the business situation of obligors reflects their payment ability to a great extent. The worse the business situation is, the weaker the payment ability will be, and vice versa. When obligors enter the states of bankruptcy, shutting-down, or ceasing-business, they have no ability to repay their loans, which will lead to high ratio of NPLs with zero recovery. While obligors run well or even if they are in financial distress, the cash flow of business is still in existence, which makes the probability

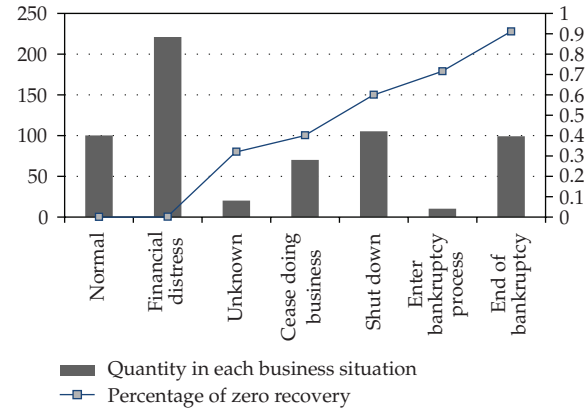


Figure 3

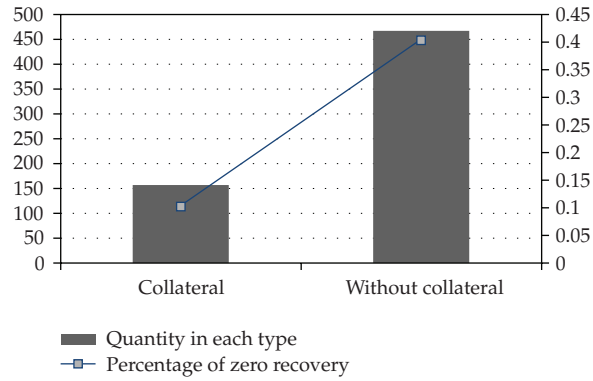


Figure 4

of recovery much larger. So when disposing NPLs, the business situation of enterprises should be paid much attention to. If the business situation can change with time, the NPLs of enterprises should have been disposed when the enterprises run relatively well as far as possible, which has positive effect.

3.3. Collateral

Due to the quality of the data, we cannot get the collateral value; so only the state variable, that is, "whether the loan has effective mortgage", is contained in the model.

In the discriminant function, whether the loan has effective collateral is significant, and the difference between the coefficients in two types of discriminant function is 1.961. This shows that effective collateral has a positive effect in obtaining positive recovery. Keep the other variables constant; when a loan changes from noncollateral to collateral, the change in the probability of positive recovery is $e^{1.961}$ times of the change in the probability of zero recovery. Figure 4 has made this point more clearly in a straight way.

As Figure 4 shows, the ratio of collateral NPLs with zero recovery is 10%; this ratio is approximate 44% for the NPLs without collateral. This result shows that with effective

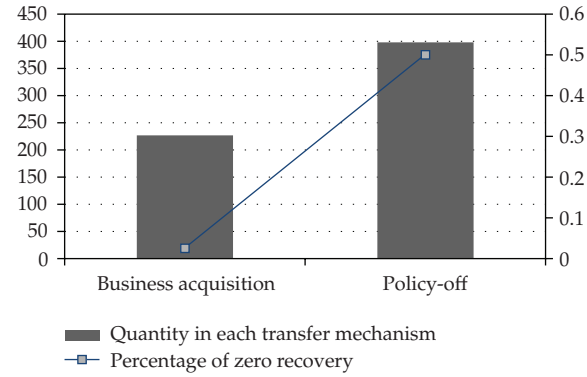


Figure 5

collateral, the probability of positive recovery will increase greatly. This is because the collateral can be used to negotiate with the obligor or sold straightly in the secondary market. When purchasing, managing, and disposing NPLs, the collateral should be paid much attention to, which might determine whether NPLs have recovery and how much could be got from the obligors.

3.4. Transfer Mechanism of NPLs

The variable "policy-off" is significant in Stepwise discrimination. The difference of coefficient between the two discriminant functions is 1.174, which implies that the probability of positive recovery will be larger than that of zero recovery when NPLs are commercially bought from banks. That is, the quality of NPLs by business acquisition may be better than the quality of NPLs by policy-off. More details can be seen in Figure 5.

As is shown in Figure 5, the ratio of NPLs by business acquisition with positive recovery is almost 100%, while the ratio is much lower for NPLs by policy-off. This implies that the power of market promotes healthy development of NPLs' management, disposal, and recovery.

3.5. Region

Some of the provinces are engaged into in the model, such as Tianjin, Fujian, and Hebei. Combined with the result of Stepwise discrimination, the factor of "province" has a relative small effect in classification. Further, when using the other significant variables except "province", the result of the classification has little change. The results of classification are well stated in the diagram below. In combination of Figure 6, we can do more in-depth analysis.

Ignoring the provinces with very few samples (Beijing, Hainan and Guangdong) in Figure 6 , the ratio of NPLs with zero recovery is relatively high, coming between 40% and 60%, provided that the obligors are located in Liaoning, Jiangxi, Shandong and Guangxi province. These provinces, other than Shandong province, are less developed regions, especially Jiangxi and Guangxi. The ratio of NPLs with zero recovery is relatively high in Shandong province because economic development is unbalanced there. Further analysis

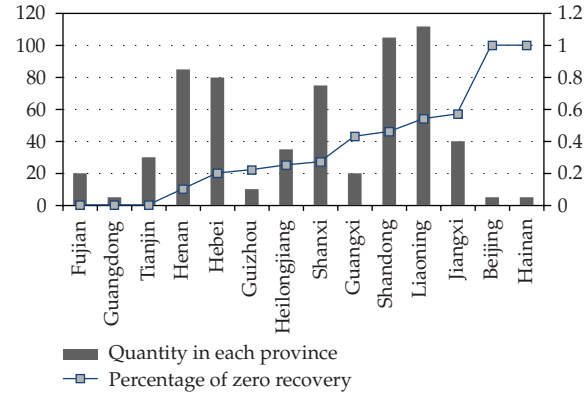


Figure 6

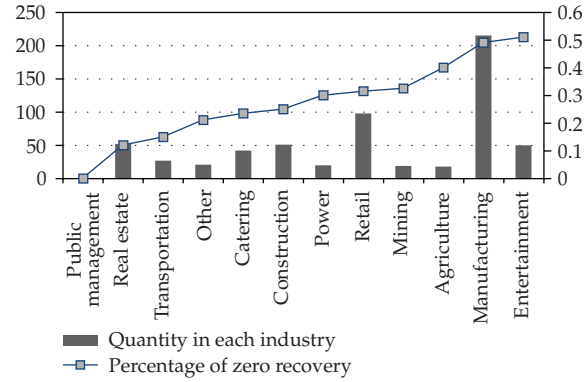


Figure 7

finds that most of the NPLs with zero recovery are from less developed cities. This may lead to the high ratio of NPLs with zero recovery.

The two highest ratios of NPLs with positive recovery are from Fujian and Tianjin. In these two provinces, NPLs commonly have positive recovery because both Fujian and Tianjin are developed regions in China. Less developed Henan province also has a relatively low ratio of NPLs with zero recovery, though it has not been referred in the model. A further analysis of the sample finds that most of the NPLs in Henan belong to subcategory and doubt-category loans and are bought from commercial banks. This is why the ratio of NPLs in Henan with positive recovery is high and variable, "Henan" is not significant in Stepwise discrimination model. Compared with region factors, the quality of loans plays a major role in determining whether they have recovery. To sum up, the economic conditions in the region surely influence the recovery of NPLs. Generally speaking, the more developed their economy is, the larger the probability of positive recovery is. Nevertheless, the factors about loan quality, such as the five-category classification of loans and the business situation of companies etc., play a more important role than region factor in determining whether loans are with positive recovery.

3.6. Industry

The results of Stepwise discrimination show that variables of industry have a small effect on the classification; only real estate industry and retail industry show a significant impact.

Table 3: The prediction results of whole samples.

			Positive recovery (predicted)	Zero recovery (predicted)	Accuracy
Whole sample	Positive recovery	Bayesian Discrimination	406	19	95.5%
		Distance	395	30	92.9%
	Zero recovery	Bayesian	19	181	90.5%
		Distance	14	186	93.0%
	Total accuracy		Bayesian		93.9%
			Distance		93.0%

Figure 7 shows that the ratio of NPLs with zero recovery in real estate is approximate 13%, which implies that most of NPLs in the real estate industry have positive recovery. This is in line with the actual situation in real estate industry, which has been developing fast over the past 10 years in China. The other industries of which the ratios are also low are not significant. This is mainly because the effect of industry on classification has been categorized into the business situation and five-category classification, as concluded by analyzing corresponding samples in these industries.

4. Predicting Models and Accuracy of Prediction

Business situation, five-category classification, collateral state, province, and industry have been proved to be important in determining whether recovery rate of NPLs is zero or positive, according to discriminant function and data samples. In the sequel, we will use these variables to establish some discrimination models and verify the predicting competence of these models. Both Bayesian discrimination and Distance discrimination will be used to build predicting models with the variables selected by the Stepwise discrimination.

The most commonlyused method of modeling is to build models with the whole sample and then compare the difference between the predicted data and the actual amount; however, the adaptability of model established in this way needs to be considered carefully. The model based on the whole samples may be well adapted to the original samples but may also act badly for new samples. This is the so-called "over-fitting" problem. In order to solve this problem, the cross-validation method is employed. The whole sample is then divided into five equally subsamples. Each subsample has 125 observations which consist of 85 observations with positive recovery and 40 with zero recovery. Four of the five subsamples are adopted to fit a model each time, and the remained one would be used for prediction. The mean accuracy of the five submodels would be considered as the standard to select the best model. In this paper, the predicting efficiency is compared between Bayesian discrimination and Distance discrimination. The reason for the sample to be divided in this way lies in consideration of balance of sample size, computing complexity and the consistency of the statistical properties.

The results of the prediction out of whole samples by Bayesian discrimination and Distance discrimination are shown in Table 3 .

The prediction results using cross-validation in the two discriminating methods are shown in the following tables.

Table 4: The prediction results of the first model using cross-validation.

				Positive recovery (predicted)	Zero recovery (predicted)	Accuracy
The first model	In-Sample	Positive recovery	Bayesian	323	17	95.0%
			Distance	319	21	93.8%
		Zero recovery	Bayesian	14	146	91.3%
			Distance	7	153	95.6%
		Total accuracy		Bayesian		93.8%
				Distance		94.4%
				Positive recovery (predicted)	Zero recovery (predicted)	
	Out-of-Sample	Positive recovery	Bayesian	77	8	90.6%
			Distance	72	13	84.7%
		Zero recovery	Bayesian	5	35	87.5%
			Distance	5	35	87.5%
		Total accuracy		Bayesian		89.6%
				Distance		85.6%

Table 5: The prediction results of the second model using cross-validation.

				Positive recovery (predicted)	Zero recovery (predicted)	Accuracy
The second model	In Sample	Positive recovery	Bayesian	323	17	95.0%
			Distance	321	19	94.4%
		Zero recovery	Bayesian	9	151	94.4%
			Distance	8	152	95.0%
		Total accuracy		Bayesian		94.8%
				Distance		94.6%
				Positive recovery (predicted)	Zero recovery (predicted)	
	Out-of-Sample	Positive recovery	Bayesian	75	10	88.2%
			Distance	73	12	85.9%
		Zero recovery	Bayesian	1	39	97.5%
			Distance	1	39	97.5%
		Total accuracy		Bayesian		91.2%
				Distance		89.6%

Some further results can be calculated based on the results given in the above tables.

In Bayesian discrimination, the total mean in-sample accuracy of the five models is 94%, with 94.7% accuracy for positive recovery and 92.5% accuracy for zero recovery; the mean total out-of-sample accuracy of the five models is 92.64% with 93.6% accuracy for positive recovery and 90.5% accuracy for zero recovery. While in Distance discrimination,

Table 6: The prediction results of the third model using cross-validation.

				Positive recovery (predicted)	Zero recovery (predicted)	Accuracy
TheThird model	In-Sample	Positive recovery	Bayesian	317	23	93.2%
			Distance	316	24	92.9%
		Zero recovery	Bayesian	10	150	93.8%
			Distance	9	151	94.4%
		Total Accuracy		Bayesian		93.4%
				Distance		93.4%
				Positive recovery (predicted)	Zero recovery (predicted)	Accuracy
	Out of Sample	Positive recovery	Bayesian	83	2	97.6%
			Distance	81	4	95.3%
		Zero recovery	Bayesian	2	38	95.0%
			Distance	2	38	95.0%
		Total Accuracy		Bayesian		96.8%
				Distance		95.2%

Table 7: The prediction results of the fourth model using cross-validation.

				Positive recovery (predicted)	Zero recovery (predicted)	Accuracy
The fourth Model	In-Sample	Positive recovery	Bayesian	324	16	95.3%
			Distance	314	26	92.4%
		Zero recovery	Bayesian	17	143	89.4%
			Distance	10	150	93.8%
		Total accuracy		Bayesian		93.4%
				Distance		92.8%
				Positive recovery (predicted)	Zero recovery (predicted)	Accuracy
	Out-of-Sample	Positive recovery	Bayesian	82	3	96.5%
			Distance	78	7	91.8%
		Zero recovery	Bayesian	2	38	95.0%
			Distance	1	39	97.5%
		Total accuracy		Bayesian		96.0%
				Distance		93.6%

the mean total in-sample accuracy of the five models is 93.88% with 93.59% accuracy for positive recovery and 94.5% accuracy for zero recovery; the mean total out-of-sample accuracy of the five models is 90.24% with 89.6% accuracy for positive recovery and 91.5% accuracy for zero recovery.

Table 8: The prediction results of the fifth model using cross-validation.

				Positive recovery (predicted)	Zero recovery (predicted)	Accuracy
The Fifth model	In-Sample	Positive recovery	Bayesian	323	17	95.0%
			Distance	321	19	94.4%
		Zero recovery	Bayesian	10	150	93.8%
			Distance	10	150	93.8%
		Total accuracy		Bayesian		94.6%
				Distance		94.2%
				Positive recovery (predicted)	Zero recovery (predicted)	Accuracy
	Out-of-Sample	Positive recovery	Bayesian	81	4	95.3%
			Distance	77	8	90.6%
		Zero recovery	Bayesian	9	31	77.5%
			Distance	8	32	80.0%
		Total accuracy	Bayesian			89.6%
			Distance			87.2%

Based on the prediction results, we can conclude the following.

(1) The in-sample accuracy and out-of-sample accuracy are both fairly high regardless of the discrimination method employed.

(2) Bayesian discrimination has a higher total accuracy than Distance discrimination. It is related to the sample structure and the characters of the two discriminations. The accuracy of Bayesian discrimination is more likely to be affected by the prior probability than the Distance discrimination. Provided we have chosen the right prior probability, Bayesian method would have better accuracy.

5. Predicting Zero Recovery for Obligors with Several Loans

As mentioned in Section 2, obligors are divided into two types: obligors having one loan or obligors with several loans. In this section, we extend the models in the above section to classify the zero recovery and positive recovery for the obligors with several loans. The main idea is described as follows.

First, assume that whether one loan is zero recovery or not is independent from other loans of the obligor. To be specific, assuming one obligor has n loans: $A_1, A_2, \ldots, A_n$, we can determine whether the obligor's recovery is zero or not by the information of this customer and all his loans and by the linear discriminating function f1 and f2 we derive from the model. Assuming that the recovery state of the ith loan is I_i, which is a discrete variable with value 1 if the loan is positive recovery and value 0 otherwise, then, the recovery state of the obligor's is $I = \sum_{i=1}^{n} I_i$. If $I > 0$, then this obligor has positive recovery and zero otherwise.

Since the in-sample and out-of-sample prediction accuracy of Bayesian discrimination is better, we select Bayesian method to predict the classification. The result is displayed in Table 9 .

Table 9: The prediction results of obligors having several loans.

	Be judged to be positive recovery	Be judged to be zero recovery	Accuracy
Positive recovery	533	59	90.1%
Zero recovery	18	211	92.0%
Total accuracy			90.6%

As is shown in Table 9 , the total accuracy is quite high. There is no big difference for the accuracy of single loan case, which proves the efficiency of the single loan model. Moreover, the result indicates that there exists significant difference in the obligor's economic condition and the quality of the loans between the obligors with positive recovery and those with zero recovery, no matter the obligors borrow one loan or several loans. So, it is essential for commercial banks and asset management companies to pay close attention to obligors instead of loans when defaults happen.

6. Conclusion and Future Work

This paper uses Stepwise discrimination to select important factors which determine whether an NPL has recovery or not. Combined with statistical analysis, we give an in-depth analysis of the significant factors on the basis of discriminant function and then we employ Distance discrimination and Bayesian discrimination to develop several models. We test the models' forecasting efficiency, and the results show that these models can reach a high accuracy in prediction.

We find that there are significant differences between the NPLs with recovery and the NPLs without recovery, which are shown in business situation, five-category classifications for loans, transfer mechanism of loan, collateral, industry, and region, respectively. The business situation of obligors and the quality of loans play a major role in classification.

Since it is difficult to access to financial data, some important information such as value of the collateral as well as company's financial statement can not be included in our models. This will no doubt affect the prediction accuracy of models. As the data in LossMetrics are collected from AMCs, the financial data were intentionally deleted by the commercial banks when they transferred NPLs to AMCs with abusing excuse of business secret. Hence it is believed that the results are possible to be improved provided that the commercial banks' data are used for modeling.

It should be pointed out that the structure and characteristics of NPLs will gradually change along with the economic development and time. For example, the policy-off NPLs occurred under a given history circumstance, might not take place again. Thus the changes make the studies on NPLs' recovery model an everlasting job.

Acknowledgments

The authors appreciate the anonymous referee's detailed and valuable suggestions on presentation of this paper. The work is supported by the National Science Foundation of China (no. 70425004) and Quantative Analysis and Computation in Financial Risk Control (no. 2007 CB 814900).

References

[1] T. Schuermann, "What do we know about Loss Given Default?" in *Credit Risk: Models and Management*, Working Paper, Risk, New York, NY, USA, 2nd edition, 2003.
[2] E. I. Altman and V. M. Kishore, "Almost everything you wanted to know about recoveries on defaulted bonds," *Financial Analysts Journal*, vol. 52, no. 6, pp. 57–64, 1996.
[3] V. V. Acharya, S. T. Bharath, and A. Srinivasan, "Understanding the recovery rates on defaulted securities," CEPR Discussion Papers 4098, Centre for Economic Policy Research, Washington, DC, USA, 2003.
[4] E. I. Altman, A. Resti, and A. Sironi, "Analyzing and explaining default recovery rates," ISDA Research Report, The International Swaps and Derivatives Association, London, UK, December 2001.
[5] L. V. Carty and D. Lieberman, "Defaulted bank loan recoveries," Moody's Special Report, Moody's Investors Service, New York, NY, USA, November 1996.
[6] G. M. Gupton, D. Gates, and L. V. Carty, "Bank-loan loss given default," Moody's Special Comment, Moody's Investors Service, New York, NY, USA, November 2000.
[7] L. Hurt and A. Felsovalyi, "Measuring loss on Latin American defaulted bank loans: a 27-year study of 27 countries," *The Journal of Lending & Credit Risk Management*, vol. 81, no. 2, pp. 41–46, 1998.
[8] G. M. Gupton, "Advancing loss given default prediction models: how the quiet have quickened," *Economic Notes*, vol. 34, no. 2, pp. 185–230, 2005.
[9] Z. Y. Chen, "Studies of loss given default," *Studies of International Finance*, no. 5, pp. 49–57, 2004 (Chinese).
[10] J. Wu, "Models of loss given default in IRB-research on core technology of New Capital Accord," *Studies of International Finance*, no. 2, pp. 15–22, 2005 (Chinese).
[11] H. F. Liu and X. G. Yang, "Estimation of loss given default: experience and inspiration from developed countries," *Journal of Financial Research*, no. 6, pp. 23–27, 2003 (Chinese).
[12] P. L. Shen, "The measurement of LCD in IRB approach," *Journal of Finance*, no. 12, pp. 86–95, 2005.
[13] X. K. Ye and H. L. Liu, "Research on structure of loss given default of NPLs in commercial bank," *Contemporary Finance*, no. 6, pp. 12–15, 2006 (Chinese).
[14] L. Bruch, *Discriminant Analysis*, Hafner Press, New York, NY, USA, 1975.

Hindawi Publishing Corporation
Journal of Applied Mathematics and Decision Sciences
Volume 2009, Article ID 868215, 14 pages
doi:10.1155/2009/868215

Research Article

Cumulative Gains Model Quality Metric

Thomas Brandenburger and Alfred Furth

Department of Mathematics and Statistics, South Dakota State University, Box 2220, Brookings, SD 57007, USA

Correspondence should be addressed to
Thomas Brandenburger, thomas.brandenburger@sdstate.edu

Received 1 December 2008; Accepted 13 April 2009

Recommended by Lean Yu

This paper proposes a more comprehensive look at the ideas of KS and Area Under the Curve (AUC) of a cumulative gains chart to develop a model quality statistic which can be used agnostically to evaluate the quality of a wide range of models in a standardized fashion. It can be either used holistically on the entire range of the model or at a given decision threshold of the model. Further it can be extended into the model learning process.

1. Introduction

In developing risk models, developers employ a number of graphical and numerical tools to evaluate the quality of candidate models. These traditionally involve numerous measures including the KS statistic or one of many Area Under the Curve (AUC) methodologies on ROC and cumulative Gains charts. Typical employment of these methodologies involves one of two scenarios. The first is as a tool to evaluate one or more models and ascertain the effectiveness of that model. Second however is the inclusion of such a metric in the model building process itself such as the way Ferri et al. [1] proposed to use Area Under the ROC curve in the splitting criterion of a decision tree.

However, these methods fail to address situations involving competing models where one model is not strictly above the other. Nor do they address differing values of end points as the magnitudes of these typical measures may vary depending on target definition making standardization difficult. Some of these problems are starting to be addressed. Marcade [2] Chief Technology officer of the software vendor KXEN gives an overview of several metric techniques and proposes a new solution to the problem in data mining techniques. Their software uses two statistics called KI and KR. We will examine the shortfalls he addresses more thoroughly and propose a new metric which can be used as an improvement to the KI and KR statistics. Although useful in a machine learning sense of developing a model,

these same issues and solutions apply to evaluating a single model's performance as related by Siddiqi [3, chapter 6] and Mays [4, section 2] with respect to risk scorecards. We will not specifically give examples of each application of the new statistics but rather make the claim that it is useful in most situations where an AUC or model separation statistic (such as KS) is used.

2. Problems with Current Methods

2.1. Overview

As previously mentioned, the references indicate that there are many ways to assess the way in which a model classifies the outcome. Mays [4, chapter 6] separates these into a few general categories. We are concerned with two of these.

(i) Separation statistics. Within this specifically we are concerned with the KS statistic. Its advantages include that it is fairly easy to understand. In the context in which we use Kolmogorov-Smirnov (KS) statistic, it is defined as the maximum separation (deviation) between the cumulative distributions of "goods" and "bads" as both Mays [4, chapter 6] and Siddiqi [3, page 123] outline it.

(ii) Ranking statistics. Siddiqi [3] outlines the C statistic or area under the ROC curve along with the Lorenz Curve, Gini index, Lift Curve, and Gains Chart. These definitions vary somewhat from source to source in industry literature. We will concern ourselves with the AUC of a Cumulative Gains chart similar to that used for the KI statistic by Marcade [2].

2.2. Details of the Cumulative Gains Chart

To illustrate some of the flaws of KS and AUC statistics, let us use two graphical examples. The figures represent example models built from actual test data of a random mailing of potential credit card recipients in the sub-prime credit market. The sample includes approximately 1400 cardholders who responded to credit offers. Models were built using logistic regression with several covariates. Figures 1 and 2 are charts from 2 candidate models.

The construction of the chart is as follows.

(1) Create a logistic model. It does not need to be a logistic model but the ability to define a level of the dependent or target variable as a successful prediction is necessary. In the case of risk the target is often a "bad" account since bad accounts have the greatest financial cost. Whereas if you were doing a model for a response to a mailing marketing campaign, a response would be "good" and that would be your target. To simplify for this example, the value of the dependent variable = 1 in this risk model will constitute a "bad" account risk. A value of the target dependent variable = 0 will be "good".

(2) Score the data set on the model and rank them in order from highest to lowest probability of being bad (target = 1).

(3) Go through the ranked data set (highest to lowest probability) in a loop counting the cumulative number of actual points in the data set which were bad (value = 1) and good (value = 0).

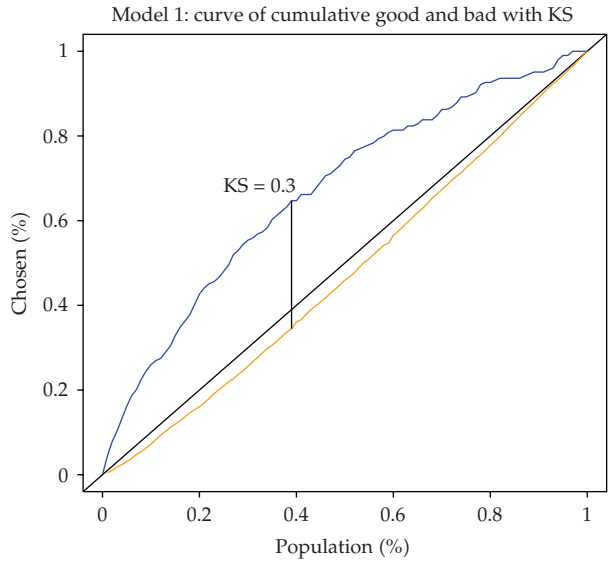


Figure 1: Model 1 Gains Chart

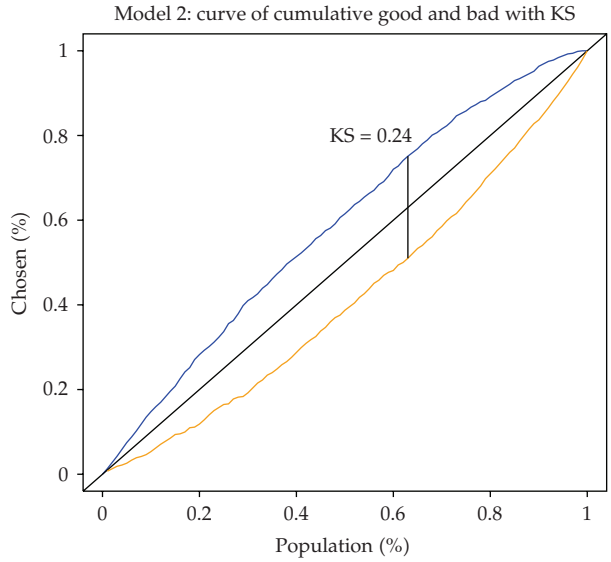


Figure 2: Model 2 Gains Chart

(4) Plot these 2 sets of values in 2 curves as a percentage of the proportion of the bad and good populations, respectively. In our risk model example the percent of the population in Model 1 which are "bad" is approximately 15 percent. In Model 2 the definition of the risk target variable is different so even though it is the same data set the bad rate is approximately 50 percent for the entire sample. So if there are 1400 data points and the proportion of bad in Model 1 is .15 then the data point on the graph corresponding to (.4,.63) would mean that by successfully picking the riskiest accounts (calculated by the model) in decreasing order it

would take 40 percent or the accounts (or 560) to find 63 percent of the bad accounts in the sample (or approximately 132 out of the 210 bad accounts in the 1400 sample size).

(5) Plot a 45-degree line. The meaning of this line is often glossed over quickly in literature and often misunderstood by analysts; however, it is very key to the development of descriptive statistics for model quality so we will detail its meaning. This line represents if you were truly random guessing. Imagine a bag full of some green balls and some red balls. A random ball is successively picked at random out of the hat without replacement. Its value is cataloged as red or green. If you pick zero balls out of the bag you would get zero red. Picking all the balls out of the hat would result in having all the red balls accounted for. Picking half of the balls out of the bag should on average net you half of the red balls which were in the bag regardless of the proportion of red balls. Hence the vertical axis is "as a percentage of the bad accounts". Choosing randomly you should get on average half the bad accounts with half the population chosen regardless of the proportion of bad accounts.

(6) Calculate KS by taking the maximum difference between these good and bad curves.

2.3. Shortcomings

When observing the cumulative gains charts in Figures 1 and 2, it is intuitive that the greater the separation between the good and bad curves, the better the model. If one model has separation across the entire ranking greater than another model with the same definition of the target variable, then as pointed out by Piatetsky-Shapiro and Steingold [5, section 1] for any cost versus benefit matrix the dominate model wins. However in practical models you may not get the case where one model is strictly dominating over the other. Consider Model 2. It is easy to imagine another model which has the same KS but the distribution of the curve curve is such that the KS occurs significantly further to the left and the curves may cross. In fact they may cross more than once. Piatetsky-Shapiro and Steingold [5, section 1], Mays [4, page 110], Siddiqi [3, page123], and Marcade [2] all point out that practically speaking a model has a functional decision cutoff, and the KS may fall outside the range of risk values used to extend credit therefore making its usefulness suspected. Further they indicate that the performance of the model in the intended range of use is highly important.

The reality of the intended practical use of the model cutoff is also important in the calculation of any AUC type statistic. Notice that in Figure 1 not only is the KS greater but it occurs further to the left. This creates a bulbous shape in the curve to the left side of the graph. When comparing 2 models it is often more important what happens in the lower deciles. For instance in a mailing campaign for marketing the decision may be to only mail the 20 percent of the population which is most likely to respond. Consider 2 models which happen to have the same KS and same AUC under the bad curve. The model which is "shifted left" with its KS occurring earlier and AUC weighted to the left will have superior financial performance since it is identifying a greater percentage of the target variable with a lower percentage of population. This concept is widely known and is easy to see visually, however there is no good quantification of this. We propose that our model statistic will account for this. A more difficult fact to comprehend from the graph is that of the proportion of bad in the population and its effect on KS and AUC. Consider 2 modeling situations. The first is the one presented here in the difference between Model 1 and Model 2. Model 1 has a much lower proportion of bad in the population as compared to Model 2 (15 percent versus 50 percent) even though the data set is the same. This is due to redefining what constitutes a bad account.

Mays [4, page 122] mentions different bad definitions but does not address its effect on the statistical metrics. The situation here is that the data is entirely the same for both models with the exception of the definiton of bad. Depending on the goal of the model the definition of bad may be different. For instance in a credit situation using the same data set one may predict the probability of an account being "charged off" as bad debt within 3 years of extending a credit offer. However for purposes of expediency (rather than waiting for 3 years for the test data to mature in performance) a surrogate endpoint may be used. As an example an endpoint of "60 days or more in arrears at 12 months" may be used instead. The two different definitions of bad will not have the same "bad rate". Nor will you be able to have the same ability to predict the indended endpoint of bad debt at 3 years if using the earlier estimate of risk. Inherently 2 models built on these 2 different endpoints would have different values of KS. The second case is of different populations. In prime versus subprime you have entirely different risk levels as a population. Mays [4, chapter 7] addresses an entire chapter on KS indicating that the KS for subprime loans may be different than prime loans but misses the opportunity to explain that much of the difference is arithmetic. The author asserts that KS statistics for subprime loans will be lower because they have a much higher bad rate and concludes that the reason for the lower KS is the inability of models to differentiate between good and bad in higher risk credit markets. As we will demonstrate this is not necessarily the case. The difference in KS between 2 models with largely different KS can be at least partially explained by a rather obvious arithmetic property.

3. Improving KS and AUC

3.1. The Perfect Model

Noting that KS can be misleading Siddiqi [3, page 123] suggusts one way to view the separation between goods and bads is to plot the separation over the whole range of the model at specified increments. This technique gives a better view of the distribution of

Separation as can be seen in Figures 3 and 4.

Notice that just like the gains chart even though there are differences in magnitude the general shape is to increase to a peak somewhere in the middle of the distribution then decrease. Piatetsky-Shapiro and Steingold [5] begin the discussion of creating a statistic which can be used to account for the irregularities we have already noted. They use the concept of an ideal or optimal model to compare the given model through "lift quality". The commercial software KXEN uses this concept heavily in the development of the statistics called KI and KR. We will concentrate on the explanation of KI as it is explained in their white paper. Consider Figures 1 and 2 redrawn in Figures 5 and 6, this time including a diagonal line with slope = 1/badrate.

As outlined in both Piatetsky-Shapiro and Steingold [5] and Marcade [2], the upper line represents the ideal model or perfect. If you rank the data set from highest to lowest probability of having target value = 1 then the best you could do is be right every time. In the analogy of the bag of red and green balls this line would be equivalent to drawing a red ball every time until the red balls were all gone. So in the case of Model 1, 15 percent of the population is bad and perfect model performance would be to choose 100 percent of these bad accounts in the first 15 percent of the population by picking them all correctly. It is by drawing this line that you then see that the model KS is constrained by the proportion of bad and good accounts in the data set. Again that proportion is determined by two things, the

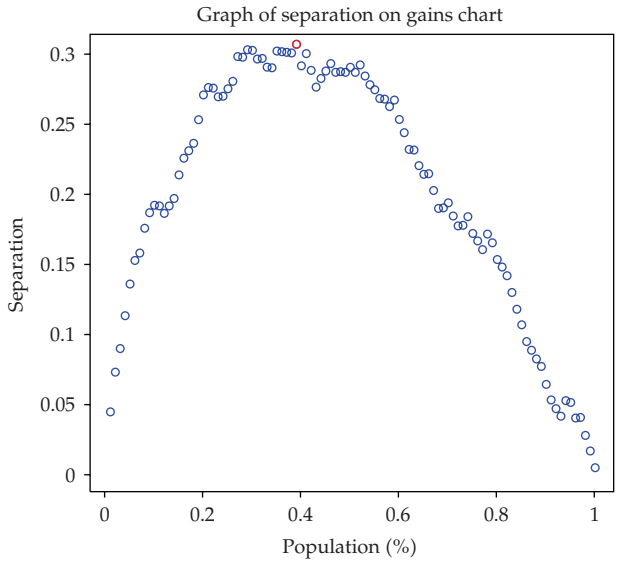


Figure 3: Model 1 Separation

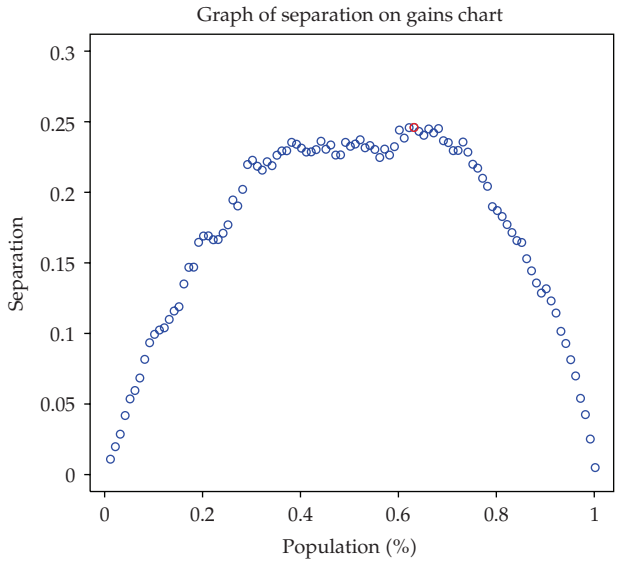


Figure 4: Model 2 Separation

definition of bad and the general risk level of the population. It can be immediately seen that the assertion by Mays [4] is attributing a lower KS to a higher rate of bad due to a model not being which is not as good at rank ordering the bad accounts starts to fall apart. You can also see why the plots of KS in Figures 3 and 4 must go up then come down in the fashion they do. What is less obvious at first is that this also generally means that the KS for the model will be to the right of the corner in the perfect graph. Note in both Model 1 and Model 2 this is true. However, depending on the application, the model's performance may be most crucial at the extremes rather than at the middle due to financial performance considerations.

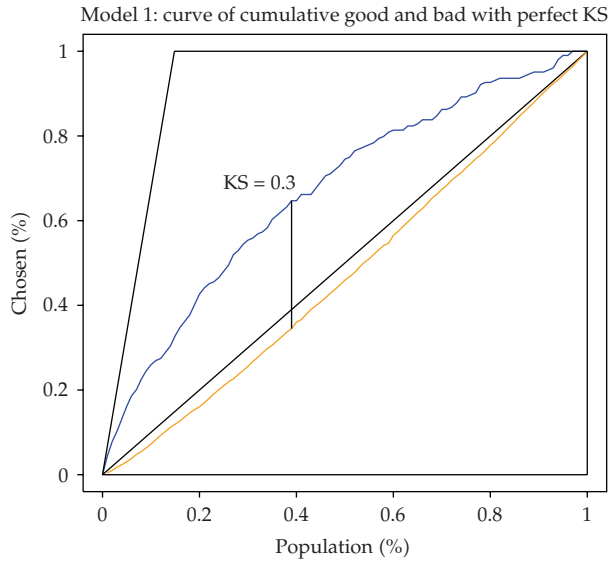


Figure 5: Model 1 Gains with Perfect Line

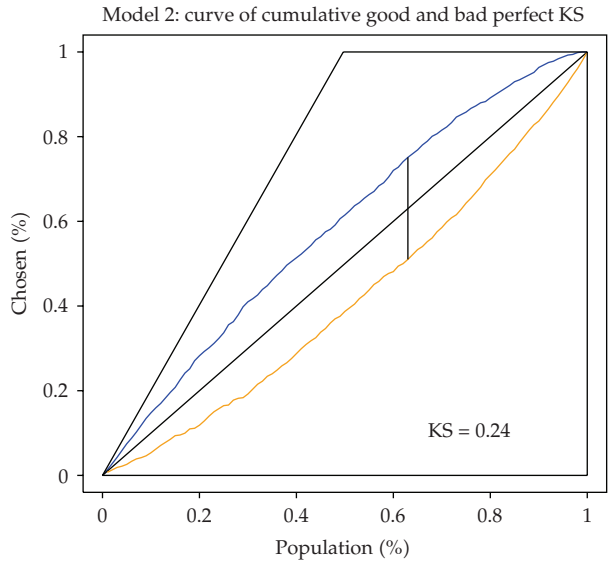


Figure 6: Model 2 Gains with Perfect Line

3.2. KI and KR

KXEN's KI statistic then uses the concept laid out by Piatetsky-Shapiro and Steingold [5] to create two statistics. We have used KXEN's analytics tool and find it to be an outstanding piece of software with powerful results which are easy to interpret. The KI statistic falls between 0 and 1, gives a value of 1 for a perfect model, and gives 0 for a completely random model. This gives it an intuitive feel for a good model metric as Marcade [2] suggests it should. KI is calculated as a "percent of perfect". The target (in this case bad) curve is plotted.

Let $p(x)$ be a function describing the distance between the line of the perfect model and the random diagonal line. Let $a(x)$ be a function describing distance between the curve of the model target and the random line

$$\text{KI} = \frac{\int_0^1 a(x)dx}{\int_0^1 p(x)dx}. \tag{3.1}$$

KR not covered in the reference is a natural extention of KI. It is a measure used on validation data to determine stability of the model. It is essentially the KI of the validation data divided by KI of the model data. The closer this ratio is to 1, the more stable the model. KI is an attempt to offer some measure of "percent of perfect" for the model gains versus a perfect model. It however leaves out a few key ideas.

(i) There is no consideration of the shape of the curve for models whose separation may be shifted left. A model which has a relatively larger separation earlier in the ranking is favorable. KI has no way to reward this.

(ii) It does not incorporate the full consideration of separation between both curves but rather considers only the target. This idea will be more fully explored in our statistic.

(iii) It is only useful as a full model metric. It cannot be used at a cutoff decision point to evaluate a model. That is, KI(x) is not defined.

3.3. Redefining KI

Looking at the definition of KI, 2 of the 3 listed drawbacks of the statistic could be mitigated by a small but significant shift in definition. Let us redefine KI by

$$\text{KI}^* = \int_0^1 \frac{a(x)}{p(x)}dx. \tag{3.2}$$

This change is small but crucial. It does 2 things. First it allows us to define a function which is a point estimate of model quality. We will call it $q(x)$. Second it allows us to create a metric for a range of the model risk rather than just a point or the whole model. Remember that KS was a point estimate, and AUC statistics are typically full model statistics. By taking the integral over a region of the model that we are practically interested in we can achieve a statistic which answers the need,

$$q(x) = \frac{a(x)}{p(x)}, \tag{3.3}$$

$$\text{KI}^* = \int_a^b \frac{a(x)}{p(x)}dx. \tag{3.4}$$

By taking (a, b) to be (0,1) you obtain a statistic on the whole model. It becomes immediately obvious that $q(x)$ can be used as a point estimate of model separation where q is a ratio between actual separation and perfect separation from random. This change by itself would

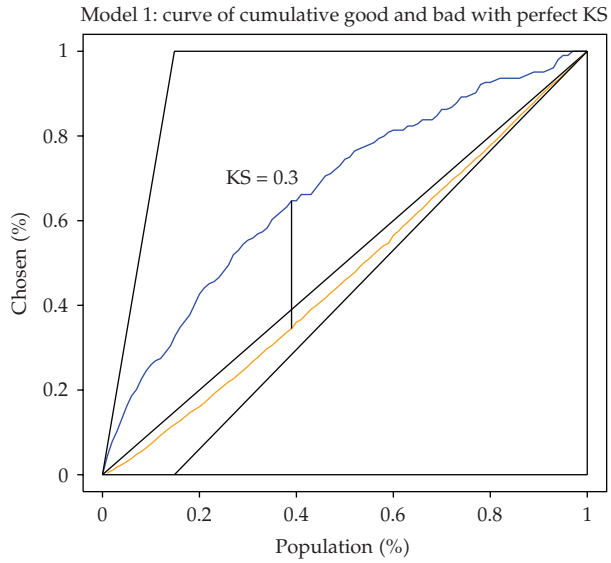


Figure 7: Model 1 Gains with Good and Bad Perfect Lines

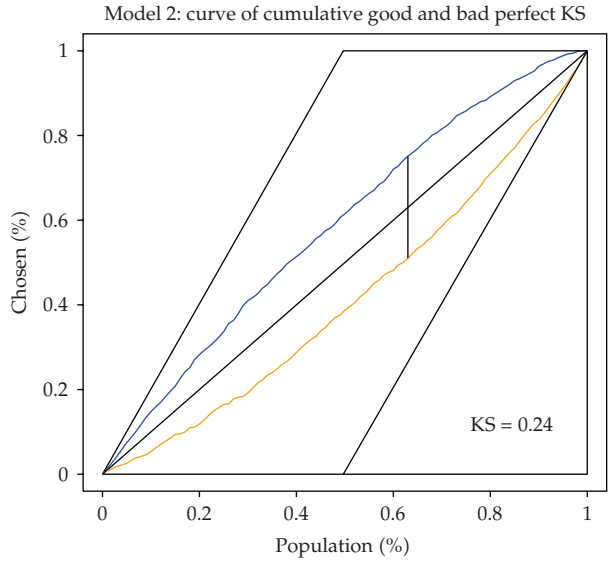


Figure 8: Model 2 Gains with Good and Bad Perfect Lines

be sufficient to be a useful improvement to KI. However there is still one more item not addressed by this metric. That is the idea of separation between goods and bads. To understand the necessity of this portion of the problem, let us once again redraw our two models . In Figures 7 and 8, we insert another diagonal line with slope $= 1/(goodrate)$.

Immediately, another visual difference between the two models is explained. Notice that in Model 1 the good curve follows closely under the diagonal line while in Model 2 there appears to be a large separation. Once the second diagonal line is drawn it becomes

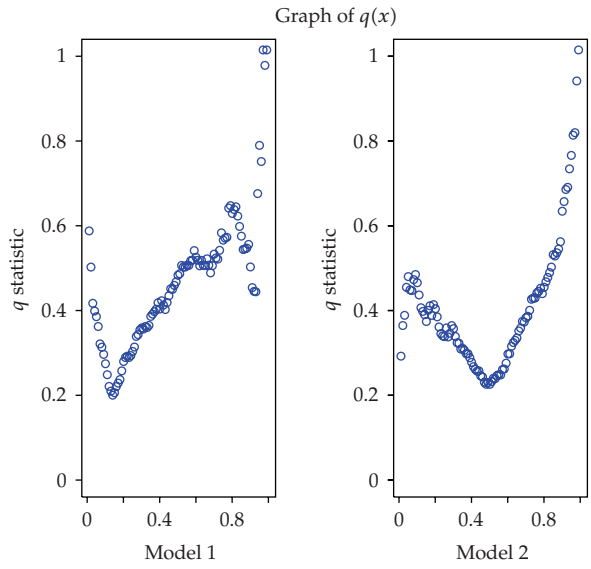


Figure 9: $q(x)$ for Both Models

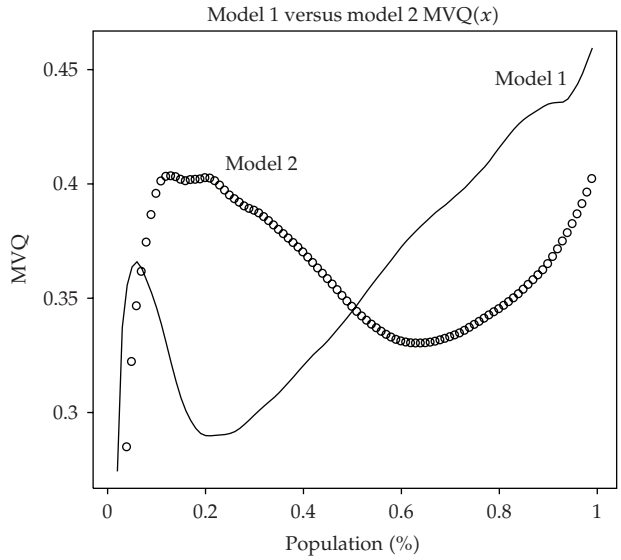


Figure 10: MVQ for Both Models

obvious that the good curve is being constrained by the lower diagonal in the same way the bad curve is constrained by the upper diagonal. What is the meaning of this diagonal? Consider the problem of the red and green balls in the bag in our analogy. The upper perfect curve represents drawing all the red balls out of the bag without error. Once you have drawn them all, the curve must stop at 1 and continue on horizontally as there are no more red balls. Consider the curve of green balls in this circumstance. It would remain at zero until all the red balls had been exhausted. At that point every subsequent ball drawn would be green until you reached 100 percent of the green balls which would naturally be obtained only after the

last ball was drawn. Thus the progression of the good curve for a perfect model would follow the x axis until reaching $x = bad\ rate$ then rising in a straight diagonal to the point $(1, 1)$ at a slope of $1/(1 - badrate)$ or $1/(goodrate)$.

3.4. Cumulative Gains Model Quality

As can be seen in Figures 7 and 8. KS = 1 for a perfect model is always obtained at $x = bad\ rate$ and everywhere else separation of good and bad must be less than this. Extending the logic used for KI, it then becomes a natural extension to consider the area between the perfect good and bad curves as being the ideal curve. Dividing the area between the actual model good and bad curves by the area between the perfect good and bad curves would seem to be a good extension of KI. However this leads to the same problems discussed with KI. This was solved by altering KI in such a way to offer a function which described a point estimate of the ratio between actual and perfect at any point. In order to do this you simply need to divide the separation at that point which we will call $\mathrm{ks}(x)$ by the perfect separation $p(x)$. Intuitively $q(x)$ becomes a normalized separation statistic always between 0 and 1:

$$q(x) = \frac{\mathrm{ks}(x)}{p(x)}. \tag{3.5}$$

In itself it becomes a good second look at separation to supplant KS. However since we are also interested in the performance of the entire model or the performance of the model over a specified range of risk, integrating this function over the whole range of the model becomes a model statistic improvement over KI and integrating over a specific range of x values becomes an improvement to KI*. We therefore define Q to be

$$Q = \int_a^b \frac{\mathrm{ks}(x)}{p(x)} dx = \int_a^b q(x) dx, \tag{3.6}$$

where (a, b) is a region of risk in which we are interested. If chosen to be (0,1), Q becomes a statistic on the whole model.

There is one last thing to consider. One reason KXEN's KI statistic is so successful is that it is easy to interpret. It is always between 0 and 1. This makes it an excellent statistic when interpreting results. As we have shown in this paper one of the primary problems with KS is that it does not land within any consistent range of values. By not being a standardized statistic, it loses ability to compare one model to another. We have specifically chosen two models which have differing KS values due to different definitions of good and bad. We could have just as easily chosen two models with different risk populations and the same bad definition yielding different bad rates. In both cases the challenge is to strive for a statistic which can compare the relative effectiveness of the models.

In this case we have achieved our goal. The distribution of q is analogous to a normalized KS point estimate of model effectiveness, and if you choose (a, b) to be (0,1) on the whole model you achieve a good metric on the whole model. Choosing (a, b) other than (0,1) also allows us to compare two models easily. However what it loses is the guarantee that Q will fall between 0 and 1 and thus loses some of what we were striving for. To fix this

consider the Mean Value Theorem from elementary calculus. By applying this theorem to our problem we can thus guarantee that all stated goals can be achieved by any segment of the model. This statistic is given as

$$\mathrm{MV}Q = \frac{Q}{b-a}. \tag{3.7}$$

4. Discussion

As we have mentioned MVQ becomes a rather useful tool to assess the quality of a model. We plot q in Figure 9, which again is the same as plotting the KS separation at each point but dividing by the separation of a perfect model.

Notice how you can now get a feel for the separation with respect to what was possible. It also levels the playing field in such a way as to indicate that maybe Model 2 actually outperforms Model 1 in certain regions of the model. Specifically in the lower deciles. As already noted there are modeling situations when this performance becomes key. We could then calculate MVQ for the entire model, which turns out to be $\mathrm{MV}Q_1 = 0.46$ and $\mathrm{MV}Q_2 = 0.40$. For illustration to understand the performance of the model over regions of the model, we plot MV$Q(x)$ taken to be MVQ on interval $(0, x)$, In Figure 10. Note that this plot accentuates the performance advantage of Model 2 in the first half of the population.

4.1. Model Stability Metric

The KXEN statistic KR is an extention of KI as already noted. By similar extension of MVQ we can similarly create a model stability metric. When creating models it is common practice to separate a data set into: data the model is built on, and data the model is validated on. One of the easiest and most elegant ways to test for model stability and overfit is to compare the lift statistics of the model build data versus the lift statistics of the validation data. In this case we want to compare MVQ(build) versus MVQ(validation). The closer the ratio of these two is to 1, the more stable the model. We define the Model Stability Metric (MSM) to be

$$\mathrm{MSM} = \frac{\mathrm{MV}Q(\mathrm{validationdata})}{\mathrm{MV}Q(\mathrm{builddata})}. \tag{4.1}$$

4.2. Applicability

We feel that this is a valuable new tool in model performance metrics. As Marcade [2] notes this type of metric is not only useful as a visual or manual interpretation but can be used in a number of different algorithmic ways in a similar fashion to entropy for machine learning. As an example Ferri et al. [1] pointed out you could use area under the ROC curve to not only assess the quality of a decision tree but in the creation of the splits during the algorithm. By extension MVQ can be used in a similar fashion. MVQ and q offer several key features of a model quality statistics including what follows.

(i) $q(x)$ can be used as a point estimate of separation of model curves which is superior to KS in that it is normalized between 0 and 1. This gives it an inherent advantage to understand model performance in that it is very intuitive. The normalization between 0 and 1 also proves superior in different risk populations of financial modeling where the underlying risk of the population may differ substantially, we have demonstrated that KS would necessarily differ from among scenarios whereas $q(x)$ would be a pure "percent of perfect". Results can more consistently be interpretted.

(ii) MVQ can be used as a whole model lift statistic by choosing the interval (a, b) to be (0,1). It can also be used to measure the useful range of the model where financial decisions will take place on any interval (a, b) without loss of generality. As an example a risk model may have pro forma financial criteria. The cut point in the population where credit will be extended is based on this. We may only be interested in comparing two models below this cut point. Above this performance cut point the two models may actually cross in performance but that is of no concern as we have predetermined that both scenarios are unprofitable above a certain risk.

(iii) MVQ and q both give weight to higher performance early in the population selection ranking, whereas other measures such as KS have no ability to discern a "percent of perfect" rating. This allows for greater weight on the left for higher performing models. This is a very key achievement for practitioners of risk management. In the case of 2 models with similar KS the model where the separation is further to the left will inherently be given a higher value of q and MVQ.

(iv) It is agnostic in its assumptions of the model. Credit risk management does not use traditional measures of model goodness such as P-value. Instead various lift metrics are often used. MVQ is free from any assumptions regarding the model or population itself and is only concerned with how accurately we successfully predicted the endpoint. In the case of a simple logistic regression, the choice of 0 and 1 and the predicted probability is obvious. Many other types of models easily lend themselves to this kind of interpretation however. In a Markov model, often used for risk management, one only has to determine which states of the Markov model are a success and which are failure creating a 0 and 1 outcome. Similarly financial models may have a continuous dependent variable such as profit. Decision criteria of the business will yield a threshold of profit which is acceptable and not acceptable. This leads to a 0 versus 1 endpoint. In practice these financial decisions are regularly made and it is only up to the model to predict which side of the threshold a particular credit account falls. It does not matter whether logistic regression, multinomial regression, Markov models, clustering, neural nets, or trees (all important models in risk management) are used. All may be condensed into whether the predicted outcome meets financial criteria or not and will submit to the 0 versus 1 test. So in fact this allows you to compare the use of say a neural net on a data set versus a logistic model versus a regression tree and have a meaningful way to compare the results.

The ability to better understand the implications of risk in the financial world is of utmost importance. It is precisely the lack of this ability which has been implicated in the current financial crisis. Practitioners are often left however with a confusing myriad of models which

do not lend themselves to traditional measures of model quality taught in classic statistics classes. Ultimately the goodness of a model can only be determined by the proportion of the time it is correct in predicting the outcome. Current methods of doing this such as KS are not complete or intuitive and do not have consistent normalized measure from model to model. MVQ answers these issues. At a practical level, the authors feel that q and MVQ are especially useful in interpreting results at the decision level of risk management where nonstatisticians become intimately involved.

References

[1] C. Ferri, P. A. Flach, and J. Hernández-Orallo, "Learning decision trees using the area under the ROC curve," in *Proceedings of the 19th International Conference on Machine Learning (ICML '02)*, pp. 139–146, Sydney, Australia, July 2002.

[2] E. Marcade, "Evaluating modeling techniques," Technical Whitepaper, KXEN, Inc., San Francisco, Calif, USA, 2003.

[3] N. Siddiqi, *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*, John Wiley & Sons, Hoboken, NJ, USA, 2006.

[4] E. Mays, *Credit Scoring for Risk Managers: The Handbook for Lenders*, South-Western, Mason, Ohio, USA, 2004.

[5] G. Piatetsky-Shapiro and S. Steingold, "Measuring lift quality in database marketing," *ACM SIGKDD Explorations Newsletter*, vol. 2, no. 2, pp. 76–80, 2000.