2018

Option pricing with short selling restrictions or bans being imposed

Guiyuan Ma
University of Wollongong

Option pricing with short selling restrictions or bans being imposed

A thesis submitted in fulfilment of the requirements for the award of the degree of

Doctor of Philosophy

from

University of Wollongong

by

Guiyuan Ma,
B.Sc (Jilin University)
GradCert (Fudan University)

School of Mathematics and Applied Statistics

Certification

I, *Guiyuan Ma*, declare that this thesis, submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Mathematics and Applied Statistics, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

Guiyuan Ma

May 7, 2018

To the Justice League

Acknowledgements

First of all, I would like to owe my deepest gratitude to my supervisor, Prof. Song-Ping Zhu, for his valuable guidance, cheerful enthusiasm and friendly encouragement. Without his continuous support during my research, it would be very hard to complete this thesis. I, in particular, appreciate his great contributions of time and ideas to enrich my Ph.D career so much. His great passion in science and rigorous attitude in research impress me deeply and I would abide them as my principle in my future work.

My appreciation goes, in particular, to Dr. Wen-Ting Chen for her great help at the first year of my research, which introduces me into such a wonderful research field. Another great appreciation is given to Dr. Ivan Guo as I benefit so much from the discussions with him about our project. I would also express my thanks to Dr. Boda Kang from University of York. During his visit to University of Wollongong, we have a great discussion and a wonderful cooperation. I wish to thank all the fellow friends in the Center of Financial Mathematics and the School of Mathematics and Applied Statistics in University of Wollongong, particular Dr. Kai Du and Dr. Xiao-Ping Lu for her encouragement in research and life and my dear friends Xiang-Chen Zeng, Xin-Jiang He, Zi-Wei Ke, Sha Lin, Dong Yan, Chengbo Yang and Peng Li for their help in my PhD career. I trust that all the others whom I have not specifically mentioned here are ware of my deep appreciation.

Finally, the financial support from the University of Wollongong with a Discovery Project University Postgraduate Award and an International Postgraduate Tuition Award (IPTA) is also gratefully acknowledged. I also like to give my thanks to the financial support from the Australian Research Council (ARC) under the ARC(DP) funding scheme (DP140102076), which funded our research project.

Abstract

The Financial Crisis 2007-2009 is considered as the worst one since the Great Depression of the 1930s. During this financial crisis, most regulatory authorities around the world imposed restrictions or bans on short selling to reduce the volatility of financial market and limit the negative impacts of a downturn market (Beber & Pagano 2013). Such interventions were implemented to restore the orderly functioning of financial markets and limit drops in stock price. However, these regulations also resulted in some new problems, one of which is how to price options in a market with short selling restrictions or bans being imposed. The motivation of this Ph.D thesis is to study the effects of short selling restrictions or bans on option pricing.

The thesis is divided into two parts, the first one of which is consist of Chapter 3 and Chapter 4, where option pricing is explored under a new hard-to-borrow stock model. Such a model was proposed by Avellaneda & Lipkin (2009) to characterize the price-evolution of stocks subject to short selling restrictions. Although an approximate semi-explicit pricing formula has been produced for European call option, its derivation requires an independence assumption, which has limited its application to more general cases. In Chapter 3, we propose a new partial differential equation (PDE) approach to price European call options under the hard-to-borrow stock model and then an alternative direction implicit (ADI) scheme is applied to solve it numerically. This new PDE approach has also laid a solid foundation for the study on option pricing of American-style options. In Chapter 4, we extend the PDE approach to the American case and reformulate it as a linear complementarity problem, which is numerically solved with the Lagrange multiplier approach. A significantly important conclusion is that it may be optimal to exercise an American call option before expiration even though the underlying stock pays no dividends. Such a conclusion supports the recent work by Jensen & Pedersen (2016) and overturns a classic result by Merton (1973).

The second part of this thesis is about option pricing with short selling bans being imposed. Recently, Guo & Zhu (2017) proposed a new *equal-risk pricing approach* to study the effects

of short selling bans on option pricing. Their analysis method appears to be but not the same as the existing utility indifference pricing methods. Only when the payoff function is monotonic, can an analytical pricing formula be produced. However, it is still difficult to apply equal-risk pricing approach to the case where the payoff function is non-monotonic. We intend to expand its application by establishing a PDE framework. Since Hamilton-Jacobi-Bellman (HJB) equation would be involved, we first explore different solution approaches to the HJB equation in Chapter 5, Chapter 6 and Chapter 7 as preliminaries before taking on the tough challenge of establishing the PDE framework for equal-risk pricing approach.

In Chapter 5, we successfully apply the homotpy analysis method to decompose the highly nonlinear HJB equation into an infinite series of linear PDEs and finally derive an exact and explicit solution for the HJB equation subject to general utility functions for the first time. In Chapter 6, a closed-form analytical solution for the Merton problem defined on a finite horizon with exponential utility function is obtained through two different methods without any one of the following assumptions: (1) the utility function belongs to the constant relative risk aversion (CRRA) class; (2) the utility function is defined over terminal wealth only and consumption is not allowed; (3) the investment horizon is infinite. In Chapter 7, a monotone numerical scheme method is presented to solve the HJB equation with general utility functions. Such a scheme is proved to be convergent through demonstrating its stability, consistency and monotonicity.

After proposing these three solution approaches to the HJB equation, we finally establish a PDE framework for equal-risk pricing approach in Chapter 8 and successfully solve the HJB equation analytically and numerically. When the payoff function is monotonic, analytical pricing formula is derived from our PDE framework, which matches perfectly with the pricing formula derived by Guo & Zhu (2017). When the payoff function is non-monotonic, such as a butterfly spread option, equal-risk price is also produced through solving the PDE system numerically, which is absent in Guo & Zhu (2017). Consequently, our PDE framework has really expanded the range of application of equal-risk pricing approach so that effects of short selling bans are discussed in more general cases.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Short selling and financial crisis

Short selling is the practice of selling financial securities that are not currently held and subsequently repurchasing them to deliver. In financial practice, it is achieved as follows: (a) the seller informs his broker that he wishes to sell a stock that he does not own; (b) the broker arranges for a buyer; (c) the trade takes place; (d) the seller delivers the stock within a stipulated amount of time.

Naked short selling is a kind of short selling that does not require borrowing or arranging to borrow the stock in advance. Through naked short selling, investors could sell any amount of stocks that they do not own without any other cost. Obviously, it perfectly satisfies the assumption in the Black-Scholes model. However, if the stock is in short supply or finding a lender is very difficult in practice, the seller may fail to deliver the stock within the settlement period, which leads to "fail-to-deliver" risk. Therefore, some regulators would impose some constraints on short selling. In this thesis, all the impediments that makes short selling still available, but naked short selling unavailable, are collectively referred to as *restrictions on short selling*. *Covered short selling* is one kind of short selling with restrictions. The short seller is required to borrow or arrange to borrow shares in advance at the time of sale and he needs to pay some borrowing fees when the stock is hard to borrow. Such restrictions make the covered

short selling costly, which does not satisfy the perfect assumption in the Black-Scholes model any more.

In fact, short selling has a long history and it still plays an important role in modern financial markets. The first records of actual short selling date back to 1609 when a group of Dutch businessmen formed a secret association to short the shares in the East India Company in anticipation of the incorporation of a rival French-chartered trading firm. They sold shares that they did not own and promised future delivery in one or two years. Over the next year, the group profited a lot as the stock of East India Company dropped by 12%, which angered the shareholders who inevitably learned of their plan. Laws that prohibited short selling were first enacted in 1610 following such a well coordinated and highly profitable "bear raid" (Bris et al. 2007).

Over the last 400 years, short sellers have always been blamed for stock market declines, leading market participants to call for regulation against short selling. Recently, it again caught the public eyes for its notoriety during the Global Financial Crisis 2007-2009, which is considered as the worst one since the Great Depression of the 1930s. According to the Statement of Richard S. Fuld Jr., the final Chairman and CEO of Lehman Brothers, before the United States House of Representatives Committee on Oversight and Government Reform on October 6, 2008[1], the collapse of the investment bank is partly due to short selling, which allegedly depressed the stock price during this financial crisis.

During the Financial Crisis 2007-2009, most regulators around the world imposed restrictions or bans on short selling to reduce the volatility of financial market and limit the negative impacts of a downturn market. For example, the U.S. Securities and Exchange Commission (SEC) temporarily banned most short selling in nearly 1,000 financial stocks in September 2008[2]. In a coordinated approach with the SEC, the UK Financial Services Authority (FSA), also introduced a temporary and total ban on all short selling (both naked and covered) in the shares of 32 financial sector companies[3]. The Australian Securities and Investment Commission

[1]See website: http://online.wsj.com/public/resources/documents/fuldtestimony20081006.pdf
[2]See website: https://www.sec.gov/news/press/2008/2008-211.htm
[3]See website: http://www.klgates.com/files/upload/FSA_ShortSelling_No2.pdf

(ASIC) announced a package of interim measures that naked short selling is not permitted and covered short selling is still permitted but needs to be disclosed[4]. Other markets also followed and announced their own restrictions or bans on short selling (Beber & Pagano 2013).

These corresponding regulations on short selling were implemented to restore the orderly functioning of financial markets and limit drops in stock price. However, they would also bring in some new problems to the financial markets. For example, option pricing now becomes a difficult problem because short selling is restricted or banned, which leads the market to be incomplete. The classical option pricing model proposed by Black & Scholes (1973) becomes invalid because the important assumption about short selling has been violated either in the case where covered short selling is still permitted or in the case where both naked and covered short selling are banned. The motivation of this thesis is to study how these restrictions or bans imposed on short selling by the regulatory authorities during the financial crisis 2007-2009 affect the option pricing in financial market. Before starting our research, we first present some literature review about the classical option pricing theory in the next section.

1.2 Literature review on option pricing

In finance, an *option* is a contract that entitles its holder the right, but not the obligation, to buy or sell an underlying asset on a specified date (*expiration date*) at a specific price (*strike price*), irrespective of the market price of the underlying on that date. Options can be classified in many ways. An option that gives to its owner the right to buy (sell) at a specific price is referred to as a *call (put) option*. An option that can only be exercised on expiration date is called a *European* one; while an option that may be exercised on any trading day before the expiration date is called an *American* one.

For any kind of option, the first and foremost topic both in academic and in practical finance is pricing problem. The history of option pricing theory dates back to 1900 when the French mathematician Louis Bachelier derived an option pricing formula on the assumption

[4]See website: http://asic.gov.au/about-asic/media-centre/find-a-media-release/2008-releases/08-204-naked-short-selling-not-permitted-and-covered-short-selling-to-be-disclosed

that stock prices follow a Brownian motion with zero drift (Bachelier 1900). The milestone of contemporary option pricing theory is the famous and classical model proposed by Black & Scholes (1973), which provided an analytical and quantitative formula for European options for the first time. In their landmark paper, the market is assumed to be consist of at least one risky asset, usually called the stock, and one riskless asset, usually called the bond or cash. Many assumptions are imposed both on the assets and on the market to derive the elegant pricing formula. Here we list all the assumptions in the following.

Assumption 1. (Assumptions in the Black-Scholes model)

1. The rate of return on the riskless asset is constant and thus called *risk-free interest rate*.
2. The underlying stock price follows a geometric Brownian motion and both its drift and volatility are constant.
3. The underlying stock does not pay a dividend.
4. There is no arbitrage opportunity in the market, i.e. the market is complete.
5. It is possible to borrow and lend any amount of cash at risk-free interest rate.
6. It is possible to buy and sell any amount of stocks. In other words, short selling is permitted without any cost.
7. The above transactions do not incur any fees or costs, i.e. it is a frictionless market.

Under these assumptions, Black & Scholes (1973) managed to replicate a European call option with a portfolio that is constructed by the underlying stocks and risk-free bond. After complicated mathematical calculations, they succeeded in deriving an analytical pricing formula for European call options, which also laid the solid foundation for the development of the modern option pricing theory. Since then, considerable research interests have been drawn to the option pricing problems. Klemkosky & Resnick (1979) furthermore derived an analytical pricing formula for European put options via the put-call parity.

Compared with European options, American options are much more valuable for they can be exercised on any trading day before the expiration date. The possibility of early exercise also makes American options more difficult to price than its European counterpart. The valuation of American options is considered as an optimal stopping time problem because the holder of American options needs to choose an optimal time to exercise the option in order to profit as much as possible. Such an optimal stopping time problem is then reformulated as a variational inequality problem (Merton et al. 1977). Since an unknown *optimal exercise boundary* is also part of the solution, the variational inequality problem becomes a free boundary problem, which is highly nonlinear. Unlike the European case where the well-known Black-Scholes equation is linear and can be solved analytically, the nonlinearity has hindered the search for an analytical pricing formula for American options.

Some numerical methods are first applied to solve the nonlinear free boundary problem. Schwartz (1977) first applied the finite difference method to solve the option pricing problem for American options. Then Cox et al. (1979) proposed a binomial tree method to approximate the valuation of American option, which is actually considered as a special case of the explicit finite difference method. To reformulate the free boundary problem as a fixed boundary one, Wu & Kwok (1997) applied a front-fixing transformation, which was first suggested by Landau (1950), and then developed an efficient finite difference method to produce the optimal exercise boundary and option value at once. Later the radial basis function method and the finite element method were also introduce to solve the valuation of American options by Hon & Mao (1999) and Allegretto et al. (2001), respectively. Moreover, Monte Carlo simulation method was also proposed to price American options. Longstaff & Schwartz (2001) proposed a simple and powerful least-squares approach for approximating the value of American options by simulation.

In addition to these numerical methods, some quasi-analytical solutions are also introduced to solve the pricing problem of American options. Carr & Faguet (1996) proposed a novel approach, referred to as *analytical method of lines*, Huang et al. (1996) presented a new method of recursive implementation of analytical formulae, referred to as *integral equation approach*, Broadie & Detemple (1996) developed lower and upper bounds on the prices of American

call and put options, referred to as *capped option approximation* and Carr (1998) adopted a technique, referred to as *randomization* to obtain a new semi-explicit approximation for American options. Later, Zhu (2006) applied the homotopy analysis method to decompose the nonlinear PDE system arising from the free-boundary problem into an infinite series of linear PDEs which can be solved analytically and derived an exact and explicit solution for the first time.

So far, the option pricing problems for European and American options under the Black-Scholes model have been studied very extensively. However, as the research on option pricing develops, more and more empirical evidences suggest that the Black-Scholes model, which was ever a breakthrough in the option pricing theory, is inadequate to describe asset return and the behavior of the option market in modern financial practice. The real markets are never as ideal as the assumptions imposed in the Black-Scholes model. Some of these assumptions are demonstrated to be inconsistent with the financial practice, especially in periods of turmoil, including the market crash in 1987, the burst of the internet bubble in early 2000s and the financial crisis 2007-2009. More and more models are proposed as the extension of the classical Black-Scholes model and some new problems are arising at the same time.

The constant volatility assumption in the Black-Scholes model is the first one that is criticized by investors. From empirical studies (Rubinstein 1985), a well-known discrepancy between the Black-Scholes price and the market price was observed and the implied volatility varied with strike price and time to expiration, which is also referred to as *volatility smile* or *volatility skew*. To remedy the drawback of the constant volatility assumption, Dupire (1997) proposed a local volatility model in which the volatility is treated as a deterministic function of both the underlying and the time. The advantage of local volatility model is that there is no additional source of uncertainty introduced to the model and theoretically the market is still complete. Some other modified models are proposed to explain the volatility smile. Instead of a deterministic function, the volatility is also assumed to follow a stochastic process. In order to capture stochastic volatility and the leverage effect, Cox (1975) proposed a constant elasticity of variance (CEV) model, which is widely used by practitioners in the financial industry. Stein

& Stein (1991) assumed the volatility was driven by an arithmetic Ornstein-Uhlenbeck process and applied analytic techniques to derive an explicit closed-form solution for option pricing problem. Heston (1993) also proposed a stochastic volatility model in which the randomness of the variance process varies as the square root of variance. Such a model stands out from other stochastic volatility models because there exists an analytical solution for European options and thus has been subsequently developed by various researchers (Dragulescu & Yakovenko 2002, Zhu & Chen 2011). Since these extended models have introduced a new source of uncertainty, the market is incomplete and the corresponding option pricing problems have also been studied extensively (Bates 1996, Zvan et al. 1998, Fouque et al. 2000, Sepp 2003, Ito & Toivanen 2009).

As demonstrated above, a large number of extended models are proposed to remedy the constant volatility assumption in the Black-Scholes model. For the similar reasons, a large number of new models would also be proposed when some other assumptions are criticized by investors in the financial practice. In practice, the risk-free interest rate is not always a constant, which has violated the assumption in the Black-Scholes model. Vasicek (1977) first proposed an Ornstein-Uhlenbeck process to describe the evolution of interest rate and such a model is the first one to capture mean reversion, an essential characteristic of the interest rate that sets it apart from other financial prices. To overcome a disadvantage of the Vasicek model that the interest rate may be negative, Cox et al. (1985) introduced a new CIR model, in which the interest rate is guaranteed to be non-negative. Hull & White (1990) further demonstrated that the one-state-variable interest rate models of Vasicek (1977) and Cox et al. (1985) can be extended so that they were consistent with both the current-term structure of spot interest rate and the current-term structure of interest-rate volatility. As these stochastic interest rate models were introduced in the literature, option pricing problems with these extended models were also studied. Heath et al. (1992) first presented a unifying theory of option pricing under a stochastic term structure of interest rates. Closed-form formulae for certain types of European options were also derived in a stochastic interest rate economy (Amin & Jarrow 1992).

From the literature review above, we have learnt some basic ideas of research on option pricing. When one of the assumptions in the Black-Scholes model does not match with the

financial practice, some extended models would be first proposed to characterize the real world precisely and then option pricing problems under these new modified models would be studied again to demonstrate how option pricing is affected by the modification of the model. In this thesis, we would follow such an idea to study how short selling restrictions affect option pricing. As the first step of our study, we review the literature about short selling restrictions.

There is an extensive literature about short selling restrictions. Early studies mainly focus on how these short selling restrictions affect stock prices. Miller (1977) first pointed out that restrictions on short selling would lead to artificially inflated prices by restricting short sellers to express their opinions. Harrison & Kreps (1978) confirmed that implementing short selling restrictions would result in overvalued stock price because pessimists are shut out of the market and optimists do not take into account the absence of pessimists in setting prices. Figlewski (1981) also provided some empirical evidences that restrictions on short selling would overprice the underlying stock by demonstrating the connection between the level of short interest and subsequent stock returns. Diamond & Verrecchia (1987) concluded that restrictions on short selling would influence the rate at which private information is revealed to the public through observable trading because short sellers are the "good people" to discover the "true" price. All the literature suggests that short selling restrictions would hinder the market efficiency of searching for the "correct price" and result in the market price being overvalued.

Restrictions on short selling would also make some stocks hard to borrow when they are in short supply. There is also a considerable amount of previous work on hard-to-borrow stocks. Jones & Lamont (2002) found that hard-to-borrow stocks would have a high return rate because they may enter the borrowing market to earn an additional lending fees when shorting demand is high. Duffie et al. (2002) had ever studied how to price the lending fees when short selling required locating stock lender and bargaining over the lending fees. Evans et al. (2009) mainly focused on how options market and short selling interact with each other. Recently, Avellaneda & Lipkin (2009) proposed a new dynamic model to study the price-evolution of stocks that are subject to restrictions on short selling. They introduced a new stochastic process to describe the asset price based on the intensity of buy-ins, special phenomena associated with hard-to-borrow

stocks according to the regulations.

Up to date, a lot of studies have been conducted to explore the relation between short selling restrictions and asset prices. However, few of them mentioned how to price option when short selling restrictions are imposed. As the first part of this thesis, we mainly focus on option pricing under a new hard-to-borrow stock model. This part makes a contribution to the literature that we demonstrate how short selling restrictions affect option pricing for the first time.

In additional to restrictions, complete bans are also imposed on short selling by some regulators in the financial crisis 2007-2009. Compared with restrictions which can be characterized with new mathematical model, complete bans are much more difficult to deal with. Even in the classic Black-Scholes model, short selling bans make the market incomplete. Consequently, option pricing with short selling bans is a special case of option pricing in an incomplete market, which has been studied extensively in the literature and is grouped into two categories in general.

Papers in the first category share a common feature that an equivalent martingale measure is chosen according to some optimal criterion. Since there are many equivalent martingale measures in incomplete markets, an investor has many choices when he intends to price options. Follmer & Schweizer (1991) proposed a criterion to choose the *minimal martingale measure*. Frittelli (2000) provided another criterion to define the *minimal entropy martingale measure*. Later, similar concepts, such as the *minimal distance martingale measure* and *minimax measure* were also proposed by Goll & Rüschendorf (2001) and Bellini & Frittelli (2002), respectively. Each measure will lead to a different price, which is "fair" according to the criteria they chose the measure. It is hard to justify which choice of these equivalent martingale measures is "correct". Calibration has to be implemented to demonstrate that their choice of equivalent martingale measures is consistent with the market data.

Papers in the second category include Karatzas & Kou (1996), Davis (1997), Rouge & El Karoui (2000), Musiela & Zariphopoulou (2004) and Hugonnier et al. (2005) . The key idea of these papers is utility indifference pricing. An investor chooses a utility function first

according to his risk preference. The utility indifference buying price p^b is the price at which the utility of the investor is indifferent between (1) paying nothing and not having the claim and (2) paying p^b now to receive the contingent claim at expire time (Henderson and Hobson, 2004). The utility indifference selling price is defined similarly. In finance literature, utility indifference price is also referred to as "private valuation", which emphasizes the proposed price is for an individual with particular risk preference and not a transactional price (Detemple & Sundaresan 1999, Tepla 2000).

Recently, Guo & Zhu (2017) proposed a completely new approach, referred to as the *equal-risk pricing approach*, which determines the derivative price by simultaneously analyzing the risk exposure of both parties involved in the contract. They aimed to find out an equal-risk price which distributes expected loss evenly between the two involved parties. Such an *equal-risk price* is interpreted as a fair price that both parties are happy to accept during the negotiation if they intend to enter into the derivative contract. However, they only produced a simple pricing formula when the payoff function is monotonic, such as European call and put options. It is difficult to extend the analysis method to general case when the payoff function is non-monotonic, such as a butterfly spread option.

1.3 Structure of thesis

The theme of this thesis is option pricing with short selling restrictions or bans being imposed. Accordingly, the thesis is organized as two parts: (1) the first part deals with option pricing problem under a new hard-to-borrow stock model where short selling is restricted; (2) the second part is to explore how short selling bans affect option pricing. Each part is also consist of several chapters that discuss one particular topic relevant to that part.

In Part 1, some restrictions have been imposed on short selling according to the Securities and Exchange Commission's regulation SHO. The short seller is required to borrow or arrange to borrow the stock that he intends to short in advance and he may also be required to buy back the stock they have shorted if there is too much risk in the market. These restrictions would

make some stocks hard to borrow and also result in a new phenomenon, i.e. *buy-in*. Avellaneda & Lipkin (2009) proposed a new dynamic model to describe the hard-to-borrow stocks with short selling restrictions being imposed. Under this new stock model, option pricing is first explored in Chapter 3 and Chapter 4.

Although an approximate semi-explicit pricing formula actually has been obtained for European call options, its derivation depends on an independence assumption, which has limited its application to general cases. In Chapter 3, we propose a new PDE approach to the European option pricing problem whether or not the assumption is valid. Through comparing the numerical results from our PDE system and the semi-explicit pricing formula, it is verified that their semi-explicit pricing formula is a good approximate solution when the independence assumption is reasonable. However, in the event that this is not the case, the full model needs to be solved with the PDE approach as demonstrated in Chapter 3.

In Chapter 4, we extend our PDE approach to American option pricing problem, which is reformulated as a linear complimentary problem (LCP). Then Lagrange multiplier approach (LMA)is applied to solve the resulting LCPs numerically. According to the numerical results, early exercise of American call option may be optimal when the option is deep in-the-money, although the underlying stock pays no dividends. Such a conclusion is significantly different from the classic result by Merton (1973). In other words, the underlying stock follows the hard-to-borrow stock model can be considered as a stock that is paying "equivalent" dividends because the owner could collect some lending fees. Our numerical results reassure the conclusion of Jensen & Pedersen (2016) under the hard-to-borrow stock model. In addition, we quantify why the early exercise would occur both from the view of financial and mathematical points. How the parameters in the hard-to-borrow stock model affect the optimal exercise price is also provided numerically.

In Part 2, short selling is completely banned in the market. Even we adopt the Black-Scholes model to describe the underlying stock price, the market is still incomplete because the assumption about short selling has been violated. Considering short selling bans as an example of convex trading constraints, Guo & Zhu (2017) proposed an *equal-risk pricing approach* to

option pricing problems in markets with trading constraints. They produced an elegant pricing formula when the payoff function is monotonic, such European call and put options. However, their analysis method cannot be applied to general cases where the payoff function is not monotonic, such the butterfly spread option. In this part, the we intend to establish a PDE framework for equal-srisk pricing approach so that it not only recovers analytical pricing formula for European call and put options, but also produces equal-risk price for general contingent claims. It is pointed that the HJB equation would be involved in the process of establishing PDE system. As a result, we first explore different solution approaches to the HJB equation in Chapter 5, Chapter 6 and Chapter 7 as preliminaries.

In Chapter 5, we derive an exact and explicit solution for the well-known HJB equation arising from the Merton problems subject to some general utility functions. The solution presented in Chapter 5 is written in the form of a Taylor's series expansion and constructed through the homotopy analysis method (HAM). The fully nonlinear HJB equation is decomposed into an infinite series of linear PDEs which can be solved analytically. To convincingly demonstrate the success of applying the HAM to solve the fully nonlinear HJB equation, which has many application even beyond mathematical finance, four examples are presented with the first two cases showing the accuracy of the HAM; while the last two demonstrating the versatility of this solution approach.

In Chapter 6, we manage to produce a closed-form analytical solution for the Merton problem defined on a finite horizon with exponential utility function. Our new solution is obtained through two distinct approaches: an indirect method and a direct method. In the former, the Merton problem with a family of parameterized utility functions is solved first and then an analytical solution is obtained as we take the limit with respect to a parameter. In the latter, the HJB equation is decomposed into two nonlinear ordinary differential equations (ODEs), which can be explicitly solved with some nonlinear transforms. We also demonstrate that these two solution are equivalent although they appear to be of different forms. A great advantage of having these two solution, particular the one obtained through the direct method, is that the optimal strategies can now be scrutinized and discussed from both mathematical and economic

viewpoint.

In Chapter 7, we propose a monotone numerical scheme to solve the HJB equation arising from the Merton problem with general utility functions. After demonstrating the stability, consistency, and monotonicity of the numerical scheme, we ensure that it really converges to the solution of the HJB equation. Some proper boundary conditions are provided for the first time when we implement the numerical scheme to solve the utility maximization problem. To demonstrate the performance of our numerical scheme, three examples are provided with different utility functions. Economic discussions are also presented based on our numerical results.

Finally, in Chapter 8, we successfully establish and implement a PDE framework for equal-risk pricing approach. Such a PDE framework can be apply to deal with option pricing problem whether or not the payoff function is monotonic. When the payoff function is monotonic, analytical pricing formula is derived, which is consistent with the results from Guo & Zhu (2017). When the payoff function is not monotonic, such as a butterfly spread option, equal-risk price is still provided through solving the HJB equation numerically, which is absent in Guo & Zhu (2017). Therefore, our PDE framework has indeed extended the range of application of equal-risk pricing approach. The effect of short selling bans are demonstrated via comparisons with the Black-Scholes price. Generally, short selling bans would decrease the price of European call option; while it has an opposite effect on European put options. As for the butterfly spread option, the effects of short selling would varies with current underlying stock price.

Chapter 2

Background

In this chapter, we review some mathematical knowledge that is employed as a useful tool for the studies on option pricing. Since they are preliminary work for our research, we provide these concepts and theorem directly without proofs. For more detailed information, the readers can refer to the reference book we provide in each section.

2.1 Stochastic calculus

In this section, some basic concepts and theorems in stochastic calculus are provided first. All of them play important roles in the theory of option pricing. More detailed information about the proofs can be found in Øksendal (2003), Shreve (2004).

The first concept is *martingale*, which can be considered as a mathematical model of fair games because it can exclude the possibility of winning strategies based on game history.

Definition 2.1.1. On a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{Q})$, a stochastic process M_t is called a *martingale* with respect to the filtration $\mathcal{F}_t$ if

1. M_t is $\mathcal{F}_t$-measurable for all t,
2. $\mathbf{E}_{\mathbb{Q}}[|M_t|] < \infty$, for all t,
3. $\mathbf{E}[M_s|\mathcal{F}_t] = M_t$ for all $s > t$.

It is remarked that both Brownian motion and Ito integral are martingale. Inversely, we also have the following theorem

Theorem 2.1.1. (Martingale representation theorem) Suppose W_t is a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{Q})$, $\mathcal{F}_t$ is the filtration generated by the Brownian motion W_t, and M_t is a martingale with respect to the filtration $\mathcal{F}_t$. Then there exists a predictable process ϕ_t, such that

$$M_t = \mathbf{E}M_0 + \int_0^t g_u dW_u. \tag{2.1.1}$$

The martingale representation theorem states that a martingale with respect to the filtration generated by Brownian motion can be expressed as an initial value plus an Ito integral with respect to the Brownian. Such a theorem only asserts the existence of the representation and does not held to find it explicitly. It is possible in some cases to determine the form of the representation using Malliavin calculus (Nualart 2009).

In financial mathematics, stock price is always assumed to follow some special stochastic processes. Here we introduce the concept of *Ito process*.

Definition 2.1.2. An Ito process is defined to be an adopted stochastic process that can be expressed as the sum of an integral with respect to Brownian motion and an integral with respect to time,

$$X_t = X_0 + \int_0^t \sigma_s dW_s + \int_0^t \mu_s ds, \tag{2.1.2}$$

where W_t is a Brownian motion, σ is a predictable W-integral process and μ is predictable and Lebesgue integral.

The classical and famous *Ito formula*, which serves as the stochastic calculus counterpart of the chain rule, is provided in the next theorem.

Theorem 2.1.2. (Ito formula) Assume X_t be an Ito process given by Equation (2.1.2) and $f(t,x) \in C^{1,2}([0,\infty) \times \mathbf{R})$, then we have

$$df(t, X_t) = (\frac{\partial f}{\partial t} + \mu_t \frac{\partial f}{\partial x} + \frac{\sigma_t^2}{2}\frac{\partial^2 f}{\partial x^2})dt + \sigma_t \frac{\partial f}{\partial x} dW_t. \tag{2.1.3}$$

The following theorem is also very important in the theory of financial mathematics as it describes how to convert from the physical measure to the risk-neutral measure.

Theorem 2.1.3. (Girsanov theorem) Let W_t be a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{Q})$, $\mathcal{F}_t$ be a filtration for this Brownian motion and θ_t be an adapted process. Define

$$\begin{cases} Z_t = exp\{-\int_0^t \theta_u dW_u - \frac{1}{2}\int_0^t \theta_u^2 du\}, \\ \bar{W}_t = W_t + \int_0^t \theta_u du. \end{cases} \tag{2.1.4}$$

and assume that

$$\mathbf{E}_{\mathbb{Q}} \int_0^T \theta_u^2 Z_u^2 du < \infty. \tag{2.1.5}$$

The new measure $\mathbb{P}$ is defined as

$$\left.\frac{d\mathbb{P}}{d\mathbb{Q}}\right|_t = Z_t, \tag{2.1.6}$$

which is also referred to as *Radon-Nikodym derivative*. Then the new process $\bar{W}_t$ is a Brownian motion under the new probability measure $\mathbb{P}$.

The following one is Feynman-Kac theorem that establishes an important relation between the conditional expectation under martingale measure $\mathbb{Q}$ and a parabolic PDE system.

Theorem 2.1.4. (Feynman-Kac theorem) Assume that X_t be an Ito process described as

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \tag{2.1.7}$$

where $b(t, x)$ and $\sigma(t, x)$ are continuous, and they satisfy the Lipschitz condition:

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq C|x - y| \tag{2.1.8}$$

with C being constant. If $f(x) \in C_0^2(\mathbf{R})$ and $r(x) \in C(\mathbf{R})$, define a function as

$$v(t, x) = \mathbf{E}[e^{-\int_t^T r(X_s)ds} g(X_T)|X_t = x]. \tag{2.1.9}$$

Then it be expressed as a solution to the following PDE system

$$\begin{cases} \dfrac{\partial v}{\partial t} + \dfrac{\sigma(t,x)^2}{2}\dfrac{\partial^2 v}{\partial x^2} + b(t,x)\dfrac{\partial v}{\partial x} - r(x)v = 0, \\ v(T,x) = g(x). \end{cases} \tag{2.1.10}$$

Feynman-Kac formula implies that a problem of calculating the condition expectation can be connected to problem of solving PDE. On the other direction, for some special PDE system, the solution may have an stochastic representation. It offers a method of solving certain PDE by simulating random paths of an Ito process.

Theorem 2.1.5. Consider the partial differential equation

$$\frac{\partial u}{\partial t}(t,x) + b(t,x)\frac{\partial u}{\partial x} + \frac{\sigma(t,x)^2}{2}\frac{\partial^2 u}{\partial x^2}(t,x) - r(x)u(t,x) + f(t,x) = 0, \tag{2.1.11}$$

subject to the terminal condition

$$u(x,T) = g(x), \tag{2.1.12}$$

where $b(t,x), \sigma(t,x), r(x)$ and $f(x)$ are known functions, T is a parameter and $u : [0,T]\times\mathbf{R} \to \mathbf{R}$ is the unknown function. Then its solution can be expressed in a conditional expectation as

$$u(t,x) = \mathbf{E}_{\mathbb{Q}}\{\int_t^T e^{-\int_t^s r(X_\tau)d\tau} f(s,X_r)ds + e^{-\int_t^T r(X_\tau)d\tau} g(X_T)|X_t = x\}, \tag{2.1.13}$$

under the probability measure $\mathbb{Q}$ such that X_t is an Ito process driven by the SDE

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dW_t^{\mathbb{Q}} \tag{2.1.14}$$

where $W_t^{\mathbb{Q}}$ is a Winner process under $\mathbb{Q}$ and the initial condition for $X(t)$ is $X(t) = x$.

2.2 Fundamental theorems of asset pricing

In this section, we review the very important theorems of asset pricing, which provide necessary and sufficient conditions for a market to be arbitrage free and a market to be complete. Before

stating these theorems, some concepts are introduced first. Consider a financial model on a complete probability space $(\Omega, \mathcal{F}, \mathbb{Q})$. A risk-neutral measure is defined as follows.

Definition 2.2.1. (Risk-neutral measure) A probability measure $\mathbb{P}$ is said to be *risk-neutral* if

1. probability measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent, i.e., for every $A \in \mathcal{F}, \mathbb{Q}(A) = 0$ if and only if $\mathbb{P}(A) = 0$,

2. under the probability measure $\mathbb{P}$, the discounted stock price $e^{-rt}S_t$ is a martingale.

An *arbitrage* is a way of trading so that one starts with zero capital and at some later time T is sure not to have lost money and furthermore has a positive probability of making profits. Such a financial concept can also be defined mathematically.

Definition 2.2.2. (Arbitrage) An arbitrage is a portfolio value process X_t satisfying $X_0 = 0$ and also satisfying for some time $T > 0$,

$$\mathbb{Q}(X_T \geq 0) = 1, \qquad \mathbb{Q}(X_T > 0) > 0. \tag{2.2.1}$$

With risk-neutral measure and arbitrage being introduced, the first fundamental theorem of asset pricing is briefly stated in the following.

Theorem 2.2.1. (First fundamental theorem of asset pricing) There exists a risk-neutral measure $\mathbb{Q}$ *if and only if* there is no arbitrage opportunity in the market.

In the theory of pricing assets, investors would never offer prices derived form a model that admits arbitrage. Such a theorem provides us a simple mathematical condition so that we can easily check whether or not the model has such a fatal flaw. In addition to arbitrage, another important concept arising from the option pricing theory is *market completeness*

Definition 2.2.3. A market is said to be *complete* if any financial derivative can be perfectly replicated. In other word, for any contingent claim Z, there exists a self-financing strategy (ϕ_t^0, ϕ_t) such that

$$Z = V_0 + \int_0^T \phi_t dS_t + \int_0^T \phi_t^0 dS_t^0, \tag{2.2.2}$$

where S_t and S_t^0 are the prices of stock and bond at time t, respectively.

Now we propose another theorem to provide an equivalent mathematical description on market completeness.

Theorem 2.2.2. (Second fundamental theorem of asset pricing) Consider a market model that has risk-neutral measures. The market is complete *if and only if* there exists a unique risk-neutral measure $\mathbb{Q}$.

The proofs of these two fundamental theorems of asset pricing can be found in Harrison & Kreps (1979), Harrison & Pliska (1981), Delbaen & Schachermayer (1994).

Generally, option pricing problem in a complete market is easy to deal with for there is a unique pricing measure. In next section, we would take the Black-Scholes model an example to demonstrate how to price options in a complete market.

2.3 The Black Scholes model

The classic Black-Scholes model is a complete model because there is a unique risk-neutral measure (Shreve 2004). In such a complete market, any option can be duplicated perfectly by some self-financing dynamic portfolio strategies. The option price corresponds to the cost of constructing such a replicating portfolio.

In the Black-Scholes model, the stock price is assumed to follow a geometric Brownian motion:

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \tag{2.3.1}$$

where μ is the drift, σ is the volatility and W_t is a standard Brownian motion. The price of risk-free bond follows

$$dP_t = rP_t dt, \tag{2.3.2}$$

where r is the risk-free interest rate. With all assumptions in the Black-Scholes model being satisfied, we construct a portfolio consisting of ϕ_t shares of stocks and ψ_t shares of bonds at

time t such that it replicates the option, i.e.,

$$V(t, S_t) = \phi_t S_t + \psi_t P_t. \tag{2.3.3}$$

Differentiating Equation (2.3.3) and applying self-financing property on the right side and Ito formula on the left side, we obtain

$$(\frac{\partial V}{\partial t} + \mu S_t \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2})dt + \sigma S_t \frac{\partial V}{\partial S} dW_t = (\mu \phi_t S_t + r\psi_t P_t)dt + \sigma \phi_t S_t dW_t. \tag{2.3.4}$$

Equating the coefficients of the dW_t terms leads to

$$\phi_t = \frac{\partial V}{\partial S}. \tag{2.3.5}$$

Then substituting $\psi_t P_t = V - \phi_t S_t$ into Equation (2.3.4), we come to the Black-Scholes PDE

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0. \tag{2.3.6}$$

The terminal conditions for such a PDE system is

$$V(T, S) = Z(S), \tag{2.3.7}$$

where $Z(S)$ is the payoff function of the European options. Black & Scholes (1973) first derived a closed-form analytical solution, which is also known as the Black-Scholes formula, for the price of a European call option. The pricing formula reads:

$$V(t, S) = SN(d_1) - Ke^{-r(T-t)}N(d_2), \tag{2.3.8}$$

where

$$\begin{cases} d_1 = \frac{\ln(S/K) + (r + \sigma^2/2)}{\sigma\sqrt{T-t}}, \\ d_2 = d_1 - \sigma\sqrt{T-t}, \end{cases} \tag{2.3.9}$$

and $N(d)$ is the standard normal distribution function defined as

$$N(d) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d} e^{-\frac{x^2}{2}} dx. \tag{2.3.10}$$

2.4 Utility indifference pricing

When the market is incomplete, portfolio replication method becomes invalid for option pricing problems. Instead, utility indifference pricing is a popular and powerful method of option pricing in incomplete markets (Hodges 1989, Henderson & Hobson 2004). *Utility indifference price* is the price at which an agent would have the same expected utility level by exercising a financial transaction as by not doing so. A formal definition is presented as follows.

Definition 2.4.1. Given a utility function U and a contingent claim Z_T with known payoff functions at expire date T. If we let the function $V : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be defined by

$$V(x, k) = \max_{X_T \in \mathcal{A}(x)} \mathbf{E}[U(X_T + kZ_T)], \tag{2.4.1}$$

where x is the initial endowment, $\mathcal{A}(x)$ is the set of all self-financing portfolios at time T with endowment x, and k is the number of the claims to be purchased or sold, then the utility indifference purchase price $v^p(k)$ for k units of Z_T is the solution of

$$V(x - v^p(k), k) = V(x, 0), \tag{2.4.2}$$

and the utility indifference selling price $v^s(k)$ is the solution of

$$V(x + v^s(k), -k) = V(x, 0), \tag{2.4.3}$$

Utility indifference prices have a number of appealing properties. First of all, in contrast to the Black-Scholes price and many other alternative pricing method in incomplete markets, utility indifference prices are nonlinear in the number of options k due to the concavity of utility

function. Secondly, utility indifference prices are equivalent to the Black-Scholes prices when the market is complete. Thirdly, utility indifference prices are monotonic and concave.

From Definition 2.4.1, two stochastic control problems are involved. The first one is the optimal investment problem when the investor has bought or sold the contingent claim; while the second one is the optimal investment problem when the investor has a zero position in the claim. These optimal investment problems date back to Merton (1969, 1971), which would be presented in the next section.

2.5 Optimal investment problem

Optimal investment problem was first reformulated by Merton (1969) and thus was also referred to as *the Merton problem*. In this section, such a classical problem was reviewed briefly.

Consider a financial market with two assets being traded continuously on a finite horizon $[0, T]$. One asset is a risk-free bond, whose price $P(t)$ evolves as

$$dP(t) = rP(t)dt, \qquad t \in [0, T], \tag{2.5.1}$$

with r being the risk-free interest rate. The other one is a risky asset with its price following a geometric Brownian motion

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \qquad t \in [0, T], \tag{2.5.2}$$

where μ is the drift rate, σ is the volatility, and $W(t)$ is a standard Brownian motion.

An investor starts with a known initial wealth x_0 and the wealth at time t is denoted as $X(t)$. At any time t, prior to T, the investor needs to make a decision on how much to consume and, in the mean time, how much to invest in stock markets, in order to maximize his expected utility from intermediate consumption and terminal wealth. The consumption rate per unit time at time t is denoted as $c(t)$ and the investment proportion $u(t)$ represents the fraction of total wealth that is invested in the risky asset at time t. The remaining fraction $1 - u(t)$ is thus

left in form of the risk-free bond within the framework of this two-asset model. The investment proportion on the underlying stock $u(t)$ may be negative, which is to be interpreted as short selling. The remaining proportion $1 - u(t)$ may also become negative and this corresponds to borrowing at the interest rate r. As a result, the total wealth $X(t)$ is governed by the following SDE:

$$dX(t) = \{[r + u(t)(\mu - r)]X(t) - c(t)\}dt + X(t)u(t)\sigma dW(t). \tag{2.5.3}$$

The objective of the Merton problem is to obtain the optimal investment and consumption policies, i.e. to determine $u(t)$ and $c(t)$, such that the expected utility from accumulated consumption and the terminal wealth is maximized. Mathematically, such an objective functional is stated as

$$\max_{(u(\cdot),c(\cdot))} \quad \mathbf{E}[\int_0^T e^{-\rho s}U(c(s))ds + e^{-\rho T}B(X_T)], \tag{2.5.4}$$

where $\mathbf{E}$ is the expectation operator; ρ is the subjective discount rate; U is a function measuring the utility from intermediate consumption $c(t)$ and B is also a function measuring the utility from terminal wealth X_T. In addition, the fact that the consumption and wealth process can not be negative in practice leads to two constraints being imposed on the optimization

$$c(t) \geq 0, \quad X(t) \geq 0, \quad t \in [0, T]. \tag{2.5.5}$$

As a result, the Merton problem has been reformulated as a stochastic optimal control problem with the objective functional (2.5.4), driven by the dynamics of the wealth (2.5.3), and subject to the constraints (2.5.5).

2.6 Stochastic control problem and the HJB equation

It is observed that stochastic control problems arise both in the utility indifference pricing problem and the Merton problem. To solve these expected utility maximization problems, the Hamilton-Jacobi-Bellman (HJB) equation, a powerful mathematical tool to deal with stochastic control problem, is introduced in this section.

Let $(\Omega, \mathcal{F}, \mathbb{Q})$ be a completed probability space satisfying the usual condition, on which is defined an m-dimensional standard Brownian motion $W(t)$. We consider the following stochastic controlled system:

$$\begin{cases} dx_t = b(t, x(t), u(t))dt + \sigma(t, x(t), u(t))dW_t, \\ x_0 = x. \end{cases} \tag{2.6.1}$$

with the cost functional

$$J(u(\cdot)) = \mathbf{E}\int_0^T f(t, x(t), u(t))dt + h(x(T)). \tag{2.6.2}$$

Define admissible control set as

$$\mathcal{U}[0,T] := \{u : [0,T] \times \Omega \to U | u \text{ is measurable and } \{\mathcal{F}_t\}_{t\geq 0}\text{-adopted}\}, \tag{2.6.3}$$

where U is a Borel set. The optimal stochastic control problem can be stated as follows:

Problem 1. Minimize (2.6.2) subject to the state equation (2.6.1) over the admissible control set $\mathcal{U}[0,T]$.

To guarantee the existence and uniqueness of the solution, some assumptions on the coefficients as follows:

Assumption 2. The maps $b : [0,T] \times \mathbf{R}^n \times U \to \mathbf{R}^n$, $\sigma : [0,T] \times \mathbf{R}^n \times U \to \mathbf{R}^{n\times m}$, $f : [0,T] \times \mathbf{R}^n \times U \to \mathbf{R}$, and $h : \mathbf{R}^n \to \mathbf{R}$ are uniformly continuous, and there exists a constant $L > 0$ such that for $\phi(t,x,u) = b(t,x,u), \sigma(t,x,u), f(t,x,u), h(x)$

$$\begin{cases} |\phi(t,x,u) - \phi(t,\bar{x},u)| \leq L|x-\bar{x}|, \quad \forall t \in [0,T], x, \bar{x} \in \mathbf{R}^n, u \in U \\ |\phi(t,0,u)| \leq L, \qquad\qquad \forall (t,u) \in [0,T] \times U. \end{cases} \tag{2.6.4}$$

According to the dynamic programming method (Yong & Zhou 1999), we introduce a se-

quence of similar problems. For any $(s, y) \in [0, T) \times \mathbf{R}^n$, consider the state equation:

$$\begin{cases} dx_t = b(t, x(t), u(t))dt + \sigma(t, x(t), u(t))dW_t, \\ x_s = y. \end{cases} \tag{2.6.5}$$

along with the cost functional

$$J(u(\cdot); s, y) = \mathbf{E} \int_s^T f(t, x(t), u(t))dt + h(x(T)). \tag{2.6.6}$$

The value function is defined as

$$V(s, y) = \min_{u(\cdot) \in \mathcal{U}[s,T]} J(u(\cdot); s, y). \tag{2.6.7}$$

Based on Bellman's principle of optimality, we have the following theorem

Theorem 2.6.1. When Assumption 2 hold, for any $(s, y) \in [0, T) \times \mathbf{R}^n$, we have

$$V(s, y) = \min_{u(\cdot) \in \mathcal{U}[s,T]} E\{\int_s^{\hat{s}} f(t, x(t; s, y, u(\cdot)), u(t))dt + V(\hat{(}s), x(\hat{s}; s, y, u(\cdot)))\}, \quad \forall 0 \leq s \leq \hat{s} \leq T. \tag{2.6.8}$$

Following the Bellman's principle of optimality, we provide the HJB equation governing the value function $V(s, y)$.

Theorem 2.6.2. When Assumption 2 holds and the value function $V(s, y) \in C^{1,2}([0, T] \times \mathbf{R}^n)$. Then $V(s, y)$ is a solution of the following terminal value problem of a second-order partial differential equation:

$$\begin{cases} -v_t + \max_{u \in U} G(t, x, u, -v_x, -v_{xx}) = 0, \\ v(T, x) = h(x), \end{cases} \tag{2.6.9}$$

where $G(t, x, u, p, P) = \frac{1}{2}\text{tr}(P\sigma(t, x, u)\sigma(t, x, u)^T) + \langle p, b(t, x, u)\rangle - f(t, x, u)$, $\forall (t, x, u, p, P) \in [0, T] \times \mathbf{R}^n \times U \times \mathbf{R}^n \times \mathcal{S}^n$ and $\mathcal{S}^n$ is the set of all $n \times n$ symmetric matrices.

The proofs of these theorems in this section can be found in Yong & Zhou (1999).

Part I: Option pricing with short selling restrictions

In this part, short selling is still allowed but restriction has been imposed. According to the U.S. Securities and Exchange Commission's Regulation SHO, the short seller is required to borrow or arrange to borrow the stock that he intends to short in advance. The availability of the stock for borrowing depends on the market conditions. While many stocks are easily borrowed; others may be in short supply. In the latter case, short selling may be costly because the stock is hard to borrow and the short seller has to pay additional borrowing fees. Sometimes, the short seller fails to deliver the stock that he has shorted, which is called "fail-to-deliver" risk. When there is too much risk, the clearing firm would forcibly require the short seller to repurchase (buy in) the stock that has been shorted following the Securities and Exchange Commission's Regulations SHO. To describe the price-evolution for hard-to-borrow stocks, Avellaneda & Lipkin (2009) proposed a new dynamic model, in which a new stochastic process was introduced to characterize the intensity of buy-ins.

The main contribution of this part is that, for the first time, we explore option pricing problems under a hard-borrow stock model by establishing a PDE system. Although an approximate semi-explicit pricing formula for European call options has been obtained, its derivation depends on an independence assumption, which has limited its application to general cases. In Chapter 3, we propose a PDE approach to European call option pricing and an ADI numerical scheme is also provided to solve the PDE system. Numerical results calculated from both our PDE system and the semi-explicit pricing formula are compared to demonstrated that our PDE approach is a broader way than the semi-explicit pricing formula. Then such a PDE approach is extended to deal with American call option pricing problem in Chapter 4. It is furthermore reformulated as a linear complementarity problem, which is solved numerical with the Lagrange multiplier approach. Numerical results are produced to demonstrate how the parameters in this hard-to-borrow stock model affect option prices and optimal exercise price.

Chapter 3

Pricing European call options under a hard-to-borrow stock model

3.1 Introduction

Option pricing is one of the most important topics in quantitative finance ever since Black & Scholes (1973) proposed an analytical and quantitative formula for pricing European options, which has laid the solid foundation for pricing financial derivatives. One of the most important assumptions in their celebrated model is that short selling is permitted without any cost, while the market regulations suggest otherwise. In most stock markets, naked short selling is forbidden for it may result in too much fail-to-deliver risk. However, short selling with some restrictions is usually allowed. If an investor wants to short a stock, he has to borrow from others in advance. The availability of stocks for borrowing depends on market conditions. While some are easily borrowed, others may be in short supply. In the latter, they are referred to as *hard-to-borrow* stocks (Avellaneda & Lipkin 2009).

In general, *buy-ins* are associated with hard-to-borrow stocks according to the Securities and Exchange Commission's Regulation SHO[1]. Following the rules, in order to cover shortfalls in delivery of such stocks, the short seller may be "forced" to repurchase the stock when the

[1]The reader can visit the website: www.sec.gov to learn more about Regulation SHO.

risk of short selling goes up to a certain level. Such an event is called a *buy-in*. To measure the risk, the *short interest* is defined as the ratio of tradable shares being shorted to shares in the market (Asquith et al. 2005). As the short interest increases, the fail-to-deliver risk accumulates more and more in the market. When the risk is almost beyond the control, a buy-in would be triggered. Once a buy-in occurs, the fail-to-deliver risk would be reduced as the short interest of the stock falls down. Then, the risk accumulates gradually with time going on and the short interest goes up again. When it comes to a certain level again, another buy-in follows. The investors have to take the possible buy-ins into consideration. These buy-ins are always considered as stochastic dividend yields or convenience yields, because the holders of hard-to-borrow stocks can obtain lending fees by lending their stocks to the investors who wish to maintain short positions and not risk buy-ins (Avellaneda & Lipkin 2009). The harder it is to borrow the stock, the more the lending fees will be.

In the literature, there is a considerable amount of research about short selling. Diamond & Verrecchia (1987) considered constraints on short selling and asset price adjustment to private information. Duffie et al. (2002) presented a model of asset valuation in which short selling is achieved by searching for security lenders and bargaining over the terms of lending fee. Jones & Lamont (2002) pointed out that, from market data, stocks are overpriced when short-sale constraints are imposed. Evans et al. (2009) mainly focused on how options market and short selling interact with each other.

In 2009, Avellaneda and Lipkin (A&L) observed various phenomena associated with hard-to-borrow stocks from market data. All these phenomena were related to buy-ins. To describe these phenomena better, they presented a new dynamic model for hard-to-borrow stocks by introducing a stochastic *buy-in rate*, an additional factor absent in standard models. The buy-in rate represents the frequency at which buy-ins take place. It has been manifested by Avellaneda & Lipkin (2009) that the features such a new model shows are in good agreement with empirical (market) observations. Since then, it has attracted much attention from both theoretical and empirical aspects (Avellaneda & Zhang 2010, Li et al. 2014, Jensen & Pedersen 2016).

The A&L model is a fully coupled system with the stock price and the buy-in rate depending on each other. After making an assumption that the buy-in rate is independent of the Brownian motion that drives the stock price, which is referred to as *the independence assumption* hereafter, they obtained a formula to price European options. The pricing formula is semi-explicit because a series of unknown weight functions are still involved. Monte Carlo simulations are necessary to estimate the weight functions when we try to obtain the numerical results from semi-explicit formula. However, they did not mention how to implement the Monte Carlo simulations when the intensity of Poisson process is also a stochastic process, instead of a constant or a deterministic function.

It is remarked that the independence assumption is significantly important because it makes the fully coupled system become semi-coupled, which has facilitated the derivation of their semi-explicit pricing formula for European call options. However, it has also limited its application to general cases. Provided that the independence assumption is indeed reasonable, the semi-explicit formula is a good approximate solution. While, in the event that this is not the case, it is the full model that needs to be solved with a PDE approach as demonstrated in this chapter. Not only the stock price but also the buy-in rate is considered as a variable of option value when the PDE system is established. To obtain the numerical results, two schemes are carefully chosen based on different approaches to the jump term. Both of them have adopted the ADI scheme to improve the computational efficiency.

The contribution of this chapter is to present a PDE approach to pricing European call options, which enlarges the application of this new dynamic model for hard-to-borrow stocks. In other words, our PDE approach would not only recover the special case where the independence assumption is valid, but also deal with the case where the independence assumption is no longer valid. In the latter, the semi-explicit pricing formula does not work any more.

The chapter is organized as follows. In Section 3.2, the A&L model is reviewed first, including the risk-neutral measure and the semi-explicit pricing formula for European call options. In Section 3.3, the PDE system is established with a set of appropriate boundary conditions. In Section 3.4, two numerical methods are presented to solve the PDE system

based on the different treatments for the jump term and then Monte Carlo simulations are also provided to estimate the unknown weight functions in order to implement the semi-explicit pricing formula. In Section 3.5, the numerical results are provided and some discussions are presented. Conclusions are given in the last section.

3.2 The A&L model for hard-to-borrow stocks

Avellaneda and Lipkin (A&L) presented a model by introducing a Poisson process $N_{\lambda_t}(t)$ to describe buy-ins associated with hard-to-borrow stocks. The buy-in rate λ_t represents the frequency at which buy-ins take place. The stock price S_t and the buy-in rate λ_t satisfy the following stochastic differential equations (SDEs) under the physical measure $\mathbb{P}$:

$$\begin{cases} \dfrac{dS_t}{S_t} = \sigma dW_t + \gamma\lambda_t dt - \gamma dN_{\lambda_t}(t) \\ dx_t = \kappa dZ_t + \alpha(\overline{x} - x_t)dt + \beta\dfrac{dS_t}{S_t}, \quad x_t = \ln(\lambda_t), \end{cases} \tag{3.2.1}$$

where dN_{λ_t} denotes the increment of a standard Poisson process with intensity λ over the interval $(t, t+dt)$. The parameters σ and γ are respectively the volatility and the price elasticity of demand due to buy-ins; W_t and Z_t are two independent standard Brownian motions which drive the stock price and the buy-in rate respectively. For convenience, x_t, the logarithm of the buy-in rate λ_t, is also called buy-in rate hereafter. The second equation describes the evolution of the buy-in rate with κ being the volatility of the rate, $\overline{x}$ the long-time equilibrium value for buy-in rate x_t, α the speed of mean-reversion and β the coupling parameter that couples the change in price with the buy-in rate.

In order to introduce a positive feedback between increase in buy-in rate and increase in stock price, β was required to be positive in Avellaneda & Lipkin (2009). When a buy-in occurs, the stock price drops down and the buy-in rate also falls to a low level simultaneously because of the coupling term $\beta\dfrac{dS_t}{S_t}$. Another buy-in is unlikely to occur immediately following the previous one because the previous one has released some fail-to-deliver risk. With the time moving on, the risk accumulates gradually and the buy-in rate goes up again. Another buy-in

occurs in the future once the risk goes up again and reaches a certain level. This demonstrates how the stock price and the buy-in rate interact with each other.

3.2.1 The risk-neutral measure

It should be pointed out that the A&L model operates in an incomplete market since an additional source of uncertainty has been introduced through the buy-in rate, which is not a tradable quantity (Tankov 2003). Therefore, it is impossible to perfectly hedge a portfolio composed of hard-to-borrow stocks and there does not exist a unique risk-neutral measure. For pricing a derivative, a risk-neutral measure needs to be defined for the processes S_t and x_t first. What Avellaneda and Lipkin did was to introduce an arbitrage-free pricing measure, which is equivalent to changing the drift of the Brownian motion associated with the underlying stock. Mathematically, to conduct measure transform, two new processes are defined as

$$\tilde{W}_t = W_t + \int_0^t \frac{\gamma\lambda_l - r}{\sigma} dl, \tag{3.2.2}$$

and

$$\tilde{Z}_t = Z_t + \int_0^t \frac{\alpha z(l, x_l, S_l)}{\kappa} dl, \tag{3.2.3}$$

where $z(t, x, S)$ is an arbitrary function. By Girsanov's theorem, $\tilde{W}$ and $\tilde{Z}$ are two independent Brownian motions under the risk-neutral measure $\mathbb{Q}$ defined by

$$\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_t = exp\Big\{-\int_0^t \big[\frac{\gamma\lambda_l - r}{\sigma} + \frac{\alpha z(l, x_l, S_l)}{\kappa}\big] dW_l - \frac{1}{2}\int_0^t \big[\frac{(\gamma\lambda_l - r)^2}{\sigma^2} + \frac{\alpha^2 z^2(l, x_l, S_l)}{\kappa^2}\big] dl\Big\}, \tag{3.2.4}$$

which is the so-called Radon-Nikodym derivative that facilitates the change of measure. Under this risk-neutral measure, the dynamics of the A&L model become

$$\begin{cases} \dfrac{dS_t}{S_t} = \sigma d\tilde{W}_t + r dt - \gamma dN_{\lambda_t}(t), \\ dx_t = \kappa d\tilde{Z}_t + [\alpha(\overline{x}^* - x_t)]dt + \beta \dfrac{dS_t}{S_t}, \end{cases} \tag{3.2.5}$$

where $\overline{x}^* = \overline{x} - z$.

Financially, any source of uncertainty needs to be compensated by the associated market price of risk or risk premium. In the classic Black-Scholes model, the market price of risk for the underlying is $\frac{\mu - r}{\sigma}$ (Wilmott et al. 1995). On the other hand, in the Heston model, an additional source uncertainty is introduced by the stochastic volatility and an additional market price of volatility risk is defined through an arbitrary function, i.e., $\lambda(t, S, v)$ in Heston (1993), which may appear in a more general form for other stochastic volatility models discussed in Fouque et al. (2000). In the A&L model, the new buy-in process also brings in an additional source of uncertainty and the corresponding market price of buy-in risk is represented by the function $z(t, x, S)$ in Equation (3.2.3). Furthermore, it should be remarked that the market price of risk for the stock in the A&L model becomes $\frac{\gamma\lambda - r}{\sigma}$, which is different from its counterpart $\frac{\mu - r}{\sigma}$ in the Black-Scholes model.

When a market is complete, the market price of risk for the underlying is unique, such as the term $\frac{\mu - r}{\sigma}$ in the Black-Scholes model. When a market is incomplete, the market price of risk is specified after a risk-neutral measure is chosen, or in a vice versa way in financial practice that a market price of risk is extracted from market data first, which then implicitly dictates the risk-neutral measure to be used in pricing a derivative. Therefore, the market price of buy-in risk in the A&L model should be determined by market data, just as the market price of volatility risk in the Heston model needs to be calibrated from market data (Bollerslev et al. 2011). For simplicity, Avellaneda & Lipkin (2009) effectively set $z(t, x, S)$ to be zero, which is a standard treatment in the Heston model as well (Rouah 2013).

3.2.2 An approximate semi-explicit pricing formula

Under the risk-neutral measure they chose, we have

$$dx_t = [\alpha(\overline{x} - x_t) + \beta r]dt + \kappa d\tilde{Z}_t + \beta\sigma d\tilde{W}_t - \beta\gamma dN_{\lambda_t}(t). \tag{3.2.6}$$

To obtain a semi-explicit pricing formula, Avellaneda & Lipkin (2009) made an independence assumption that the buy-in rate is independent of Brownian motion that drives the stock price.

This key assumption about independence has really facilitated the derivation of their pricing formula for a European call option. As a result, Equation (3.2.6) becomes

$$dx_t = [\alpha(\bar{x} - x_t) + \beta r]dt + \kappa d\tilde{Z}_t - \beta\gamma dN_{\lambda_t}. \tag{3.2.7}$$

An approximate semi-explicit pricing formula for a European call can be expressed as

$$C(S, K, T) = \sum_{0}^{\infty} \Pi(n, T) C^{BS}(S(1-\gamma)^n, K, T, r, \sigma), \tag{3.2.8}$$

where $C^{BS}(S, K, T, r, \sigma)$ is the value of a European call option calculated from the Black-Scholes formula with being S the price of underlying stock, T the time to maturity, and K strike price; the weight functions are defined as:

$$\Pi(n, T) \quad = \quad P(\int_0^T dN_{\lambda_t} = n). \tag{3.2.9}$$

However, the pricing formula (3.2.8) actually involves a series of unknown functions $\Pi(n, T)$ which are unnecessary for computing option price from a completely explicit formula, such as the famous Black-Scholes formula. It is for this reason that pricing formula (3.2.8) is called a semi-explicit one. If the intensity of Poisson process is a constant or a deterministic function, the expression (3.2.9) will be so easily computed that the form (3.2.8) becomes a completely explicit formula according to Merton (1976). When the intensity is driven by a stochastic process, the computation of the weight functions requires simulations for the intensity process.

In addition, the original model (3.2.1) represents a fully coupled system with the stock price and the buy-in rate depending on each other; while the independence assumption makes the fully coupled SDEs become semi-coupled, as the buy-in rate x_t no longer depends on the stock price S_t. This independence assumption has facilitated the derivation of the semi-explicit pricing formula for European call options under the A&L model. However, it has limited the application to more general cases. In this chapter, we present a PDE approach to price European call options. Given that the intensity of Poisson process is described by a stochastic

process x_t instead of a constant or a deterministic function, the buy-in rate x is also considered as a variable of option value in addition to stock price S. The PDE approach can also be extended to the American case in the next chapter, while the semi-explicit pricing formula proposed by Avellaneda & Lipkin (2009) is not easily extended. In other words, our PDE approach has a wide range of application.

3.3 The PDE system for European call options

In this section, the PDE system governing the value of European call options is derived first for the case where the independence assumption is removed. To obtain a properly-closed PDE system, some boundary conditions, especially those in the buy-in rate direction, are presented from mathematical and financial aspects. A simplified PDE system is also provided when the independence assumption is taken into consideration.

The PDE system governing the option value is demonstrated in following proposition.

Proposition 3.3.1. If the underlying asset follows the dynamic processes (3.2.5) under the risk-neutral measure $\mathbb{Q}$ and the value of a European call option written on the underlying asset at time t is defined as

$$u(x, S, t) = \mathbb{E}[e^{-r(T-t)}h(S_T)|S_t = S, x_t = x],$$

where the payoff function is $h(S) = (S - K)^+$ and the expectation $\mathbb{E}$ is under the risk-neutral measure $\mathbb{Q}$, then $u(x, S, t)$ is governed by the following PDE system

$$\begin{cases} -\dfrac{\partial u}{\partial t} = (\mathcal{L}_1 + \mathcal{L}_2)u, \\ u(x, S, T) = (S - K)^+, \qquad (x, S, t) \in \mathbf{R} \times [0, \infty) \times [0, T], \end{cases} \tag{3.3.1}$$

where the diffusion operator $\mathcal{L}_1$ and the jump operator $\mathcal{L}_2$ are defined as

$$\begin{cases} \mathcal{L}_1 u = \dfrac{\kappa^2 + \beta^2\sigma^2}{2}\dfrac{\partial^2 u}{\partial x^2} + \dfrac{1}{2}\sigma^2 S^2 \dfrac{\partial^2 u}{\partial S^2} + \beta\sigma^2 S \dfrac{\partial^2 u}{\partial x \partial S} + [\alpha(\overline{x} - x) + \beta r]\dfrac{\partial u}{\partial x} + rS\dfrac{\partial u}{\partial S} - ru, \\ \mathcal{L}_2 u = e^x[u(x - \beta\gamma, S(1-\gamma), t) - u(x, S, t)]. \end{cases} \tag{3.3.2}$$

The details of the proof are left in Appendix A.1.

Remark 3.3.1. It should be pointed out that the operator $\mathcal{L}_2$ would be in a form of integration if the jump size γ was a random variable. In that case, Equation (3.3.1) would become an integro-PDE. However, in the A&L model, γ is a deterministic constant and Equation (3.3.1) is still a PDE, instead of an integro-PDE.

The boundary conditions along the stock price direction are very easy to impose. They are similar to those in the standard Black-Scholes model. The stock price stays at zero once it hits zero. In this case, the call option becomes worthless even if there is a long time to expiry. Hence we have $u(x, 0, t) = 0$.

On the other hand, as the stock price becomes large, it is more likely that the call option will be exercised at the expiry. The corresponding boundary condition is imposed as:

$$\lim_{S\to\infty} \frac{u(x, S, t)}{S} = 1.$$

Then we turn to the boundary conditions along the buy-in rate direction. When x tends to $-\infty$ (i.e.$\lambda \to 0$), there is no jump in the stock price. In this case, the model is equivalent to the Black-Scholes model. Therefore, we set the value of the call option as the counterpart in the Black-Scholes model, i.e.

$$\lim_{x\to-\infty} u(x, S, t) = \lim_{\lambda\to 0} u(S, \lambda, t) = C^{BS}(S, K, T - t, r, \sigma).$$

Finally, we come to the boundary condition on $x \to \infty$. One needs to understand how the buy-in rate affects the option price first. Roughly speaking, the buy-in rate is a measure of the frequency of buy-ins. When the buy-in rate increases, the buy-ins occur more often, resulting

in higher lending fees. On the other hand, the holder of a European call option needs to hedge the risk by shorting stocks. If the lending fees are very large, hedging the call options would become very expensive. Fewer and fewer investors would buy the European call options as the buy-in rate increases. When the buy-in rate $x = \ln(\lambda)$ becomes large enough, it has few effect on the option value. In other words, the option value would be expected to be insensitive to the buy-in rate change when x has been very large, which is similar to the boundary condition imposed by Clarke & Parrott (1999). Accordingly, a Neumann boundary condition can be given as

$$\lim_{x\to\infty} \frac{\partial u}{\partial x}(x, S, t) = 0.$$

In a brief summary, the properly-closed PDE system governing the value of European call options under the A&L model can be written as:

$$\begin{cases} -\dfrac{\partial u}{\partial t} = (\mathcal{L}_1 + \mathcal{L}_2)u, \\ u(x, S, T) = (S - K)^+, \\ \\ u(x, 0, t) = 0, \\ \lim\limits_{S\to\infty} \dfrac{u(x, S, t)}{S} = 1, \\ \lim\limits_{x\to-\infty} u(x, S, t) = C^{BS}(S, K, T - t, r, \sigma), \\ \lim\limits_{x\to\infty} \dfrac{\partial u}{\partial x}(x, S, t) = 0, \end{cases} \tag{3.3.3}$$

for $(x, S, t) \in \mathbf{R} \times [0, \infty) \times [0, T]$, where the operator $\mathcal{L}_1$ and $\mathcal{L}_2$ are defined as (3.3.2).

It should be noted that the above PDE system is established without the independence assumption. Therefore, the results obtained from the PDE system (3.3.3) cannot be directly compared with those calculated from the semi-explicit pricing formula where the independence assumption has been made. Actually, our PDE approach can also recover the case where the independence assumption is taken into consideration. Similar to Proposition 3.3.1, a simplified PDE system can be obtained as

$$-\frac{\partial u}{\partial t} = (\widetilde{\mathcal{L}}_1 + \mathcal{L}_2)u, \tag{3.3.4}$$

where

$$\widetilde{\mathcal{L}}_1 u = \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 u}{\partial S^2} + \frac{\kappa^2}{2}\frac{\partial^2 u}{\partial x^2} + [\alpha(\overline{x} - x) + \beta r]\frac{\partial u}{\partial x} + rS\frac{\partial u}{\partial S} - ru, \tag{3.3.5}$$

with all boundary conditions remaining the same as those in the PDE system (3.3.3). The results calculated from the PDE (3.3.4) now can be used to verify those computed from the pricing formula for the independence assumption has also been taken into account.

It should also be noted that in our PDE approach, the only difference, with or without this independence assumption, lies in the coefficients of the term $\frac{\partial^2 u}{\partial x^2}$ and the term $\frac{\partial^2 u}{\partial x \partial S}$. The coefficient of term $\frac{\partial^2 u}{\partial x^2}$ changes from $\frac{\kappa^2+\beta^2\sigma^2}{2}$ to $\frac{\kappa^2}{2}$ and the coefficient of term $\frac{\partial^2 u}{\partial x \partial x}$ changes from $\beta\sigma^2 S$ to 0 when the assumption of is taken into consideration. Obviously, it makes perfect sense to make such an assumption when β is small, which indicates that the buy-in rate is not strongly affected by the change of stock price. However, when the coupling parameter β is sufficiently large, the independence assumption is unacceptable and the PDE system (3.3.3) needs to be solved directly instead of the semi-explicit pricing formula. Unlike the semi-explicit pricing formula (3.2.8), which depends heavily on the independence assumption, our PDE approach can deal with option pricing problem successfully whether the independence assumption is made or not. In addition, our PDE approach can also be extended to the American case while it is difficult to extend the semi-explicit pricing formula.

3.4 Numerical schemes

In this section, two numerical schemes are presented to solve the PDE system (3.3.3). The PDE system (3.3.4), which is actually is a special case of the PDE system (3.3.3), can also be solved with these two numerical schemes.

Upon establishing the properly-closed PDE system (3.3.3), it is observed that the operator in the PDE is split into two terms, a diffusion term and a jump term. To solve the PDE system numerically, two approaches to the jump term are presented. One is to approximate the jump term with bilinear interpolation and the other is to adopt a second-order Taylor expansion to estimate it. Based on the different approaches to the jump term, two numerical methods for the

PDE system are presented, Method 1 and Method 2. In order to implement the semi-explicit pricing formula, the weight functions need to be estimated through Monte Carlo simulations. The scheme of Monte Carlo simulation is also provided in this section.

3.4.1 Numerical scheme for the PDE system

For the convenience of numerical implementation, by introducing transforms $\tau = T - t, y = \ln(S), u(x, y, \tau) = u(x, S, t)$, we obtain the PDE system expressed as :

$$\begin{cases} \dfrac{\partial u}{\partial \tau} = \overline{\mathcal{L}_1} u + \overline{\mathcal{L}_2} u, \\ u(x, y, 0) = (e^y - K)^+, \\ \lim\limits_{y \to -\infty} u(x, y, \tau) = 0, \\ \lim\limits_{y \to \infty} \dfrac{u(x, y, \tau)}{e^y} = 1, \\ \lim\limits_{x \to -\infty} u(x, y, \tau) = C^{BS}(e^y, K, \tau, r, \sigma), \\ \lim\limits_{x \to \infty} \dfrac{\partial u}{\partial x}(x, y, \tau) = 0, \end{cases} \tag{3.4.1}$$

for $(x, y, \tau) \in \mathbf{R} \times \mathbf{R} \times [0, T]$, where

$$\begin{aligned} \overline{\mathcal{L}_1} u &= \frac{\kappa^2 + \beta^2\sigma^2}{2} \frac{\partial^2 u}{\partial x^2} + \frac{1}{2}\sigma^2 \frac{\partial^2 u}{\partial y^2} + \beta\sigma^2 \frac{\partial^2 u}{\partial x \partial y} + [\alpha(\overline{x} - x) + \beta r] \frac{\partial u}{\partial x} + (r - \frac{1}{2}\sigma^2) \frac{\partial u}{\partial y} - ru, \\ \overline{\mathcal{L}_2} u &= e^x [u(x - \beta\gamma, y + \ln(1 - \gamma), \tau) - u]. \end{aligned}$$

The domain is truncated as

$$(x, y, \tau) \in [X_{\min}, X_{\max}] \times [Y_{\min}, Y_{\max}] \times [0, T]. \tag{3.4.2}$$

Theoretically, to eliminate the boundary effect, $X_{\max}(Y_{\max})$ should be sufficiently large and $X_{\min}(Y_{\min})$ should be sufficiently small. However, truncation is necessary if we want to adopt the finite difference method. According to Wilmott et al. (1995), the upper bound of stock price is always three or four times the strike price. Therefore we set $Y_{\max} = \ln(5K)$ and $Y_{\min} = -Y_{\max}$

so that $S_{\min} = \frac{1}{5K} \approx 0$. As for the buy-in rate, we set $\lambda_{\max} = 252$, which means that buy-ins occur every day at most. Therefore, $X_{\min} = -X_{\max} = -\ln(252)$ so that $\lambda_{\min} = \frac{1}{252} \approx 0$. The space (x, y, τ) is divided into a uniform grid with

$$\begin{aligned} x_i &= X_{\min} + (i-1) \cdot \Delta x, i = 1, \cdots, N_x; \\ y_j &= Y_{\min} + (j-1) \cdot \Delta y, j = 1, \cdots, N_y; \\ \tau_l &= (l-1) \cdot \Delta \tau, l = 1, \cdots, N_T; \end{aligned}$$

where N_x, N_y, N_T are the number of grid points in the x direction x, y, τ, respectively. The option values at the grid points thus are $u^l_{i,j} = u(x_i, y_j, \tau_l)$.

After meshing the computational domain, we provide two different approaches to the jump term $\overline{\mathcal{L}_2 u}$ in PDE (3.4.1). They are marked as Method 1 and Method 2, respectively.

Method 1

Upon discretizing the computational domain, the discretization associated with the jump term requires values of the unknown functions being evaluated off grid points. In Method 1, we apply a bilinear interpolation to approximate $u(x - \beta\gamma, y + \ln(1-\gamma), \tau)$. For convenience of notation, the value on point P is denoted as $u_P(x, y)$, while the grid points around point P are denoted as $Q_{1,1} = (x_1, y_1), Q_{1,2} = (x_1, y_2), Q_{2,1} = (x_2, y_1), Q_{2,2} = (x_2, y_2)$ as shown in Figure 3.1.

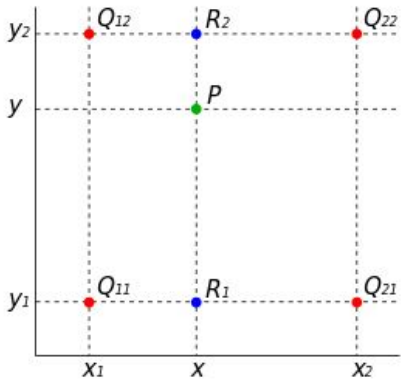


Figure 3.1: Bilinear interpolation

According to the bilinear interpolation method, the value on point P can be approximated as

$$u_P(x,y) \approx \frac{1}{(x_2-x_1)(y_2-y_1)}(x_2-x, x-x_1)\begin{pmatrix} u_{Q_{1,1}} & u_{Q_{1,2}} \\ u_{Q_{2,1}} & u_{Q_{2,2}} \end{pmatrix}\begin{pmatrix} y_2-y \\ y-y_1 \end{pmatrix}, \tag{3.4.3}$$

where $u_{Q_{i,j}}$ are the values on the grid points $Q_{i,j}, i,j = 1,2$. The jump term in the n-th time step is approximated as:

$$\overline{\mathcal{L}_2}u^n \approx e^x[u_P^n(x-\beta\gamma, y+\ln(1-\gamma)) - u^n]. \tag{3.4.4}$$

The Alternative Direction Implicit (ADI) scheme is applied to the operator $\overline{\mathcal{L}_1}$ and an explicit scheme to the operator $\overline{\mathcal{L}_2}$. When discretizing the operator $\overline{\mathcal{L}_1}$ in first step, only the derivatives with respect to x are evaluated in terms of the unknown values of u^{2n+1}, while the other derivatives are replaced in terms of known values of u^{2n}. The difference equation obtained in first step is implicit in the x-direction and explicit in y-direction. This procedure is then repeated at next step with the difference equation implicit in the y-direction and explicit in the x-direction. As for the operator $\overline{\mathcal{L}_2}$ and the cross-derivative $\dfrac{\partial u^2}{\partial x \partial y}$, both steps are explicit. Thus, we obtain two difference equations :

$$\begin{aligned}
\frac{u_{i,j}^{2n+1}-u_{i,j}^{2n}}{\Delta\tau} &= a\frac{u_{i+1,j}^{2n+1}-2u_{i,j}^{2n+1}+u_{i-1,j}^{2n+1}}{\Delta x^2} + c_i\frac{u_{i+1,j}^{2n+1}-u_{i-1,j}^{2n+1}}{2\Delta x} - \frac{r}{2}u_{i,j}^{2n+1} \\
&\quad + b\frac{u_{i,j+1}^{2n}-2u_{i,j}^{2n}+v_{i,j-1}^{2n}}{\Delta y^2} + d\frac{u_{i,j+1}^{2n}-u_{i,j-1}^{2n}}{2\Delta y} - \frac{r}{2}u_{i,j}^{2n} \\
&\quad + \rho\frac{u_{i+1,j+1}^{2n}-u_{i-1,j+1}^{2n}-u_{i+1,j-1}^{2n}+u_{i-1,j-1}^{2n}}{4\Delta x\Delta y} + \overline{\mathcal{L}_2}u_{i,j}^{2n},
\end{aligned} \tag{3.4.5}$$

$$\begin{aligned}
\frac{u_{i,j}^{2n+2}-u_{i,j}^{2n+1}}{\Delta\tau} &= b\frac{u_{i,j+1}^{2n+2}-2u_{i,j}^{2n+2}+u_{i,j-1}^{2n+2}}{\Delta y^2} + d\frac{u_{i,j+1}^{2n+2}-u_{i,j-1}^{2n+2}}{2\Delta y} - \frac{r}{2}u_{i,j}^{2n+2} \\
&\quad + a\frac{u_{i+1,j}^{2n+1}-2u_{i,j}^{2n+1}+u_{i-1,j}^{2n+1}}{\Delta x^2} + c_i\frac{u_{i+1,j}^{2n+1}-u_{i-1,j}^{2n+1}}{2\Delta x} - \frac{r}{2}u_{i,j}^{2n+1} \\
&\quad + \rho\frac{u_{i+1,j+1}^{2n+1}-u_{i-1,j+1}^{2n+1}-u_{i+1,j-1}^{2n+1}+u_{i-1,j-1}^{2n+1}}{4\Delta x\Delta y} + \overline{\mathcal{L}_2}u_{i,j}^{2n+1},
\end{aligned} \tag{3.4.6}$$

where $a = \frac{\kappa^2+\beta^2\sigma^2}{2}, b = \frac{\sigma^2}{2}, c_i = \alpha(\overline{x} - x_i) + \beta r, d = r - \frac{\sigma^2}{2}, \rho = \beta\sigma^2$. The corresponding matrix form for the above equations can be simply written as

$$\mathbf{H_1} u_j^{2n+1} = P_j^{2n} + \mathbf{xBnd}_j^1, \tag{3.4.7}$$

$$\mathbf{H_2} u_i^{2n+2} = Q_i^{2n+1} + \mathbf{yBnd}_i^1, \tag{3.4.8}$$

with the details of $\mathbf{H_1}$, $\mathbf{H_1}$, P_j, Q_i, $\mathbf{xBnd}^1$ and $\mathbf{yBnd}^1$ being defined in Appendix A.3. Note that the matrix $\mathbf{H_1}$ and $\mathbf{H_1}$ are both tridiagonal, so the Thomas algorithm can be adopted to accelerate the computational speed (Strikwerda 2004).

Method 2

The other method to deal with the jump term $\overline{\mathcal{L}_2 u}$ is a second-order Taylor expansion. After applying Taylor expansion, we have

$$\begin{aligned}
\overline{\mathcal{L}_2 u} &= e^x[u(x - \beta\gamma, y + \ln(1-\gamma), t) - u(x, y, t)] \\
&= e^x[-\beta\gamma\frac{\partial u}{\partial x} + \ln(1-\gamma)\frac{\partial u}{\partial y} + \frac{\beta^2\gamma^2}{2}\frac{\partial^2 u}{\partial x^2} + \frac{\ln^2(1-\gamma)}{2}\frac{\partial^2 u}{\partial y^2} - \beta\gamma\ln(1-\gamma)\frac{\partial^2 u}{\partial x \partial y}] \\
&\quad + O(\beta^2\gamma^2 + \ln^2(1-\gamma)).
\end{aligned}$$

With the high-order terms being dropped out, the PDE becomes

$$\frac{\partial u}{\partial \tau} = a\frac{\partial^2 u}{\partial x^2} + b\frac{\partial^2 u}{\partial y^2} + c\frac{\partial u}{\partial x} + d\frac{\partial u}{\partial y} + \rho\frac{\partial^2 u}{\partial x \partial y} - ru, \tag{3.4.9}$$

with new coefficients:

$$a = \frac{\kappa^2 + \beta^2\sigma^2 + \beta^2\gamma^2 e^x}{2}, b = \frac{\sigma^2 + \ln^2(1-\gamma)e^x}{2},$$

$$c = \alpha(\overline{x} - x) + \beta r - e^x\beta\gamma, d = r - \frac{\sigma^2}{2} + e^x\ln(1-\gamma), \rho = \beta\sigma^2 - \ln(1-\gamma)\beta\gamma e^x.$$

Since the PDE (3.4.9) does not include the jump term any longer, the ADI scheme can be directly applied. The finite difference equation for PDE system (3.4.9) is of the form :

$$(I - \theta A_1)(I - \theta A_2)u^{n+1} = [I + A_0 + (1-\theta)A_1 + (1-\theta)A_2 + \theta^2 A_1 A_2]u^n. \tag{3.4.10}$$

The definition of these operators A_0, A_1, A_2 and the details of the derivation for Equation (3.4.10) are left in Appendix A.2.

The simplest ADI scheme, Douglas-Rachford (DR) method (Douglas & Rachford 1956), is applied to solve the finite difference equation (3.4.10). The DR method involves two steps, in which the original operator in (3.4.10) is split into two that are applied in two spatial directions, respectively. First, we calculate an intermediate variable Z from

$$(I - \theta A_1)Z = [I + A_0 + (1-\theta)A_1 + A_2]u^n, \tag{3.4.11}$$

with values in the y direction fixed. After obtaining the intermediate variable Z, the second step is to calculate u^{n+1} from

$$(I - \theta A_2)u^{n+1} = Z - \theta A_2 u^n, \tag{3.4.12}$$

by fixing the value in x direction. The finite difference equation (3.4.10) in two directions now is decomposed into two separate one, Equation (3.4.11) and (3.4.12). The former is calculated implicitly only in x direction; while the latter is computed implicitly in y direction only. As a result, (3.4.11) and (3.4.12) form a classic ADI scheme. The corresponding matrix form for (3.4.11) and (3.4.12) are simply represented as :

$$AZ_j = P_j + \mathbf{xBnd}_j^2, \tag{3.4.13}$$

$$Bu_i = Q_i + \mathbf{yBnd}_i^2, \tag{3.4.14}$$

the $A, P_j, B, Q_i, \mathbf{xBnd}^2$ and $\mathbf{yBnd}_i^2$ are defined in Appendix A.3.

The von Neumann stability analysis is restricted to the PDE with constant coefficients in general. It is extended to the PDE with variable coefficients with the frozen coefficient technique (Zhu & Chen 2011). By fixing the coefficients at their values attained at each gird point in the computational domain, the variable coefficients problem becomes constant coefficient ones. If each frozen coefficient problem is stable, then the variable coefficient problem is also stable (Strikwerda 2004). The next proposition demonstrate the stability of Method 2.

Proposition 3.4.1. When $\theta \geq \frac{1}{2}$, the DR scheme (3.4.11) and (3.4.12) for the PDE (3.4.9) is unconditionally stable in von Neumann sense.

The proof is left in Appendix A.4.

3.4.2 The Monte Carlo simulation

Although A&L obtained an approximate semi-explicit pricing formula with the independence assumption, their formula still involves the calculation of a series of weight functions. In fact, it is for this reason that the pricing formula is called a semi-explicit one. In this subsection, we demonstrate how to implement a Monte Carlo scheme to approximate the weight functions.

Recall the definition of the weight functions,

$$\Pi(n,T) \quad = \quad P(\int_0^T dN_{\lambda_t} = n) = \mathbb{E}\Big\{e^{-\int_0^T \lambda(t,\omega)dt} \frac{[\int_0^T \lambda(t,\omega)dt]^n}{n!}\Big\}. \tag{3.4.15}$$

If the intensity λ_t of Poisson process is a constant or a deterministic function, the weight functions can be calculated through the above formula directly. When the intensity is described by a stochastic process, we have to approximate the probability by frequency over a large number of paths. Here, we present a scheme of Monte Carlo simulation for the Poisson process whose intensity is driven by the stochastic process:

$$dx_t = [\alpha(\bar{x} - x_t) + \beta r]dt + \kappa d\tilde{Z}_t - \beta\gamma dN_{\lambda_t} \quad x_t = \ln(\lambda_t). \tag{3.4.16}$$

In our scheme, the time axis in the domain [0,T] is first discretized into a set of discrete

nodal points, $0 = t_0 < t_1 < \cdots < t_L = T$, with $\Delta t = t_{i+1} - t_i = \frac{T}{L}$ and L being the total number of the nodes. All the values of λ_t or x_t on the nodes $\{t_i\}_{i=1}^{L}$ need to be obtained via Monte Carlo simulations. Let us take the calculation of λ_{t_1} or x_{t_1} as an example to illustrate the simulation process. The pseudocode is presented in Algorithm 1. Once obtaining the value of x_{t_1}, we can move on to next node x_{t_2} and repeat Algorithm 1. Similarly, we can simulate a path for x_t and obtain the total number of jumps for each path.

Algorithm 1 Monte Carlo simulation for SDE (3.4.16)

Set $\lambda_0 = e^{x_0}$, $k = 0$;
Generate a random number τ from exponential distribution with parameter λ_0;
$k = k + 1$ and $\tau_1 = \tau$;
if $\tau_1 > t_1$ **then**
  Generate a random number $\tilde{Z}$ from normal distribution $\mathcal{N}(0, t_1)$;
  Set $x_{t_1} = x_{t_0} + [\alpha(\overline{x} - x_{t_0}) + \beta r](t_1 - t_0) + \kappa\tilde{Z}$;
else if $\tau_1 < t_1$ **then**
  repeat
    Generate a random number τ from exponential distribution with parameter λ_{k-1};
    $\tau_k = \tau_{\tau_{k-1}} + \tau$ and $k = k + 1$;
    Generate a random number $\tilde{Z}$ from normal distribution $\mathcal{N}(0, \tau)$;
    $x_{\tau_k} = x_{\tau_{k-1}} + [\alpha(\overline{x} - x_{\tau_{k-1}}) + \beta r](\tau_k - \tau_{k-1}) + \kappa\tilde{Z} - \beta\gamma$;
  until for some k, $\tau_k > t_1$
  Generate a random number $\tilde{Z}$ from normal distribution $\mathcal{N}(0, t_1 - \tau_{k-1})$;
  Set $x_{t_1} = x_{t_{\tau_{k-1}}} + [\alpha(\overline{x} - x_{\tau_{k-1}}) + \beta r](t_1 - \tau_{k-1}) + \kappa\tilde{Z}$;
end if

Upon simulating paths, we have a summary of the number of jumps for each path. Using these data, we can approximate the weight functions as follows:

$$\Pi(n, T) \approx \frac{\text{Number of paths with } n \text{ jumps}}{\text{Number of total paths}}. \tag{3.4.17}$$

The remaining calculation is simple and straightforward by substituting the estimated weight functions (3.4.17) into the formula (3.2.8) to obtain the option price. When Monte Carlo simulations are implemented, the standard deviation and confidence interval of the option value can also been calculated. The numerical results are reported in next section.

3.5 Numerical results and some discussions

This section consists of three subsections. In first subsection, the Monte Carlo scheme is implemented to estimate the weight functions which are involved in the semi-explicit pricing formula. In the second subsection, we first demonstrate the convergence of the two different numerical schemes for the PDE approach. Then comparisons are made between the numerical results obtained from the semi-explicit pricing formula and those from our PDE method. Some discussions are presented in the last subsection when the independence assumption is unacceptable.

3.5.1 The implementation of the semi-explicit pricing formula

In order to produce numerical results from the semi-explicit formula, a series of weight functions $\{\Pi(n,T)\}_{n\geq 0}$ need to be computed first. This is achieved through Monte Carlo simulations. A simulated path for λ_t is shown in Figure 3.2. When a buy-in occurs, the intensity or buy-in rate drops down to a low level. With time moving forward, it accumulates gradually to an average level.

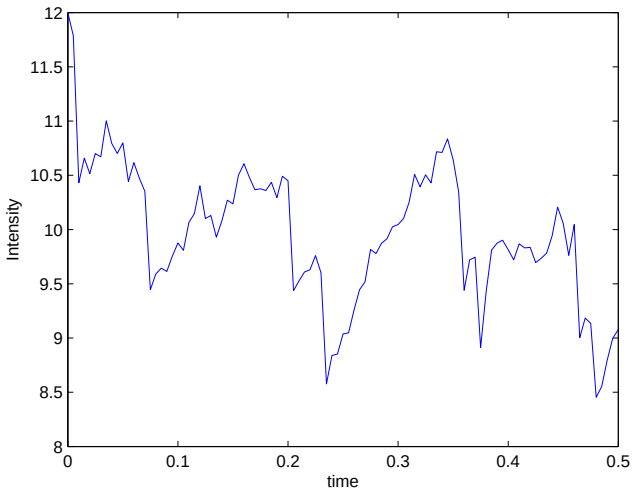


Figure 3.2: A simulated path for λ_t with $T = 0.5, \alpha = 2, \kappa = 0.2, \overline{x} = \ln(10), x_0 = \ln(12), \gamma = 0.01, \beta = 1$.

The intensity x_t or λ_t is simulated Q times consecutively starting with a lower $Q = 100$ until the convergency has been achieved when Q reaches $1,000,000$. To demonstrate the convergency

of the calculated $\Pi(n,T)$, the variation of values as a function of the number of Monte Carlo simulation paths is tabulated in Table 3.1 for $n = 6,7,8,9,10$. Clearly, for all these five functions, an accuracy at the 3rd decimal place has been achieved when the number of Monte Carlo simulations has reached $1,000,000$. Therefore, the results obtained when $Q = 1,000,000$ are adopted to carry out the subsequent calculations of the option price using the semi-explicit pricing formula. These weight functions listed in Table 3.1 are just some examples of the series of weight functions. All of them can be computed with the approximation (3.4.17).

Q	$\Pi(6,T)$	$\Pi(7,T)$	$\Pi(8,T)$	$\Pi(9,T)$	$\Pi(10,T)$
100	0.16	0.11	0.09	0.07	0.05
1,000	0.162	0.114	0.081	0.045	0.032
10,000	0.1609	0.1277	0.083	0.043	0.0306
100,000	0.1598	0.1241	0.0829	0.0497	0.0268
1,000,000	0.1593	0.1228	0.0825	0.0490	0.0262

Table 3.1: Convergence of the weight functions. Model parameters are $T = 0.5, \sigma = 0.45, \alpha = 2, \kappa = 0.2, \overline{x} = \ln(10), x_0 = \ln(12), \gamma = 0.01, \beta = 1$.

The subsequently calculated option values from the semi-explicit pricing formula and the standard derivations for our Monte Carlo Simulations are both tabulated in Table 3.2. The confidence interval of option values can also be reported. For example, a 95% confidence interval of the option value at $S_0 = 10$ is $(1.07647, 1.07693)$. In other words, we are 95% confident that the option value at $S_0 = 10$ lies between 1.07647 and 1.07693.

stock price	$S_0 = 8$	$S_0 = 9$	$S_0 = 10$	$S_0 = 11$	$S_0 = 12$
option value	0.3299	0.6405	1.0767	1.6302	2.2842
standard deviation	0.00045	0.00079	0.00117	0.00157	0.00197

Table 3.2: The value calculated from pricing formula. Model parameters are $K = 10, r = 0.05, T = 0.5, \sigma = 0.45, \alpha = 2, \kappa = 0.2, \overline{x} = \ln(10), x_0 = \ln(12), \gamma = 0.01, \beta = 1$.

3.5.2 Numerical results for the PDE approach

Option values from the semi-explicit pricing formula have been obtained through Monte Carlo simulations. In this subsection, we reported the numerical results for the PDE system (3.3.4) where the independence assumption has been adopted so that they can be compared with those from the semi-explicit pricing formula.

Both Method 1 and Method 2 are implemented to solve the PDE system (3.3.4) on different girds so that in order to observe the convergence of the computed results as grid space becomes fine. To measure the convergence, the l_2 error between the value on a fine grid and the coarse grid at the previous level is also listed in Table 3.3. As shown in Table 3.3, the l_2 error decreases both in Method 1 and Method 2 as the size of grid spacing diminishes, indicating a significant convergence of both methods.

Method	(N_x, N_y, N_T)	$S_0 = 8$	$S_0 = 9$	$S_0 = 10$	$S_0 = 11$	$S_0 = 12$	l_2 error
Method 1	(20,80,100)	0.3642	0.6829	1.1256	1.6864	2.3471	
	(40,160,200)	0.3425	0.6560	1.0967	1.6518	2.3040	0.0712
	(80,320,300)	0.3342	0.6456	1.0834	1.6372	2.2913	0.0270
	(160,640,400)	0.3304	0.6408	1.0770	1.6303	2.2840	0.0134
	(320,1280,500)	0.3299	0.6403	1.0762	1.6294	2.2831	0.0017
Method 2 $\theta = 0.5$	(20,40,40)	0.3127	0.6490	1.0144	1.6409	2.2583	
	(40,80,50)	0.3384	0.6488	1.0846	1.6401	2.2965	0.0840
	(80,160,60)	0.3313	0.6409	1.0781	1.6311	2.2824	0.0208
	(160,320,70)	0.3301	0.6401	1.0766	1.6298	2.2834	0.0027
	(320,640,80)	0.3298	0.6400	1.0761	1.6293	2.2830	0.0001

Table 3.3: Comparison of values calculated with different grids. Model parameters are $K = 10, r = 0.05, T = 0.5, \sigma = 0.45, \alpha = 2, \kappa = 0.2, \overline{x} = \ln(10), x_0 = \ln(12), \gamma = 0.01, \beta = 1$.

To demonstrate the convergence rate of our numerical methods, the ratio of consecutive l_2 error is listed in Table 3.4 as the grid is refined by a factor of two in x and y directions. The ratio of Method 2 is obviously higher than that of Method 1, which implies that the explicit treatment of the jump term in Method 1 really slows down the convergence.

Methods	(N_x, N_y, N_T)	l_2 error	ratio
Method 1	(40,160,200)	0.0712	
	(80,320,300)	0.0270	2.7
	(160,640,400)	0.0134	2.0
Method 2 $\theta = 0.5$	(20,40,40)	0.0840	
	(40,80,50)	0.0208	4.1
	(80,160,60)	0.0027	7.7

Table 3.4: Ratio of convergence

On the other hand, Method 1 requires much finer grid than Method 2 and thus more time in order to reach almost the same level of convergence. As far as the total computational time associated with each method is concerned, the total consumed CPU time for a particular run is

adopted to measure computational efficiency, which is complemented by a measure of relative error defined as

$$\text{Relative Error} = \frac{\| V_{\text{PDE}} - V_{\text{A\&L}} \|_2}{\| V_{\text{A\&L}} \|_2},$$

where V_{PDE} is the option value obtained from the PDE system (3.3.4) with Method 1 or Method 2 and $\text{V}_{\text{A\&L}}$ is the value calculated from the semi-explicit formula. Recorded in Table 3.5 are the CPU time and the relative error.

Methods	(N_x, N_y, N_T)	CPU times (s)	Relative Error(%)
Method 1	(20,80,100)	0.229	3.87
	(40,160,200)	1.386	1.56
	(80,320,300)	9.102	0.69
	(160,640,400)	101.7	0.26
	(320,1280,500)	1130	0.20
Method 2 $\theta = 0.5$	(20,40,40)	0.239	2.18
	(40,80,50)	0.461	0.93
	(80,160,60)	1.53	0.27
	(160,320,70)	8.17	0.22
	(320,640,80)	84.9	0.19
A&L	$Q = 1,000,000$	250.1	

Table 3.5: Comparison of CPU time and relative error. Model parameters are $K = 10, r = 0.05, T = 0.5, \sigma = 0.45, \alpha = 2, \kappa = 0.2, \overline{x} = \ln(10), x_0 = \ln(12), \gamma = 0.01, \beta = 1$.

From Table 3.5, we clearly observe that the relative error decreases close to zero as the size of grid is diminished, which indicates that the numerical results from the PDE system (3.3.4) are consistent with those from the semi-explicit pricing formula. Furthermore, Method 1 is more time-consuming than Method 2, in order to reach almost the same level of accuracy. This is associated with the adopted numerical scheme. Method 1 is a modified version of implicit-explicit (IMEX) scheme for the PDE, which applies an ADI scheme to the operator $\overline{\mathcal{L}_1}$ and an explicit scheme for the operator $\overline{\mathcal{L}_2}$ and cross derivative term $\frac{\partial^2 u}{\partial x \partial y}$. Such a scheme with hybrid feature is more efficient than a fully explicit scheme because the ADI scheme can speed up the process of convergence. However, the explicit treatment of jump term slows down the total speed of convergence. The efficiency of Method 1 may be improved a bit by implementing an iterative scheme as suggested in d'Halluin et al. (2005). While the computational efficiency is not our focus in this chapter, one can refer to that paper if interested. As for the Method 2, an

ADI scheme is applied after adopting the second-order Taylor expansion. It has been proved that the ADI scheme is unconditionally stable when $\theta \geq \frac{1}{2}$. Therefore, the results of Method 2 exhibit faster convergence than those of Method 1.

However, in Method 2, a second-order Taylor expansion is used to approximate the jump term. Actually, this indicates that we have implicitly assumed that γ is of a small value, so that the high-order terms can be dropped out. This implicit assumption directly affects the accuracy of Method 2. The numerical error of Method 2 includes not only the truncation error led by the finite difference scheme but also the approximation error introduced by adopting the second-order Taylor expansion. The truncation error can be eliminated gradually when the size of girds become small, while the approximation error cannot. Therefore, its accuracy will be significantly affected by the value of γ with large relative error for great γ. As for Method 1, the error source is the truncation error and interpolation error which do not heavily depend on the value of γ.

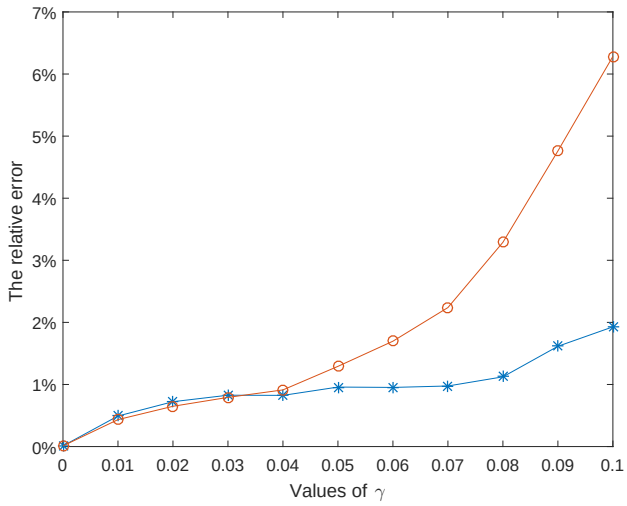


Figure 3.3: The relative error between Method 1 and Method 2. Model parameters are $K = 10, r = 0.05, T = 0.5, \sigma = 0.45, \alpha = 2, \kappa = 0.2, \overline{x} = \ln(10), x_0 = \ln(12), \beta = 1$.

Figure 3.3 demonstrates the relationship between relative error and the value of γ. From this figure, it can be observed that the relative error for the Method 2 increases more rapidly than that of Method 1 as the value of γ becomes large. Although Method 2 shows good convergence rate in Table 3.4 and more efficiency in Table 3.5, Figure 3.3 shows us that Method 1 is more

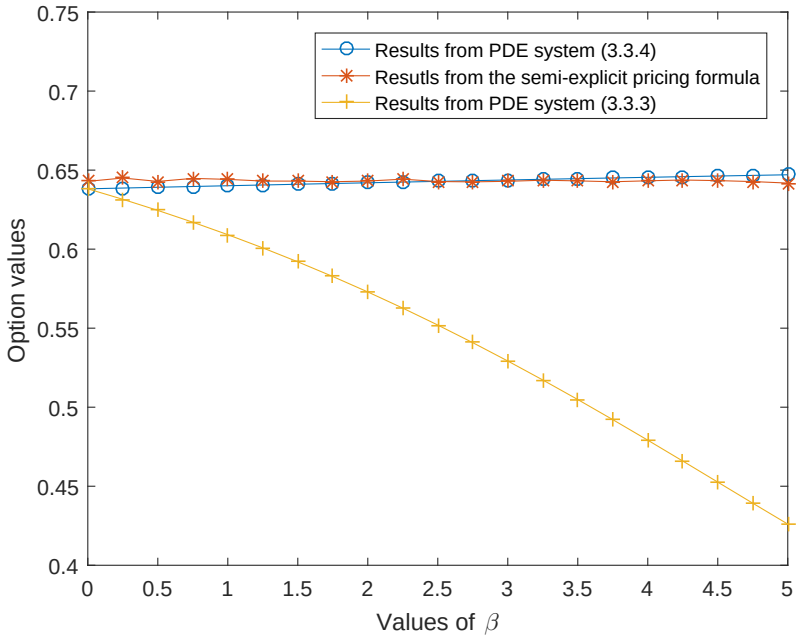
accurate when the value of γ is large.

3.5.3 When the independence assumption is unacceptable

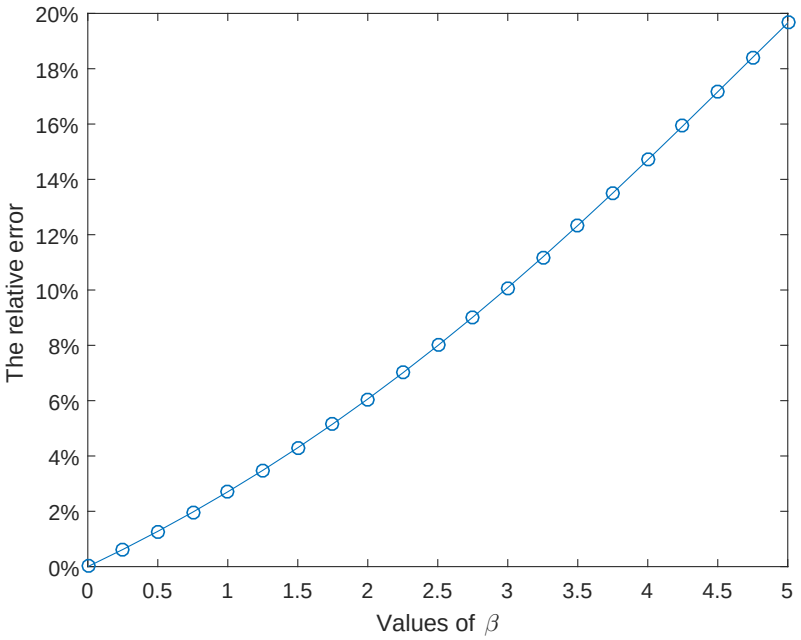
In the subsection above, we have demonstrated that the results from PDE system (3.3.4) and those from the semi-explicit formula are consistent with each, which implies that our PDE approach recovers the special case in which the independence assumption is valid. However, when the coupling parameter β is sufficiently large, which means the buy-in rate depends heavily on the change of stock price, it is unreasonable to make such an independence assumption. The PDE system (3.3.3) without the independence assumption should be solved numerically.

In order to illustrate the difference between the results with the assumption of independence and those without the assumption, the option values obtained from the PDE systems (3.3.3), (3.3.4) and the semi-explicit pricing formula with different values of β are presented in Figure 3.4(a). The relative error between the numerical results calculated from PDE systems (3.3.3) and (3.3.4) is plotted in Figure 3.4(b).

It is observed from Figure 3.4(a) that the option value calculated from the PDE system (3.3.3) varies as the coupling parameter β changes. However, both the results obtained from the PDE (3.3.4) and those from the semi-explicit formula are hardly affected by the coupling parameter β. This indicates that the key independence assumption has almost ignored the role of the coupling parameter β in the A&L model. For this set of parameters, the relative error is acceptable when β is a small number. As the value of β increases gradually, the relative error grows sharply. The relative error has reached 20% when β is around five, which is really a significant error. In fact, an example with $\beta = 30$ was included in Avellaneda & Lipkin (2009). It is obvious that the independence assumption has become unacceptable in that case. Our PDE approach without the independent assumption is more reasonable than the semi-explicit formula for the case where β is of a great value.



(a) Option values



(b) The relative difference

Figure 3.4: The option values calculated from PDE (3.3.3), PDE (3.3.4) and semi-explicit formula. Model parameters are $S = 9, K = 10, r = 0.05, T = 0.5, \sigma = 0.45, \alpha = 2, \kappa = 0.2, \overline{x} = \ln(10), x_0 = \ln(12), \gamma = 0.01$.

3.6 Conclusions

In this chapter, we have applied a PDE approach to price European call options under the A&L model. Two numerical methods are proposed to solve the PDE system based on different approaches to the jump term. We have also numerically realized the semi-explicit pricing formula via Monte Carlo simulations for the buy-in rate.

Our PDE approach is a broader way than the approximate semi-explicit pricing formula. We demonstrate that the results from PDE system (3.3.4) are consistent with those from the semi-explicit formula, which implies that our PDE approach really can recover the special case in which the independence assumption is valid. We also demonstrate that our PDE approach can deal with option pricing problem successfully without the independence assumption; while the semi-explicit pricing formula cannot because its derivation depends heavily on the independence assumption. In addition, the PDE approach presented in this chapter has laid a solid foundation for the study on option pricing of American-style option under the A&L model.

Chapter 4

Pricing American call options under a hard-to-borrow stock model

4.1 Introduction

Option pricing is one of the most important topics in quantitative finance ever since Black & Scholes (1973) proposed an analytical and quantitative formula for pricing European options; their seminal work has laid the foundation of the modern theory of pricing financial derivatives. In the past few decades, a great number of papers have been devoted to the development of methods for pricing options. Along an amazing development of new models as well as research methodology, the Black-Scholes model is now considered to be inadequate to capture the dynamics of stocks traded in open exchanges. One of the most important assumptions in the celebrated Black-Scholes model is that short selling is permitted without any cost, while the market conditions suggest otherwise.

In most developed stock markets, naked short selling is forbidden for it may lead to significant systematic risk. Generally, when an investor wants to short a stock, he has to borrow it from others in advance. The availability of the stock for borrowing depends on market conditions. Some stocks can be easily borrowed without lending fees, a case where the particular assumption about short selling in the Black-Scholes model holds. Others may be in short sup-

ply and the investor has to pay some lending fees to borrow them. In the latter, the stock is called a *hard-to-borrow* one (Avellaneda & Lipkin 2009). The assumption that short selling is permitted without any cost does not hold any more for hard-to-borrow stocks.

Another key concept about short selling is *short interest* defined as the percentage of the float currently held short in the market. As the short interest increases, the risk of fail-to-deliver accumulates in market. Too much fail-to-deliver risk may lead to high systematic risk, which is a disaster for stock markets. To avoid or at least reduce the possible systematic risk, investors in short positions may be "forced", in some extreme events, by the clearing firm to buy back the stock, which is called a *buy-in*. When a buy-in occurs, it introduces an excess demand on the stock which is unmatched by supply at the current price, resulting in a temporary upward impact on the price. Upon finishing such a buy-in, the excess demand disappears quickly, causing the stock price to jump back to the original level. More details about buy-in can be found at the Securities and Exchange Commission's Regulation SHO[1].

In the literature, there is a considerable amount of research about short selling. Diamond & Verrecchia (1987) considered constraints on short selling and asset price adjustment to private information. Nielsen (1989) studied the asset market equilibrium with short selling. Duffie et al. (2002) presented a model of asset valuation in which short selling is achieved by searching for security lenders and bargaining over the terms of lending fee. Jones & Lamont (2002) pointed out that, from market data, stocks are overpriced when short-sale constraints are imposed. Evans et al. (2009) mainly focused on how options market and short selling interact with each other.

Following the previous work, Avellaneda & Lipkin (2009) proposed a new model for hard-to-borrow stocks by introducing a Poisson process to characterize buy-ins. For convenience, the model they proposed is referred to as the A&L model hereafter in this chapter. The A&L model provides a feedback mechanism that couples the dynamics of stock price with the frequency at which buy-ins take place. The frequency is also called the *buy-in rate*. Using their model, Avellaneda and Lipkin also obtained a semi-explicit pricing formula for European call options

[1]The reader can visit the website: www.sec.gov to learn more about Regulation SHO.

with a series of weight functions. Although they did not work out the American case either theoretically or empirically in their paper, they instinctively "forecast" that it is optimal for an American call option to exercise early when the underlying stock is hard-to-borrow. They have also pointed out that pricing American call options under their model could entail a high-dimensional numerical calculation and have left this rather complicated task for others to pursue in the future.

Recently, Jensen & Pedersen (2016) studied how market frictions, such as transaction costs and short-sale costs, result in early exercise of an American call option. Although their empirical study may be motivated by Avellaneda & Lipkin (2009), Jensen and Pedersen still adopted the classic Black-Scholes model with short-sale costs in their paper, instead of the new A&L model for hard-to-borrow stocks. After analyzing the data collected from OptionMetrics database on U.S. option prices and the data of actual exercise behavior from the Options Clearing Corporation, they concluded that early exercise can be optimal for an American call option written on a stock that pays no dividends, taking the market frictions into account. They have actually used a very eye-catching title "never say never" to overturn the classic conclusion from Merton (1973) that, except just before expiration or dividend payments, one should never exercise an American call option.

These pioneer works have highly motivated us to adopt a much more comprehensive model such as the A&L model to examine whether or not it is optimal to exercise an American call option early when the underlying stock is hard-to-borrow. Such a theoretical study under the A&L model would complement the Jensen and Pedersen's study where the Black-Scholes model with short-sale costs was considered. As a preparation for this paper, Ma et al. (2016) first established the PDE system to price European call options under the A&L model. By comparing the results obtained from the PDE system with those from the semi-explicit pricing formula obtained by Avellaneda and Lipkin, the correctness of the PDE system has been demonstrated. We now extend the PDE method, in this chapter, from the European case to the American case.

To price American call options under the A&L model, the properly-closed PDE system with

a free boundary is established first.Then it is reformulated as a linear complementarity problem (LCP), which is numerically solved with the Lagrange multiplier approach (LMA). Through our numerical results, we figure out that there is an early exercise premium of an American call option when the underlying stock follows the A&L model. In other words, early exercise is optimal for an American call option than holding to expiration. We point out that it is the lending fee that leads to the early exercise of American call options under the A&L model. In addition, we also numerically analyze how and to what extent the lending fees affect the optimal exercise price.

The rest of this chapter is organized as follows. In Section 4.2, we review the A&L model and its risk-neutral measure. Then the PDE governing the value of American call options is established with a free boundary. Furthermore, it is reformulated as a linear complementarity problem (LCP). In Section 4.3, the finite difference method is adopted to discretize the PDE system, which leads to a sequence of LCPs. These LCPs are solved numerically with the Lagrange multiplier approach (LMA). In Section 4.4, numerical results are provided to demonstrate the convergence of our schemes first. Then some discussions about the early exercise of an American call option are presented. Conclusions are given in the last section.

4.2 Pricing American call options under the A&L model

For the completeness of the paper and the convenience to the readers, we first briefly review the A&L model, including its corresponding risk-neutral measure. Then the PDE system is established with free boundary. Furthermore, it is reformulated as a linear complimentary problem (LCP), which is numerically solved later.

4.2.1 The A&L model for hard-to-borrow stocks

In most stock markets, an investor has to borrow a stock in advance if he wants to short a stock since naked short selling is generally forbidden. The availability of the stock depends on the market condition: some can be easily borrowed, while others may be in short supply. In the

latter, they are referred to as *hard-to-borrow* stocks. Usually, a significantly intensified short interest indicates that a large amount of fail-to-deliver risk has accumulated for the particular stock in the market. On the other hand, too much fail-to-deliver risk may even result in high systematic risks for the entire market. To avoid the systematic risk, at some extreme occasions, short sellers are demanded by the clearing firm to buy back the stock they have shorten within a given time according to the Securities and Exchange Commission's Regulation SHO. In order to meet the buy-in requirement, short sellers would borrow stocks in market by paying lending fees.

Avellaneda & Lipkin (2009) proposed a dynamic model to describe the price of hard-to-borrow stocks, in which S_t and λ_t denote respectively the price and the buy-in rate at time t. In their model, market dynamics are described by the following stochastic differential equations under the physical measure $\mathbb{P}$:

$$\begin{cases} \dfrac{dS_t}{S_t} = \sigma dW_t + \gamma\lambda_t dt - \gamma dN_{\lambda_t}(t) \\ dx_t = \kappa dZ_t + \alpha(\overline{x} - x_t)dt + \beta\dfrac{dS_t}{S_t}, \quad x_t = \ln(\lambda_t), \end{cases} \tag{4.2.1}$$

where $dN_\lambda(t)$ denotes the increment of a standard Poisson process with intensity λ over the interval $(t, t+dt)$. The parameters σ and γ are respectively the volatility and the price elasticity of demand due to buy-ins; W_t and Z_t are two independent standard Brownian motions which drive the stock price and the buy-in rate respectively. The logarithm of λ_t is modeled as a mean-reversion process with κ being the volatility, $\overline{x}$ the long-time equilibrium value, α the speed of mean-reversion and β the coupling parameter that couples the change in price and the logarithm of buy-in rate. For convenience, x_t, the logarithm of buy-rate λ_t, is still referred to as buy-in rate hereafter.

It should be pointed out that the A&L model operates in an incomplete market since an additional source of uncertainty has been introduced through the buy-in rate, which is not a tradable quantity (Tankov 2003). Therefore, it is impossible to perfectly hedge a portfolio composed of hard-to-borrow stocks and there does not exist a unique risk-neutral measure. For pricing a derivative, a risk-neutral measure needs to be defined for the processes S_t and x_t first.

What Avellaneda and Lipkin did was to introduce an arbitrage-free pricing measure, which is equivalent to changing the drift of the Brownian motion associated with the underlying stock. Mathematically, to conduct measure transform, two new processes are defined as

$$\tilde{W}_t = W_t + \int_0^t \frac{\gamma\lambda_l - r}{\sigma} dl, \tag{4.2.2}$$

and

$$\tilde{Z}_t = Z_t + \int_0^t \frac{\alpha z(l, x_l, S_l)}{\kappa} dl, \tag{4.2.3}$$

where $z(t, x, S)$ is an arbitrary function. By Girsanov's theorem, $\tilde{W}$ and $\tilde{Z}$ are two independent Brownian motions under the risk-neutral measure $\mathbb{Q}$ defined by

$$\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_t = exp\Big\{-\int_0^t \big[\frac{\gamma\lambda_l - r}{\sigma} + \frac{\alpha z(l, x_l, S_l)}{\kappa}\big] dW_l - \frac{1}{2}\int_0^t \big[\frac{(\gamma\lambda_l - r)^2}{\sigma^2} + \frac{\alpha^2 z^2(l, x_l, S_l)}{\kappa^2}\big] dl\Big\}, \tag{4.2.4}$$

which is the so-called Radon-Nikodym derivative that facilitates the change of measure. Under this risk-neutral measure, the dynamics of the A&L model become

$$\begin{cases} \dfrac{dS_t}{S_t} = \sigma d\tilde{W}_t + r dt - \gamma dN_{\lambda_t}(t), \\ dx_t = \kappa d\tilde{Z}_t + [\alpha(\overline{x}^* - x_t)]dt + \beta \dfrac{dS_t}{S_t}, \end{cases} \tag{4.2.5}$$

where $\overline{x}^* = \overline{x} - z$.

Financially, any source of uncertainty needs to be compensated by the associated market price of risk or risk premium. In the classic Black-Scholes model, the market price of risk for the underlying is $\dfrac{\mu - r}{\sigma}$ (Wilmott et al. 1995). On the other hand, in the Heston model, an additional source uncertainty is introduced by the stochastic volatility and an additional market price of volatility risk is defined through an arbitrary function, i.e., $\lambda(t, S, v)$ in Heston (1993), which may appear in a more general form for other stochastic volatility models discussed in Fouque et al. (2000). In the A&L model, the new buy-in process also brings in an additional source of uncertainty and the corresponding market price of buy-in risk is represented by the function $z(t, x, S)$ in Equation (4.2.3). Furthermore, it should be remarked that the market

price of risk for the stock in the A&L model becomes $\frac{\gamma\lambda - r}{\sigma}$, which is different from its counterpart $\frac{\mu - r}{\sigma}$ in the Black-Scholes model.

When a market is complete, the market price of risk for the underlying is unique, such as the term $\frac{\mu - r}{\sigma}$ in the Black-Scholes model. When a market is incomplete, the market price of risk is specified after a risk-neutral measure is chosen, or in a vice versa way in financial practice that a market price of risk is extracted from market data first, which then implicitly dictates the risk-neutral measure to be used in pricing a derivative. Therefore, the market price of buy-in risk in the A&L model should be determined by market data, just as the market price of volatility risk in the Heston model needs to be calibrated from market data (Bollerslev et al. 2011). For simplicity, Avellaneda & Lipkin (2009) effectively set $z(t, x, S)$ to be zero, which is a standard treatment in the Heston model as well (Rouah 2013).

In addition to the A&L model, there are other ways to study how lending fees or short-sale costs affect option price. Recently, Jensen & Pedersen (2016) have explored how market frictions, including transaction costs and short-sale costs, affect the American call option pricing. The model in their paper was still the classic Black-Scholes one with short-sale costs. After analyzing the empirical data collected from OptionMetrics database on U.S. option prices and the data of actual exercise behavior from the Options Clearing Corporation, they have come to a conclusion that market frictions can lead to early exercise of American call options although the underlying stock pays no dividends.

In this chapter, the A&L model, which specially describes the dynamics of hard-to-borrow stocks, is adopted to replace the classic Black-Scholes model in Jensen and Pedersen's paper. We would like to explore whether or not Jensen and Pedersen's conclusion still holds under the A&L model.

4.2.2 The PDE system with boundary conditions

Let $u(t, x, S)$ denote the value of an American call option with S being the price of underlying stock, x being the buy-in rate and t being the time. After changing measure, the dynamics under the risk-neutral measure are shown in Equation (4.2.5). By Feynman-Kac theorem, the

value of a call option $u(t, x, S)$ satisfies the following bivariate PDE

$$\mathcal{L}u := \frac{\partial u}{\partial t} + \mathcal{L}_1 u + \mathcal{L}_2 u = 0, \tag{4.2.6}$$

where

$$\begin{cases} \mathcal{L}_1 u = \dfrac{\kappa^2 + \beta^2\sigma^2}{2}\dfrac{\partial^2 u}{\partial x^2} + \dfrac{1}{2}\sigma^2 S^2 \dfrac{\partial^2 u}{\partial S^2} + \beta\sigma^2 S \dfrac{\partial^2 u}{\partial x \partial S} + [\alpha(\overline{x} - x) + \beta r]\dfrac{\partial u}{\partial x} + rS\dfrac{\partial u}{\partial S} - ru, \\ \mathcal{L}_2 u = e^x[u(t, x - \beta\gamma, S - S\gamma) - u(t, x, S)]. \end{cases} \tag{4.2.7}$$

The operator $\mathcal{L}_1$ corresponds to the continuous part of stock price, while $\mathcal{L}_2$ corresponds to the jump part in the A&L model. The terminal condition for the PDE (4.2.6) is given by the payoff function $g(S) = (S - K)^+$ with K being the strike price.

Remark 4.2.1. It should be pointed out that the operator $\mathcal{L}_2$ would be in a form of integration if the jump size γ was a random variable. In that case, Equation (4.2.6) would become an integro-PDE. However, in the A&L model, γ is a deterministic constant and Equation (4.2.6) is still a PDE, instead of an integro-PDE.

The boundary conditions along the S direction are similar with those in the Black-Scholes model. Once the stock price reaches zero, it will stay at zero forever. Consequently, the American call option would become worthless, i.e. $u(t, x, 0) = 0$, even though there is still a long time to expiration.

On the other hand, just as in the Black-Scholes model, there is an optimal exercise price $S_f(t, x)$, above or equal to which it is optimal to exercise the call option. Conditions along the optimal exercise price S_f are also needed. We assume that option price and its first derivative are continuous across the optimal exercise price because of the so-called *smooth pasting condition* (Wilmott 2013)

$$\begin{cases} u(t, x, S_f) = S_f - K, \\ \dfrac{\partial u}{\partial S}(t, x, S_f) = 1. \end{cases} \tag{4.2.8}$$

It should be noted that the above two conditions are similar to those in the Black-Scholes model. The main difference is that the optimal exercise price S_f in the A&L model depends

not only on time but also on buy-in rate. In other words, the optimal exercise price is a two dimensional function of both t and x .

The boundary conditions in the x direction are similar to those in the European case provided by Ma et al. (2016). When x approaches $-\infty$, it indicates that buy-in never occurs and thus there are no jumps in the stock price. In this case, the A&L model degenerates to the Black-Scholes model. Consequently, we set the value of an American call option as its counterpart in the Black-Scholes model. In addition, under the Black-Scholes model, the value of an American call option is the same as that of a corresponding European one when there are no dividends. Consequently, the boundary condition in this direction can be prescribed as

$$\lim_{x\to-\infty} u(t,x,S) = C^{BS}(S,K,T-t,r,\sigma), \tag{4.2.9}$$

where $C^{BS}(S,K,\tau,r,\sigma)$ is the value of a European call option calculated from the Black-Scholes formula.

On the other hand, the value of an American call option would be insensitive to the change of the buy-in rate when the buy-in rate tends toward infinity, which is similar to the boundary condition imposed by Clarke & Parrott (1999). Consequently, an asymptotic boundary condition is imposed as follows:

$$\lim_{x\to\infty} \frac{\partial u}{\partial x}(t,x,S) = 0. \tag{4.2.10}$$

As a brief summary, the properly-closed PDE system governing the value of an American

call option can be written as:

$$\begin{cases} \mathcal{L}u = 0, \\ u(T,x,S) = (S-K)^+, \\ u(t,x,0) = 0, \\ \lim\limits_{x\to-\infty} u(t,x,S) = C^{BS}(S,T-t,K,r,\sigma), \\ \lim\limits_{x\to\infty} \dfrac{\partial u}{\partial x}(t,x,S) = 0, \\ u(t,x,S_f) = S_f - K, \\ \dfrac{\partial u}{\partial S}(t,x,S_f) = 1, \qquad (t,x,S) \in [0,T]\times\mathbf{R}\times[0,S_f(t,x)]. \end{cases} \tag{4.2.11}$$

The PDE system (4.2.11) is nonlinear because there is an unknown function $S_f(t,x)$ implied by free boundary conditions (4.2.8). Instead of solving such a nonlinear PDE system directly, we would reformulate it as a linear complementarity problem first.

4.2.3 Pricing an American call option as a LCP

The PDE system (4.2.11) with a free boundary S_f can be solved with many methods. One of them is to reformulate it as a linear complementarity problem (LCP), which was first presented by Merton et al. (1977).

In this approach, the pricing domain $\Omega := \{(t,x,S)|(t,x,S) \in [0,T]\times\mathbf{R}\times[0,\infty)\}$ is divided into two subregions by the free boundary S_f: the continuation region Ω^C and the exercise region Ω^E. Financially, such a free boundary is also called optimal exercise price. In the continuation region Ω^C, the PDE (4.2.6) holds and the value of holding the option is greater than that of exercising it right away. As a result, the continuation region can be expressed exactly as

$$\Omega^C = \{(t,x,S)|\mathcal{L}u = 0, u(t,x,S) \geq g(S)\}.$$

On the other hand, in the exercise region Ω^E, the call option should be exercised at the current

price. As a result, the exercise region is mathematically defined as

$$\Omega^E = \{(t, x, S) | u(t, x, S) = g(S)\}.$$

After we combine these two regions together, the value of an American call option satisfies the following linear complementarity problem

$$\begin{cases} \mathcal{L}u(t, x, S) \leq 0, \\ u(t, x, S) \geq g(S), \\ [u(t, x, S) - g(S)]\mathcal{L}u(t, x, S) = 0, \end{cases} \tag{4.2.12}$$

where $(t, x, S) \in \Omega = \Omega^C \cup \Omega^E$.

It is obvious that the free boundary S_f is no longer explicitly involved in the LCP (4.2.12). When solving the LCP (4.2.12) in the pricing domain Ω, we just need the boundary conditions (4.2.13).

$$\begin{cases} u(T, x, S) = (S - K)^+, \\ u(t, x, 0) = 0, \\ \lim\limits_{S\to\infty} \dfrac{u(t, x, S)}{S} = \lim\limits_{S\to\infty} \dfrac{S - K}{S} = 1, \\ \lim\limits_{x\to-\infty} u(t, x, S) = C^{BS}(S, T - t, K, r, \sigma), \\ \lim\limits_{x\to\infty} \dfrac{\partial u}{\partial x}(t, x, S) = 0. \end{cases} \tag{4.2.13}$$

We solve the LCP numerically to obtain the option values $u(t, x, S)$. Then as a part of solution, the free boundary $S_f(t, x)$ can be implicitly recovered from the option values through the free boundary conditions (4.2.8).

4.3 Numerical schemes

In this section, we present the numerical scheme to solve the LCP (4.2.12) obtained above. First of all, space discretization is performed using finite difference method, which leads to a family of LCPs. Then the Lagrange multiplier approach (LMA) proposed by Ito & Kunisch

(2006) is adopted to treat the resulting LCPs numerically.

4.3.1 Spacial discretization

For the convenience of our numerical implementation, we introduce the following transforms $\tau = T - t, y = \ln(S)$. As a result, functions $u(t, x, S)$ and $S_f(t, x)$ now become $u(\tau, x, y)$ and $S_f(\tau, x)$, respectively and the LCP (4.2.12) becomes:

$$\begin{cases} \dfrac{\partial u}{\partial \tau} \geq \overline{\mathcal{L}_1} u + \overline{\mathcal{L}_2} u, \\ u(\tau, x, y) \geq g(y), & (\tau, x, y) \in \bar{\Omega} := [0, T] \times \mathbf{R} \times \mathbf{R}, \\ [u(\tau, x, y) - g(y)](\dfrac{\partial u}{\partial \tau} - \overline{\mathcal{L}_1} u - \overline{\mathcal{L}_2} u) = 0, \end{cases} \tag{4.3.1}$$

where

$$\begin{cases} \overline{\mathcal{L}_1} u = \frac{\kappa^2 + \beta^2 \sigma^2}{2} \frac{\partial^2 u}{\partial x^2} + \frac{1}{2}\sigma^2 \frac{\partial^2 u}{\partial y^2} + \beta \sigma^2 \frac{\partial^2 u}{\partial x \partial y} + [\alpha(\overline{x} - x) + \beta r] \frac{\partial u}{\partial x} + (r - \frac{1}{2}\sigma^2) \frac{\partial u}{\partial y} - ru, \\ \overline{\mathcal{L}_2} u = e^x [u(\tau, x - \beta\gamma, y + \ln(1 - \gamma)) - u(\tau, x, y)]. \end{cases} \tag{4.3.2}$$

The boundary conditions (4.2.13) become:

$$\begin{cases} u(0, x, y) = (e^y - K)^+, \\ \lim\limits_{y \to -\infty} u(\tau, x, y) = 0, \\ \lim\limits_{y \to \infty} \dfrac{u(\tau, x, y)}{e^y} = 1, \\ \lim\limits_{x \to -\infty} u(\tau, x, y) = C^{BS}(e^y, \tau, K, r, \sigma), \\ \lim\limits_{x \to \infty} \dfrac{\partial u}{\partial x}(\tau, x, y) = 0. \end{cases} \tag{4.3.3}$$

A second-order Taylor expansion is adopted to approximate the jump term $\overline{\mathcal{L}_2} u$ as follows:

$$\begin{aligned} \overline{\mathcal{L}_2} u &= e^x [u(\tau, x - \beta\gamma, y + \ln(1 - \gamma)) - u(\tau, x, y)] \\ &= e^x [-\beta\gamma \frac{\partial u}{\partial x} + \ln(1 - \gamma) \frac{\partial u}{\partial y} + \frac{\beta^2 \gamma^2}{2} \frac{\partial^2 u}{\partial x^2} + \frac{\ln^2(1 - \gamma)}{2} \frac{\partial^2 u}{\partial y^2} - \beta\gamma \ln(1 - \gamma) \frac{\partial^2 u}{\partial x \partial y}] \\ &\quad + O(\beta^2 \gamma^2 + \ln^2(1 - \gamma)). \end{aligned}$$

With the high order terms being dropped out, the variational inequality becomes

$$\frac{\partial u}{\partial \tau} \geq a(x)\frac{\partial^2 u}{\partial x^2} + b(x)\frac{\partial^2 u}{\partial y^2} + c(x)\frac{\partial u}{\partial x} + d(x)\frac{\partial u}{\partial y} + \rho(x)\frac{\partial^2 u}{\partial x \partial y} - ru, \tag{4.3.4}$$

where

$$a(x) = \frac{\kappa^2 + \beta^2\sigma^2 + \beta^2\gamma^2 e^x}{2}, b(x) = \frac{\sigma^2 + \ln^2(1-\gamma)e^x}{2},$$

$$c(x) = \alpha(\overline{x} - x) + \beta(r - \gamma e^x), d(x) = r - \frac{\sigma^2}{2} + e^x \ln(1-\gamma), \rho(x) = \beta\sigma^2 - \ln(1-\gamma)\beta\gamma e^x.$$

The LCP (4.3.1) is defined on an unbounded domain :

$$\bar{\Omega} := [0, T] \times \mathbf{R} \times \mathbf{R}. \tag{4.3.5}$$

In order to use finite difference approximation for spacial variables, we need to truncate the infinite domain into a finite domain:

$$\{(\tau, x, y) \in [0, T] \times [-X_{\max}, X_{\max}] \times [-Y_{\max}, Y_{\max}]\}. \tag{4.3.6}$$

Theoretically, to eliminate the boundary effect, $X_{\max}$ and $Y_{\max}$ should be sufficiently large. However, truncation is necessary if we want to adopt the finite difference method. Wilmott et al. (1995) pointed out that the upper bound of stock price should be three or four times the strike price. Therefore it is reasonable to set $Y_{\max} = \ln(5K)$. As for the buy-in rate, x, we set $X_{\max} = \ln(252)$, which implies that buy-in occurs every day at most. The space $\overline{\Omega}$ is divided into a uniform grid with

$$\begin{aligned} x_i &= -X_{\max} + (i-1)\cdot\Delta x, i = 1, \cdots, M, \\ y_j &= -Y_{\max} + (j-1)\cdot\Delta y, j = 1, \cdots, N, \\ \tau_l &= (l-1)\cdot\Delta\tau, l = 1, \cdots, L, \end{aligned}$$

where M, N, L are the number of grid points in the x, y, τ directions, respectively. The value of a call option at each grid point thus is denoted as $u^l_{i,j} = u(x_i, y_j, \tau_l)$.

In the interior of pricing domain, the central differences are applied to approximate the first and second derivatives in the variational inequality (4.3.4). Thus we have

$$\begin{array}{rclrcl} \frac{\partial u}{\partial x} & = & \frac{u_{i+1,j}-u_{i-1,j}}{2\Delta x}, & \frac{\partial^2 u}{\partial x^2} & = & \frac{u_{i+1,j}-2u_{i,j}+u_{i-1,j}}{\Delta x^2} \\ \frac{\partial u}{\partial y} & = & \frac{u_{i,j+1}-u_{i,j-1}}{2\Delta y}, & \frac{\partial^2 u}{\partial y^2} & = & \frac{u_{i,j+1}-2u_{i,j}+u_{i,j-1}}{\Delta y^2} \end{array}$$

Following Zhu & Chen (2011), the cross-derivative term is discretized as

$$\frac{\partial^2 u}{\partial x \partial y} = \frac{u_{i+1,j+1} + u_{i-1,j-1} - u_{i+1,j-1} - u_{i-1,j+1}}{4\Delta x \Delta y}. \tag{4.3.7}$$

According to Equation (4.2.13), the boundary conditions on the truncated domain are

$$\begin{cases} u(\tau, x, y_1) = C^{\mathrm{BS}}(e^y, \tau, K, r, \sigma), \\ u(\tau, x, y_N) = e^y - K, \\ u(\tau, x_1, y) = 0, \\ u(\tau, x_N, y) = u(x_{N-1}, y, \tau). \end{cases} \tag{4.3.8}$$

The spatial discretization leads to a semi-discrete equation which has the matrix form

$$\frac{\partial U}{\partial \tau} \geq \mathbf{A}U, \tag{4.3.9}$$

where the vector U is the value on mesh points and $\mathbf{A}$ is a sparse coefficient matrix that arises from the finite difference equations. The size of U and $\mathbf{A}$ is MN and $MN \times MN$, respectively.

A general θ-scheme is applied to discretize the semi-discrete problem (4.3.9) as

$$(\mathbf{I} - \theta\Delta\tau\mathbf{A})U^{l+1} \geq [\mathbf{I} + (1-\theta)\Delta\tau\mathbf{A}]U^l, l = 1, \cdots, L, \tag{4.3.10}$$

where L is the number of time steps and $\mathbf{I}$ is the identity matrix. It becomes a fully explicit scheme when θ becomes zero; while a fully implicit one when θ is one. When $\theta = \frac{1}{2}$, it

corresponds to the famous Crank-Nicolson scheme.

The LCP (4.3.1) now has become a sequence of LCPs,

$$\begin{cases} \mathbf{B}U^{l+1} \geq \mathbf{C}U^{l}, \\ U^{l+1} \geq g, \\ (\mathbf{B}U^{l+1} - \mathbf{C}U^{l})(U^{l+1} - g) = 0, \end{cases} \tag{4.3.11}$$

for $l = 0, \cdots, L-1$. The matrix $\mathbf{B}$ and $\mathbf{C}$ are defined by the left hand and right hand of Equation (4.3.10).

When we implement this numerical scheme presented above. The Crank-Nicolson scheme is adopted for most loops. However, to avoid the oscillations with Crank-Nicolson scheme, the fully implicit scheme is adopted for the first three loops and then switch to the Crank-Nicolson scheme. In next section, we present how to solve these LCPs (4.3.11) numerically.

4.3.2 The Lagrange multiplier approach (LMA)

In this section, in order to solve these LCPs (4.3.11) numerically, we adopt the Lagrange multiplier approach (LMA) which was first proposed by Hintermüller et al. (2002) and further considered to treat the complementarity conditions by Ito & Kunisch (2006) and Ito & Toivanen (2009).

To demonstrate the algorithm for the LCPs (4.3.11) conveniently, we denote $z = U^{l+1}$ and $f = CU^{l}$. The initial point z^0 is chosen to be U^{l}. The standard form of these problems (4.3.11) becomes:

$$\begin{cases} \mathbf{B}z \geq f, \\ z \geq g, \\ (\mathbf{B}z - f)(z - g) = 0, \end{cases} \tag{4.3.12}$$

with initial guess z^0.

The LCPs (4.3.12) are approximated by a nonlinear equation with Lagrange multipliers ϵ and ν as follows:

$$\mathbf{B}z = f + \max\{\nu + \epsilon(g - z), 0\}, \tag{4.3.13}$$

where ϵ is a prescribed positive constant and ν is a given vector with nonnegative components. The penalty term $\epsilon(g - z)$ is to force the approximate solution to approach the feasible one. The main purpose to introduce the vector ν is that it leads to a feasible solution g when it is chosen as

$$\nu = \max\{\mathbf{B}g - f, 0\}. \tag{4.3.14}$$

Since Equation (4.3.13) is nonlinear and non-smooth, a semi-smooth Newton method is applied to solve this nonlinear equation by Zvan et al. (1998). The iteration method is adopted to solve the nonlinear equations (4.3.13) with the iteration form as follows:

$$z^{k+1} = z^k + d^k, \tag{4.3.15}$$

where the vector d^k is the solution of the following linear equation

$$\mathbf{J}(z^k)d^k = f + \max\{\nu + \epsilon(g - z^k), 0\} - \mathbf{B}z^k \triangleq r^k. \tag{4.3.16}$$

The matrix $\mathbf{J}(z^k)$ in Equation (4.3.16) belongs to the generalized Jacobian at z^k and it is chosen to be

$$[\mathbf{J}(z^k)]_{i,j} = \mathbf{B}_{i,j} + \begin{cases} \epsilon & \text{if } i = j \text{ and } \nu_i + \epsilon(g_i - z_i^k) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Here we come to the accuracy of the approximation (4.3.13) to the original LCP (4.3.12). Ito & Toivanen (2009) pointed out that the parameter ϵ controlled the accuracy of the approximation (4.3.13) as follows:

$$|z^\epsilon - z|_{L^\infty(\Omega)} \leq \frac{C}{\epsilon}, \tag{4.3.17}$$

where C is a constant and z^ϵ is the solution to Equation (4.3.13) with a given ν defined in Equation (4.3.14). It indicates that accuracy of the Lagrange multiplier approach is at most the order of $\frac{1}{\epsilon}$. As the Lagrange multiplier ϵ approaches infinity, the approximate solution converges to the true solution. In this chapter, we choose ϵ as 10^3 when we implement our numerical scheme.

Finally, we briefly summarize the Lagrange multiplier approach as follows:

Algorithm 2 Lagrange multiplier approach

Require:

Set $x^0 = g$; $k = 0$;$\mathbf{J} = \mathbf{B}$;

Ensure:

while iteration error is not satisfied **do**
  $k = k + 1$;
  for $i = 1, \cdots, m$ **do**
    if $i = j$ and $\nu_i + \epsilon(g_i - z_i) > 0$ **then**
      $\mathbf{J}_{i,i} = \mathbf{B}_{i,i} + \epsilon$;
    end if
  end for
  $r^k = f + \max\{\nu + \epsilon(g - z^k, 0)\} - \mathbf{B}z^k$;
  solve linear equation $\mathbf{J}d^k = r^k$;
  $z^{k+1} = z^k + d^k$;
end while

4.4 Numerical results and discussions

In this section, we report the results of numerical experiments with the Lagrange multiplier approach (LMA). First of all, it is verified that the A&L model degenerates to the Black-Scholes model when γ becomes zero. Then the convergence on different girds is demonstrated when γ is non-zero. By observing the early exercise premium, it is demonstrated that early exercise is optimal for an American call option when γ is non-zero. The reason of early exercise is also demonstrated from both financial and mathematical views. To what extent the early exercise decision would be affected and how the optimal exercise price would be affected by lending fees are also provided.

All the computations reported in this chapter were performed on a machine with 64-bit Intel Xeon 3.50 GHz system and 16GB of RAM. We iteratively solve the LCPs (4.3.11) with the iteration error being set to be 10^{-4}.

4.4.1 Convergence

We begin this section by discussing the implementation for the degenerate case ($\gamma = 0$) of the A&L model. It indicates that there is no jump in the stock price and the A&L model degenerates to the Black-Scholes model. In this degenerate case, it is optimal to hold an American call option to the expiration, which is consistent with the conclusion obtained by Merton (1973). Since the option is exercised at the expiration, the value of an American call option should be the same with that of the corresponding European one. Consequently, the values obtained from Black-Scholes formula for European options can be regarded as the benchmark solution.

To implement our numerical scheme, the other parameters are set as follows:

$$K = 10, r = 0.05, \sigma = 0.45, T = 0.5, \alpha = 2, \kappa = 0.2, \overline{x} = \ln(10), x_0 = \ln(12), \beta = 1. \quad (4.4.1)$$

The values of an American call option under the degenerate A&L model on different grids are listed in Table 4.1. The values calculated from the Black-Scholes formula are considered as the benchmark to verify the numerical scheme.

	Grid (M, N, L)	Asset price 8	9	10	11	12	l_2 error
	(40,100,20)	0.4595	0.8553	1.3869	2.0332	2.7792	0.0171
	(50,150,30)	0.4553	0.8503	1.3806	2.0299	2.7804	0.0082
	(60,200,40)	0.4548	0.8471	1.3745	2.0287	2.7783	0.0031
	(70,250,50)	0.4547	0.8474	1.3766	2.0277	2.7786	0.0029
Black-Scholes		0.4537	0.8460	1.3758	2.0269	2.7767	

Table 4.1: Convergence of values of an American call option with $\gamma = 0$.

To measure the convergence of the numerical results, the l_2 errors between our numerical results and the benchmark results are provided in the last column of the table. From Table 4.1, it is observed that the l_2 error decreases gradually as the size of girds diminishes, which demonstrates that the degenerate A&L model is really equivalent to the Black-Scholes model. In this sense, the A&L model can be considered as a generalized Black-Scholes model.

When the A&L model is not degenerate, the numerical results for an American call option are provided in Table 4.2 with $\gamma = 0.03$. The other parameters are kept the same with those

shown in Equation (4.4.1). In this case, a reference solution is needed to demonstrate that the numerical results we obtained are converged. Since there are still no analytic solutions under the non-degenerate A&L model in the literature, the reference solution is constructed on fine grids defined as $(M, N, L) = (60, 400, 50)$. Again, the l_2 error between our numerical results and the reference solution are listed in the last column of Table 4.2. From Table 4.2, it is

	Grid (M, N, L)	Asset price 8	9	10	11	12	l_2 error
LMA	(20,50,10)	0.2028	0.4350	0.7602	1.3744	2.0327	0.0618
	(30,100,20)	0.1965	0.4285	0.8000	1.3318	2.0554	0.0071
	(40,150,30)	0.1965	0.4290	0.7998	1.3354	2.0652	0.0059
	(50,200,40)	0.1984	0.4285	0.7945	1.3359	2.0646	0.0047
reference	(60,400,50)	0.1977	0.4276	0.7964	1.3340	2.0609	

Table 4.2: Convergence of values of an American call option with $\gamma = 0.03$.

observed that the decreasing l_2 error has indicated that our numerical scheme also converges when $\gamma = 0.03$.

The convergence of our numerical schemes under both the degenerate and non-degenerate A&L model has been shown in Table 4.1 and Table 4.2. Next, we would focus on some discussions about the these numerical results calculated from our reliable numerical scheme, especially whether or not it is optimal to exercise an American call option under the A&L model.

4.4.2 Discussions on early exercise of an American call option

The significant difference between American and European options relates to when the option can be exercised: investors have no choices but to hold European options to the expiration; while they can exercise American options at any time before the expiration. At any time t, prior to expiration, the investor holding an American call option would consider to exercise it right away or continue to hold it. Here, we define the *early exercise premium* as the difference between the price of an American call option and the price of the corresponding European one.

The possibility of early exercise for an American call option indicates that the early exercise premium may be positive. In other words, the positive premium really exists if the American one is supposed to be exercised early; while the premium would disappear if it is held to expiration, as is the same to the European one. Instead of obtaining the optimal exercise price

explicitly, we could also infer whether or not an American call option should be exercised early through exploring the early exercise premium.

In this chapter, we would first consider the early exercise premium to determine whether or not an American call option should be exercised early and then produce the optimal exercise price to make it clear when it should be exercised. The values of an American call option written on the hard-to-borrow stock can be calculated as we presented in the previous section; while the values for a European call option have been obtained by Ma et al. (2016). These values are shown and compared in Figure 4.1.

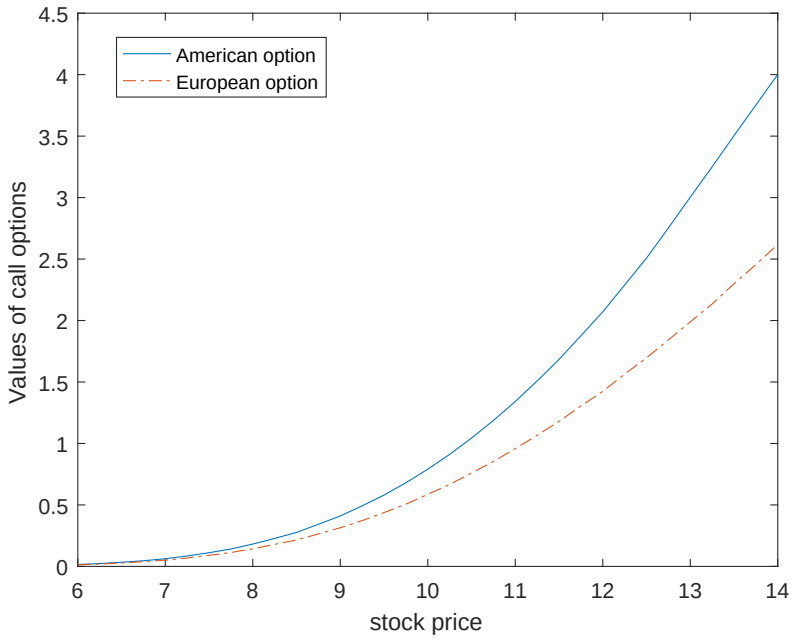


Figure 4.1: The price of American and European call options with $\gamma = 0.03$. The other parameters are $K = 10, r = 0.05, \sigma = 0.45, T = 0.5, \alpha = 2, \kappa = 0.2, \overline{x} = \ln(10), x_0 = \ln(12), \beta = 1$.

From Figure 4.1, it is obvious to note that the American call option really has premium with respect to the corresponding European one, which indicates that an American call option under the A&L model with $\gamma = 0.03$ may be exercised early if the option is deep in-the-money although the underlying hard-to-borrow stock does not pay dividends in practice.

This observation is obviously not consistent with the conclusion from Merton (1973) that one should never exercise a call option before the expiration or dividend payment. In fact,

Jensen & Pedersen (2016) have argued that "never say never" and demonstrated that early exercise of an American call option can be optimal and rational in light of financial frictions, such as short-sale costs and transaction costs, although the underlying stock does not pay dividends. Jensen and Pedersen obtained their conclusion by considering the Black-Scholes model with frictions and enough empirical data. In this chapter, we adopt the A&L model, a new one that especially describes the hard-to-borrow stocks, to verify and complement Jensen and Pedersen's work.

Furthermore, to explore how the A&L model would lead to early exercise for an American call option, we would analyze from both financial and mathematical views.

From financial perspective, the motivation to exercise an American call option early is to avoid the possible buy-in risk and lending costs. In the market, a trader who longs a call option needs to short the underlying stock to hedge the risk. However, when the stock is hard-to-borrow following the A&L model, the trader could possibly suffer repeated buy-ins, which implies that the clearing firm may require him to buy back the stock he had shorten. To meet the requirement of buy-in, the investor has to borrow stock in the market by paying extra lending fees. Early exercise becomes a good choice for the holder of an American call option to avoid the risk associated with the possible buy-ins in the future. In other words, it is possible lending fee that drives the investor to exercise the option early.

From the mathematical view, we focus on the operator in Equation (4.2.7). Once applying the Taylor expansion to the jump operator $\mathcal{L}_2$, we obtain a term $S(r-\lambda\gamma)\dfrac{\partial u}{\partial S}$ in the PDE. While in the Black-Scholes model where the stock pays dividends continuously, the dividend payment is reflected by a term $S(r-q)\dfrac{\partial u}{\partial S}$ in the PDE with q being the rate of continuous dividends. While $\lambda\gamma$ is viewed by Avellaneda & Lipkin (2009) as the cost-of-carry for borrowing the stock or the lending fees for the hard-to-borrow stock, mathematically, it is considered as the "equivalent" dividend rate q in the classic Black-Scholes model, which also explains the early exercise of call option perfectly from mathematical point. In other words, the lending fees paid by the borrower to cover the possible buy-in can be understood as equivalent dividends for the lender of hard-to-borrow stock. For the Black-Scholes model with continuous dividends, the

rate of dividends q would affect the option price, which can be analyzed via the Black-Scholes formula. Now we would like to explore to what extent the option price would be affected by the lending fees under the A&L model. To demonstrate these relations more clearly, we present some more numerical results in Figure 4.2 with different values of γ.

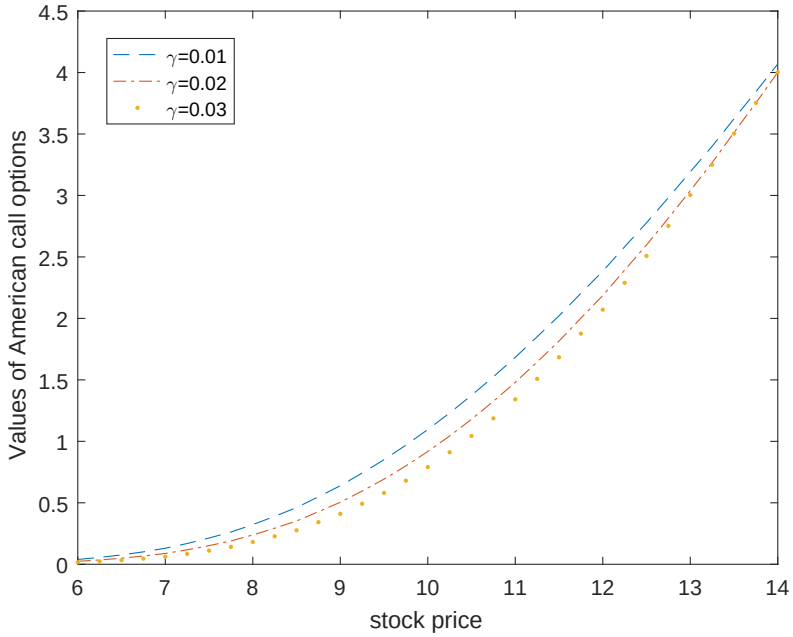
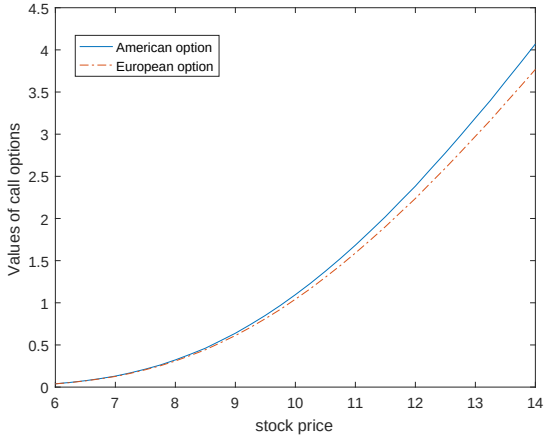


Figure 4.2: The price of an American call option with different values of γ. The other parameters are $K = 10, r = 0.05, \sigma = 0.45, T = 0.5, \alpha = 2, \kappa = 0.2, \overline{x} = \ln(10), x_0 = \ln(12), \beta = 1$.

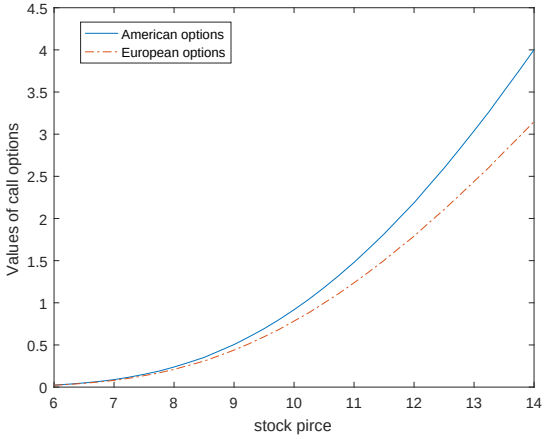
It is observed that the value of an American call option decreases gradually as the value of γ increases, which is consistent with the fact that the option price decreases with dividend rate q increasing. In order to demonstrate how the parameter γ in the A&L model affect the early exercise premium, more curves are plotted with different values of γ in Figure 4.3.

From Figures 4.3(a), 4.3(b) and 4.3(c), it is observed that the early exercise premium always exists when γ is non-zero. In addition, the variation of the early exercise premium with different values of γ is also presented in Figure 4.3(d), from which it is clear that the premium becomes larger and larger as the value of γ increases. It implies that the investor is more likely to exercise the call option early with a larger value of γ .

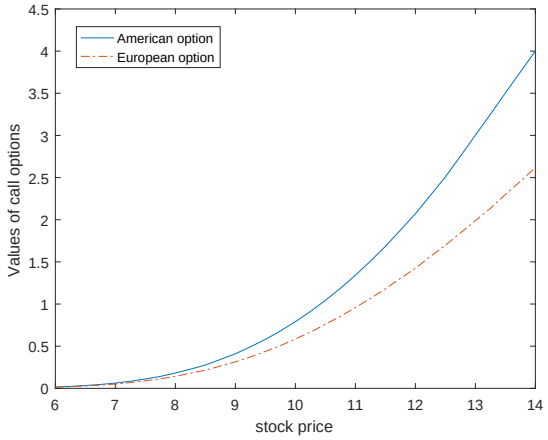
It has been demonstrated that an American call option may be optimally exercised early when γ is non-zero. The early exercise decision is made based on whether the spot price of



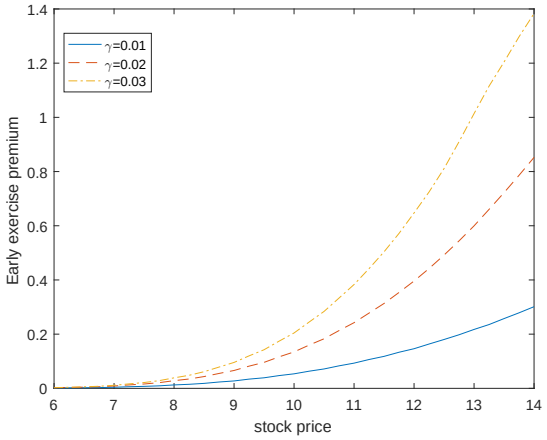
(a) $\gamma = 0.01$



(b) $\gamma = 0.02$



(c) $\gamma = 0.03$



(d) Early exercise premium

Figure 4.3: The price of European and American call options with different values of γ and the corresponding early exercise premium. The other parameters are $K = 10, r = 0.05, \sigma = 0.45, T = 0.5, \alpha = 2, \kappa = 0.2, \overline{x} = \ln(10), x_0 = \ln(12), \beta = 1$.

the underlying has exceeded the optimal exercise price $S_f(\tau, x)$ or not. So, a critical piece of information needed by traders is optimal exercise price, which mathematically is a free boundary to be determined as part of the solution. Under the LCP, this part of the solution is indirectly recovered from the option price $u(\tau, x, S)$ which is obtained through solving the LCP (4.3.11) directly. The process of recovering the free boundary $S_f(\tau, x)$ from option price is quite standard (Wilmott et al. 1995). In the continuation region Ω^C, the American call option should not be exercised and we have

$$u(\tau, x, S) > S - K, \quad S < S_f(\tau, x). \tag{4.4.2}$$

On the other hand, in the exercise region Ω^E, the option should be exercised right away. Mathematically, it is written as

$$u(\tau, x, S) = S - K, \quad S \geq S_f(\tau, x). \tag{4.4.3}$$

The recovery of the free boundary, which divides these two regions, can thus be simply realized through a root-finding problem of finding the minimum root S that satisfies the equation $u(\tau, x, S) = S - K$. With the option price $u(\tau, x, S)$ obtained through the LCP on a set of discrete grids, such a minimum root can be numerically settled on the nearest point to the free boundary in continuation region Ω^C, i.e.,

$$S_f(\tau_l, x_i) = \max_j \{S_j, j = 1, \cdots, n | u(\tau_l, x_i, S_j) > S_j - K\}. \tag{4.4.4}$$

Remark 4.4.1. It should be pointed out that non-smoothness is usually associated with the recovered free boundary, which is part of "cons" of the LCP among many "pros" of the approach (Ikonen & Toivanen 2007, 2008). The standard algorithm stated above is of no exception. To produce a smooth free boundary, some smoothing techniques need to be applied. In the following figures, we have adopted the moving average function[2] in the MATLAB to plot these

[2]Detailed information can be found at the website: http://au.mathworks.com/help/curvefit/smooth.html.

smooth curves.

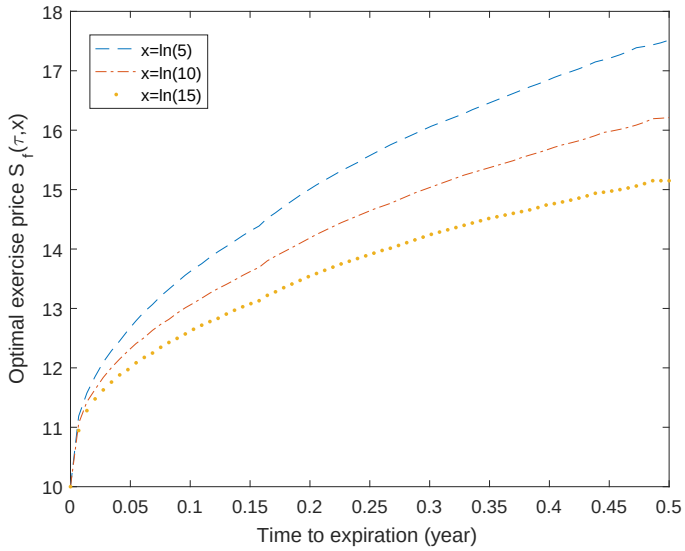


Figure 4.4: Optimal exercise price with different values of x_0. The other parameters are $K = 10, r = 0.05, \sigma = 0.45, T = 0.5, \alpha = 2, \kappa = 0.2, \overline{x} = \ln(10), \gamma = 0.01, \beta = 1$.

Depicted in Figure 4.4 is the optimal exercise price $S_f(\tau, x)$ as a function of time to expiration τ and buy-in rate x. As expected, the optimal exercise price $S_f(\tau, x)$ is a monotonically increasing function with time to expiration τ; while it is a decreasing function of buy-in rate x. As the value of buy-in rate x increases, there are more possible buy-ins and larger lending costs in the future. Consequently, the holder of an American call option would choose to exercise it early to avoid the possible buy-in risk instead of holding them to the expiration. It is reflected in Figure 4.4 that the optimal exercise price drops down as the buy-in rate x goes up.

To demonstrate how the other parameter γ affects the early exercise decision, the optimal exercise price with different values of γ are plotted in Figure 4.5.

It is observed from Figure 4.5 that the optimal exercise price decreases simultaneously as the parameter of jump size γ increases. Financially, the larger the jump size is, the higher buy-in risk is. To avoid the more possible buy-in risk, the investor would exercise the option early at a lower price, when the jump size γ is larger.

Both a large current buy-in rate x and a large jump size γ would lead to possible early exercise of an American call option because either of them indicates that the possible risk

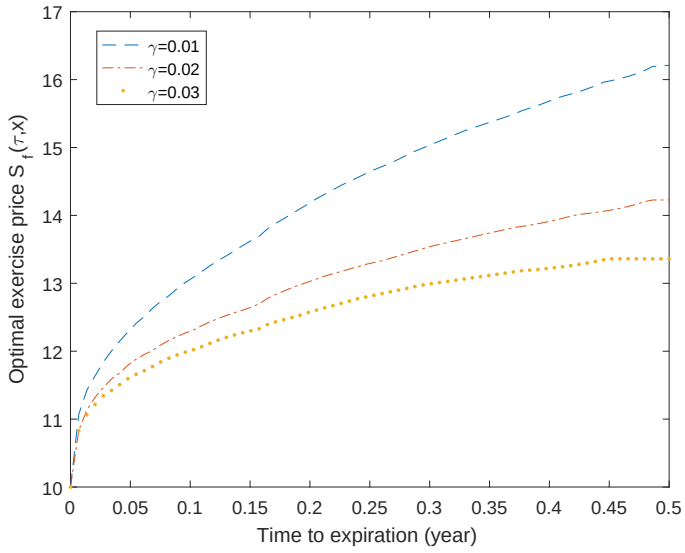


Figure 4.5: Optimal exercise price with different values of γ. The other parameters are $K = 10, r = 0.05, \sigma = 0.45, T = 0.5, \alpha = 2, \kappa = 0.2, \overline{x} = \ln(10), x = \ln(12), \beta = 1$.

associated with buy-ins in the future has been in a high level. Once these signals are observed in the market, early exercise of American call options may be a good choice if they are deep in-the-money.

4.5 Conclusions

In this chapter, we present how to price an American call option under the A&L model. The PDE system with free boundary is established first and then it is reformulated as a linear complementarity problem, which is numerically solved with Lagrange multiplier approach.

According to our numerical results, early exercise of American call options may be optimal when the options are deep in-the-money, although the underlying stock pays no dividends. Our results reassure the conclusion of Jensen & Pedersen (2016) under the A&L model. In addition, we quantify why the early exercise would occur both from the view of financial and mathematical points. How the parameters in A&L model affect the optimal exercise price is also provided numerically.

Part II: Option pricing with short selling bans

In Part 2, short selling is completely banned in the financial market, which makes the market become incomplete. Option pricing with short selling bans is considered as a special case of option pricing in incomplete market. Recently, Guo & Zhu (2017) proposed a new *equal-risk pricing approach* to price options with convex trading constraints. Only when the payoff function is monotonic, can they obtain the analytical pricing formula; while it does not work when the payoff function is non-monotonic. In this part, we intend to establish the PDE framework for equal-risk pricing approach so that the range of its application is expanded.

Since the HJB equation is involved in the process of establishing the PDE framework, we first explore different solution approaches to the HJB equations as preliminaries. Series solution approach, analytical solution approach and numerical solution approach are presented accordingly in Chapter 5, Chapter 6 and Chapter 7. In Chapter 5, we present an exact and explicit solution for the HJB equation with general utility functions. The solution is written in the form of a Taylor's series expansion and constructed through the homotopy analysis method. The fully nonlinear HJB equation is decomposed into an infinite series of linear PDEs that can be solved analytically. In Chapter 6, a closed-form analytical solution for the Merton problem defined on a finite horizon with exponential utility function. The solution is obtained through two different methods: an indirect one and direct one. Two solutions are demonstrated to be equivalent although they appear to be of different forms. Some discussions are provided based on our analytical solutions. In Chapter 7, a monotone numerical scheme is presented to solve the HJB equation. The stability, consistency, and monotonicity of the numerical scheme are also demonstrated to guarantee the convergence of the scheme. Finally, in Chapter 8, we establish a PDE framework for equal-risk pricing approach. Our PDE framework not only recovers the analytical pricing formula when the payoff function is monotonic, but also produce numerical results when the payoff function is non-monotonic. As a result, our PDE framework has indeed expanded the range of application of equal-risk pricing approach.

Chapter 5

An analytical solution to the HJB equation arising from the Merton problem

5.1 Introduction

Optimal investment and consumption problem (also referred to as the *Merton problem*) is a well-known and classical topic in mathematical finance. Consider a financial market consisting of only two kinds of assets. One is risk-free asset (such as bond), the price of which grows at a fixed rate and the other is risky asset (such as stock), the price of which follows a geometric Brownian motion. An investor has two choices to allocate his wealth: consumption (such as buying foods and clothes) and investment to accumulate wealth. In order to maximize his expected utility from intermediate consumption and terminal wealth, he needs to choose how much to consume and how to allocate his wealth between the risky asset and risk-free one. This problem has attracted very high attention in academia as well as of financial practitioners and there has been an exponential growth of literature since the seminal papers (Merton 1969, 1971) were published. So far, the problem has only been completely solved with the utility function being some special forms and there is still no consensus in terms of a convincing approach to

solve the cases with general utility functions.

In the landmark paper (Merton 1969), Merton formulated the optimal investment and consumption problem as a stochastic optimal control problem which could be solved with the dynamic programming method. As a result, it leads to the well-known Hamilton-Jacobi-Bellman (HJB) equation, a fully nonlinear partial differential equation (PDE). The main difficulty of the Merton problem is the nonlinearity of the HJB equation, which makes it tough to solve analytically.

The utility function, which describes the investor's risk aversion, is playing an important role in the Merton problem. In expected utility theory, there are many utility functions, such as power utility, logarithmic utility, exponential utility, quadratic utility and so on (Ingersoll 1987). Up to date, only when utility function is of a couple of specific forms, can an analytical solution be obtained. For examples, when the utility function belongs to the constant relative risk aversion (CRRA) class, including power utility function and logarithmic utility function, Merton (1969) produced the analytical solution for the HJB equation on a finite horizon. When utility function is of exponential form, there are some special cases, such as the terminal wealth problem without consumption (Henderson 2005) and the infinite horizon problem (Merton 1969), that do admit a closed-form analytical solution. However, for the general case where both consumption and terminal wealth are taken into account and the horizon is finite, there is still no analytical solution (Wachter 2002).

In reality, the relative risk aversion of an investor may vary with wealth instead of being constant, which gives us motivations to consider some utility functions beyond the CRRA class in the classic Merton problem. Brunnermeier & Nagel (2008) and Liu et al. (2014) have discussed how the relative risk aversion varies with wealth from empirical data. Recently, Fouque et al. (2015) provided a mixed power utility function, of which the relative risk aversion varies with wealth. Under this new utility function, it is still hard to obtain the explicit solution even though consumption is not taken into consideration and we only maximize the utility from terminal wealth, which actually is a degenerate Merton problem. In other words, the existing literature did not provide a convincing approach to solve the HJB equation with general utility

functions, such as this new mixed power utility function.

Up to date, some numerical solutions have also been provided for the HJB equation that arises from the stochastic optimal control problem. These numerical methods may solve the nonlinear HJB equation when utility function is of general forms. Kushner (1990) developed the Markov chain approximation method, of which an extensive overview can be found in Fleming & Soner (2006) and Kushner & Dupuis (2013). Krylov (2000, 2005) presented some finite difference approximations for the HJB equations and estimated the rate of convergence. Barles & Souganidis (1991) also proposed some monotone approximation schemes for the HJB equation in the theory of viscosity solutions, which has been developed further in Barles (1997) and Barles & Jakobsen (2002, 2007). Fahim et al. (2011) recently provided a probabilistic numerical scheme for the fully nonlinear HJB equation and demonstrated that their scheme could be explained naturally as a combination of Monte Carlo and finite difference schemes without appealing to the theory of backward stochastic differential equation (BSDE) (Cheridito et al. 2007). However, most of the numerical methods still require discretizations and truncations, which lead to numerical errors. In addition, intensive computation is also unavoidable before a solution of reasonable accuracy can be obtained and sometimes the method may not even converge, such as the explicit finite difference scheme.

The main contribution of this paper is that we present an analytical solution to the highly nonlinear HJB equation subject to general utility functions for the first time. It is written in the form of a Taylor's series expansion and constructed through the homotopy analysis method (HAM), which was initially suggested by Ortega & Rheinboldt (1970) and was further developed by Liao (2003*a*). Such a method has been successfully applied to solve a growing number of nonlinear ODEs and PDEs in science and engineering. Zhu (2006) first applied this method to the valuation of American put options and managed to produce a completely analytical and exact solution for the optimal exercise boundary and the option price. Then Zhao & Wong (2012) extended the method to investigate American option pricing under general diffusion process and obtained an exact and explicit solution. The essential step of the homotopy analysis method (HAM) is to construct a continuous homotopic deformation through a series

expansion of the unknown function. The series solution of the unknown function is of infinite terms, but is nevertheless of exact and explicit. The fully nonlinear PDE is decomposed into an infinite series of linear PDEs, which can be solved analytically.

As pointed out by Zhu (2006), the series solution constructed through the homotopy analysis method (HAM) is exact and explicit according to the definition given by Gukhal (2001). By 'exact' it means that no approximation is made whatsoever; the partial differential equations and the initial conditions can all be satisfied. By 'explicit' it means that the solution for the unknown functions can be determined explicitly in terms of all the inputs to the problem.

The rest of the chapter is organized as follows. In Section 5.2, the Merton problem is briefly reviewed, to give a complete background reference to the financial context of the HJB equation. In Section 5.3, we apply the homotopy analysis method to solve the fully nonlinear HJB equation arising from the Merton problem subject to general utility functions. In Section 5.4, four examples are presented to demonstrate the accuracy and versatility of the homotopy analysis method. Some conclusions are provided in the last section.

5.2 The Merton problem and the HJB equation

5.2.1 The Merton problem

Consider a financial market with two assets being traded continuously on a finite horizon $[0, T]$. One asset is a risk-free bond, whose price $\{P(t), t \geq 0\}$ evolves according to the ordinary differential equation (ODE)

$$dP(t) = rP(t)dt, \qquad t \in [0, T], \tag{5.2.1}$$

with r being the risk-free interest rate. The other one is a risky asset with its price following a geometric Brownian motion

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \qquad t \in [0, T], \tag{5.2.2}$$

where μ is the drift rate, σ is the volatility, and $W(t)$ is a standard Brownian motion.

An investor starts with a known initial wealth x_0 and the wealth at time t is denoted as $X(t)$. At any time t, prior to T, the investor needs to make a decision on how much to consume and, in the mean time, how much to invest in stock markets, in order to maximize his expected utility from intermediate consumption and terminal wealth. The consumption rate per unit time at time t is denoted as $c(t)$ and the investment proportion $u(t)$ represents the fraction of total wealth that is invested in the risky asset at time t. The remaining fraction $1-u(t)$ is thus left in form of the risk-free bond within the framework of this two-asset model. The investment proportion on the underlying stock $u(t)$ may be negative, which is to be interpreted as short selling. The remaining proportion $1-u(t)$ may also become negative and this corresponds to borrowing at the interest rate r.As a result, the total wealth $X(t)$ is governed by the following SDE:

$$dX(t) = \{[r + u(t)(\mu - r)]X(t) - c(t)\}dt + X(t)u(t)\sigma dW(t). \tag{5.2.3}$$

The objective of the Merton problem is to obtain the optimal investment and consumption policies, i.e. to determine $u(t)$ and $c(t)$, such that the expected utility from accumulated consumption and the terminal wealth is maximized. Mathematically, such an objective functional is stated as

$$\max_{(u(\cdot),c(\cdot))} \quad \mathbf{E}[\int_0^T e^{-\rho s}U(c(s))ds + e^{-\rho T}B(X_T)], \tag{5.2.4}$$

where $\mathbf{E}$ is the expectation operator; ρ is the subjective discount rate; U is a function measuring the utility from intermediate consumption $c(t)$ and B is also a function measuring the utility from terminal wealth X_T. In addition, some constraints may be imposed on the wealth process and consumption process to make sure the objective functional well defined. When the utility function is defined on $\mathbf{R}^+$, such as power and logarithmic function adopted by Merton (1969), we need the following constraint

$$c(t) \geq 0, \quad X(t) \geq 0 \quad t \in [0, T]. \tag{5.2.5}$$

If the domain of utility function is $\mathbf{R}$, such as exponential utility, the above constraint is not

necessary.

In a brief summary, the Merton problem has been reformulated as a stochastic optimal control problem with the objective functional (5.2.4), driven by the dynamics of the wealth (5.2.3), and subject to the constraints (5.2.5).

5.2.2 The HJB equation

We apply the dynamic programming method to find out a solution of the stochastic optimal control problem in previous subsection, for the wealth process $X(t)$ is obviously Markovian. The basic idea of the dynamic programming is to consider a family of stochastic optimal control problems with different initial times and states, to establish relationships among these problems and finally to solve all of them.

Let $(t, x) \in [0, T) \times \mathbf{R}^+$ and consider the following control system over $[t, T]$

$$\begin{cases} dX(s) = \{[r + u(s)(\mu - r)]X(s) - c(s)\}ds + X(s)u(s)\sigma dW(s), \\ X(t) = x, \end{cases} \tag{5.2.6}$$

with the same constraints (5.2.5). The cost functional is

$$\mathbf{J}(c(\cdot), u(\cdot); t, x) = \mathbf{E}[\int_t^T e^{-\rho s}U(c(s))ds + e^{-\rho T}B(X(T))]. \tag{5.2.7}$$

Define the value function as

$$V(t, x) = \max_{(u(\cdot), c(\cdot))} \mathbf{J}(c(\cdot), u(\cdot); t, x). \tag{5.2.8}$$

According to the dynamic programming method (Yong & Zhou 1999), the value function $V(t, x)$ satisfies the so-called HJB equation:

$$\begin{cases} \max\limits_{(u,c)} \phi(u, c; t, x) = 0, \\ V(T, x) = e^{-\rho T}B(x), \quad \forall (t, x) \in [0, T] \times [0, \infty), \end{cases} \tag{5.2.9}$$

where

$$\phi(u,c;t,x)=\frac{\partial V}{\partial t}+\{[u(\mu-r)+r]x-c\}\frac{\partial V}{\partial x}+\frac{1}{2}x^2\sigma^2u^2\frac{\partial^2V}{\partial x^2}+e^{-\rho t}U(c). \tag{5.2.10}$$

The first-order optimal conditions for a regular interior maximum to (5.2.9) are

$$\begin{cases}\dfrac{\partial\phi}{\partial c}=-\dfrac{\partial V}{\partial x}+e^{-\rho t}\dfrac{\partial U}{\partial c}=0,\\ \dfrac{\partial\phi}{\partial u}=(\mu-r)x\dfrac{\partial V}{\partial x}+ux^2\sigma^2\dfrac{\partial^2V}{\partial x^2}=0.\end{cases} \tag{5.2.11}$$

The optimal pair (u^*,c^*) can be obtained in terms of the value function $V(t,x)$

$$\begin{cases}u^*=-\dfrac{\mu-r}{x\sigma^2}\dfrac{V_x}{V_{xx}},\\ c^*=(\dfrac{\partial U}{\partial c})^{-1}(e^{\rho t}V_x),\end{cases} \tag{5.2.12}$$

where $V_{xx}:=\frac{\partial^2V}{\partial x^2}, V_x:=\frac{\partial V}{\partial x}$, and $f(x)^{-1}$ denotes the inverse function of $f(x)$[1]. After substituting the optimal pair (u^*,c^*) into the HJB equation (5.2.9), we have another nonlinear PDE system

$$\begin{cases}V_t+rxV_x-(\dfrac{\partial U}{\partial c})^{-1}(e^{\rho t}V_x)V_x-\dfrac{\lambda^2}{2}\dfrac{V_x^2}{V_{xx}}+e^{-\rho t}U(c^*)=0,\\ V(T,x)=e^{-\rho T}B(x),\quad\forall(t,x)\in[0,T]\times[0,\infty),\end{cases} \tag{5.2.13}$$

where $V_t:=\dfrac{\partial V}{\partial t}$ and $\lambda=\dfrac{\mu-r}{\sigma}$ is the market price of risk of the underlying stock. This PDE system (5.2.13) was also obtained by Merton when the utility function was of power or logarithm form (Merton 1969). Although the maximization operator in the HJB equation (5.2.9) has disappeared, the new PDE system (5.2.13) is still nonlinear. In the following section, we will demonstrate how to solve this nonlinear PDE system (5.2.13) based on the homotopy analysis method (HAM).

[1]We assume utility function U is smooth. Since utility function is concave, $\frac{\partial^2U}{\partial c^2}<0$ always holds. As a result, $\frac{\partial U}{\partial c}$ is always a monotonic function and its inverse function always exists.

5.3 Solution for the HJB equation based on the HAM

Obviously, the PDE system (5.2.13) consists of a nonlinear parabolic PDE with a given terminal condition. Generally, there is no standard procedure to obtain the analytical solution to such a nonlinear PDE system unless utility function is of some special forms. Various numerical schemes have also been developed in the literature (Kushner 1990, Barles & Souganidis 1991, Krylov 2000, Fahim et al. 2011). However, when numerical schemes are adopted, the discretization errors are unavoidable and the intensive computations are needed. In this section, we apply the homotopy analysis method (HAM) to solve the nonlinear PDE system (5.2.13). An explicit series solution is obtained in terms of a Taylor's expansion.

By introducing $\tau = T - t$, the PDE system (5.2.13) can be rewritten as :

$$\begin{cases} \mathcal{L}V(\tau, x) = \mathcal{A}V(\tau, x), \\ V(0, x) = g(x), \quad \forall(\tau, x) \in [0, T) \times [0, \infty), \end{cases} \tag{5.3.1}$$

where $\mathcal{L}$ is a linear differential operator defined as

$$\mathcal{L} = \frac{\partial}{\partial \tau} - rx\frac{\partial}{\partial x}, \tag{5.3.2}$$

and $\mathcal{A}$ is a nonlinear operator defined as

$$\mathcal{A}V(\tau, x) = -(\frac{\partial U}{\partial c})^{-1}(e^{\rho(T-\tau)}V_x)V_x - \frac{\lambda^2}{2}\frac{V_x^2}{V_{xx}} + e^{-\rho(T-\tau)}U(c^*), \tag{5.3.3}$$

with $g(x) = e^{-\rho T}B(x)$. The nonlinearity of the PDE system (5.3.1) is focused on the operator $\mathcal{A}$. Let us now construct a new unknown function $\bar{V}(\tau, x, p)$ which is governed by the following deformation equation

$$\begin{cases} (1-p)\mathcal{L}[\bar{V}(\tau, x, p) - V_0(\tau, x)] = -p(\mathcal{L} - \mathcal{A})\bar{V}(\tau, x, p), \\ \bar{V}(0, x, p) = (1-p)V_0(0, x) + pg(x), \end{cases} \tag{5.3.4}$$

where $V_0(\tau, x)$ is the initial guess for the true solution to the PDE system (5.3.1) and the

parameter $p \in [0,1]$. Actually, we are introducing a continuous map such that the solution $V(\tau,x)$ becomes the result of the continuous deformation from the initial and known function $V_0(\tau,x)$. When $p=0$, it is straightforward that

$$\begin{cases} \mathcal{L}[\bar{V}(\tau,x,0) - V_0(\tau,x)] = 0, \\ \bar{V}(0,x,0) = V_0(0,x). \end{cases} \tag{5.3.5}$$

Clearly, by the uniqueness of the linear PDE, we have $\bar{V}(\tau,x,0) = V_0(\tau,x)$. On the other hand, if $p=1$, the deformation equation becomes

$$\begin{cases} \mathcal{L}\bar{V}(\tau,x,1) = \mathcal{A}\bar{V}(\tau,x,1), \\ \bar{V}(0,x,1) = g(x). \end{cases} \tag{5.3.6}$$

Obviously, it follows that $\bar{V}(\tau,x,1) = V(\tau,x)$. In other words, as p increases from 0 to 1, $\bar{V}(\tau,x,p)$ varies from the initial guess $V_0(\tau,x)$ to the true solution $V(\tau,x)$ of the PDE system (5.3.1).

Here we make an assumption that the unknown function $\bar{V}(\tau,x,p)$ is analytic with respect to p, which implies that it is smooth enough and thus can be expanded in a Taylor's series expansion of p as follows:

$$\bar{V}(\tau,x,p) = V_0(\tau,x) + \sum_{n=1}^{+\infty} \frac{V_n(\tau,x)}{n!} p^n, \tag{5.3.7}$$

where

$$V_n(\tau,x) = \frac{\partial^n}{\partial p^n} \bar{V}(\tau,x,p)|_{p=0}. \tag{5.3.8}$$

To obtain the coefficients in the expression (5.3.7), we need to derive a set of PDEs governing the unknown function $V_n(\tau,x)$ and some appropriate initial conditions for them. They can be achieved by differentiating the PDE equation (5.3.4) with respect to embedding parameter p n times and then setting p to zero.

First of all, by differentiating the PDE equation (5.3.4) only once and setting $p=0$, we

have

$$\begin{cases} \mathcal{L}V_1(\tau,x) = -\mathcal{L}V_0(\tau,x) + \mathcal{A}V_0(\tau,x), \\ V_1(0,x) = g(x) - V_0(0,x). \end{cases} \tag{5.3.9}$$

Obviously, the right hand of Equation (5.3.9) is known to us, which indicates it is a linear PDE. Therefore, $V_1(\tau,x)$ can be obtained by solving such a linear PDE.

Then we differentiate the PDE equation (5.3.4) twice and set $p=0$, leading to

$$\begin{cases} \mathcal{L}V_2(\tau,x) = 2\dfrac{\partial}{\partial p}(\mathcal{A}\bar{V}(\tau,x,p))|_{p=0}, \\ V_2(0,x) = 0. \end{cases} \tag{5.3.10}$$

Although the right hand of equation is in terms of the unknown function $\bar{V}(\tau,x,p)$ and appears to be very complicated, it is actually a known function. To demonstrate this fact briefly, we choose $\rho=0$ and utility function $U(c)=2\sqrt{c}$ without loss the generality. Then the nonlinear operator becomes

$$\mathcal{A}V(\tau,x) = -(\frac{\partial U}{\partial c})^{-1}(V_x)V_x - \frac{\lambda^2}{2}\frac{V_x^2}{V_{xx}} + U(c^*) = \frac{1}{V_x} - \frac{\lambda^2}{2}\frac{V_x^2}{V_{xx}}. \tag{5.3.11}$$

It follows from chain rules that

$$\begin{aligned} \frac{\partial}{\partial p}(\mathcal{A}\bar{V})|_{p=0} &= \frac{\partial}{\partial p}[\frac{1}{\bar{V}_x} - \frac{\lambda^2}{2}\frac{\bar{V}_x^2}{\bar{V}_{xx}}]|_{p=0} \\ &= [-\frac{\partial^2\bar{V}}{\partial x\partial p}\frac{1}{\bar{V}_x^2} - \frac{\lambda^2}{2}(\frac{\partial\bar{V}_x^2}{\partial p}\bar{V}_{xx} - \frac{\partial\bar{V}_{xx}}{\partial p}\bar{V}_x^2)\frac{1}{\bar{V}_{xx}^2}]|_{p=0} \\ &= -\frac{\partial V_1}{\partial x}\frac{1}{(\frac{\partial V_0}{\partial x})^2} - \frac{\lambda^2}{2}[2\frac{\partial V_0}{\partial x}\frac{\partial V_1}{\partial x}\frac{\partial^2 V_0}{\partial x^2} - \frac{\partial^2 V_1}{\partial x^2}(\frac{\partial V_0}{\partial x})^2]\frac{1}{(\frac{\partial^2 V_0}{\partial x^2})^2}. \end{aligned}$$

The last equality holds because $\bar{V}(\tau,x,p)$ is assumed to be analytic with respect to p and expressed in Equation (5.3.7). Since V_0 and V_1 are both known functions, the PDE system (5.3.10) now becomes linear. When ρ is non-zero and the utility function takes another form, the calculations above may become more complicated but it can also be demonstrated similarly that the PDE system (5.3.10) is linear.

Generally, after differentiating the PDE system (5.3.4) n times and setting $p = 0$, we have

$$\begin{cases} \mathcal{L}V_n(\tau, x) = n\dfrac{\partial^{n-1}}{\partial p^{n-1}}(\mathcal{A}\bar{V}(\tau, x, p))|_{p=0}, \\ V_n(0, x) = 0. \end{cases} \tag{5.3.12}$$

Accordingly, the right hand of Equation (5.3.12) only depends on the previous known functions, $\{V_0(\tau, x), V_1(\tau, x), \cdots, V_{n-1}(\tau, x)\}$, which implies that $V_n(\tau, x)$ is governed by a linear PDE.

In a brief summary, for any fixed n, $V_n(\tau, x)$ satisfies the linear PDE

$$\begin{cases} \mathcal{L}V_n(\tau, x) = f_n(\tau, x), \\ V_n(0, x) = \psi_n(x), \qquad n \geq 1, \end{cases} \tag{5.3.13}$$

where

$$f_n(\tau, x) = \begin{cases} -\mathcal{L}V_0(\tau, x) + \mathcal{A}V_0(\tau, x), & \text{if } n = 1, \\ n\frac{\partial^{n-1}}{\partial p^{n-1}}(\mathcal{A}\bar{V}(\tau, x, p))|_{p=0}, & \text{if } n > 1, \end{cases}$$

and

$$\psi_n(x) = \begin{cases} g(x) - V_0(0, x), & \text{if } n = 1, \\ 0, & \text{if } n > 1. \end{cases}$$

Remark 5.3.1. The crux of the homotopy analysis method (HAM) is to decompose the nonlinear PDE into an infinite series of linear PDEs. Although function $f_n(\tau, x)$ appears to be written in terms of an unknown function $\bar{V}(\tau, x, p)$, it should be remarked that $f_n(\tau, x)$ is actually a known function because the process of taking $n-1$ times partial derivative with respect to p and then setting $p = 0$ has made all the unknown terms $\{V_n(\tau, x), V_{n+1}(\tau, x), V_{n+2}(\tau, x), \cdots\}$ disappear. In other words, $f_n(\tau, x)$ only depends on the known functions $\{V_0(\tau, x), V_1(\tau, x), \cdots, V_{n-1}(\tau, x)\}$ when it is calculated! Of course, such a computation of $f_n(\tau, x)$ would not have been possible without the aid of symbolic calculation software such as the Maple 17 that we adopted to carry out all the cumbersome calculations presented in this chapter.

Now, we apply the method of characteristics to solve these linear PDEs. The characteristics

equations associated with the PDE system (5.3.13) are as follows:

$$\begin{cases} \frac{d\tau}{ds} = 1, \\ \frac{dx}{ds} = -rx, \\ \frac{dV_n}{ds} = f_n(\tau(s), x(s)), \end{cases} \tag{5.3.14}$$

with initial conditions

$$\begin{cases} \tau(0) = 0, \\ x(0) = x_0, \\ V_n(0, x_0) = \psi_n(x_0). \end{cases} \tag{5.3.15}$$

After some simple calculations, the solution for the PDE system (5.3.13) is obtained as follows:

$$V_n(\tau, x) = \psi_n(xe^{r\tau}) + \int_0^\tau f_n(s, xe^{r(\tau-s)})ds. \tag{5.3.16}$$

Both functions ψ_n and f_n are known for each fixed n, the expression (5.3.16) is explicit and analytical, which can also be implemented in Maple 17.

In the theory of the homotopy analysis method, the initial guess, $V_0(\tau, x)$, can be virtually any continuous function, which provides us great freedom to choose an initial point. In this chapter, the initial guess $V_0(\tau, x)$ is given by the following linear PDE system

$$\begin{cases} \mathcal{L}V_0(\tau, x) = 0, \\ V_0(0, x) = g(x). \end{cases} \tag{5.3.17}$$

This PDE system can be solved analytically with the solution being $V_0(\tau, x) = g(xe^{r\tau})$. Such an initial guess makes the initial condition $\psi_n(x)$ vanish for any n, which simplifies the linear PDE system (5.3.13) and the solution (5.3.16).

The following proposition shows that our series solution constructed through the HAM is indeed an exact and explicit solution to the HJB equation once it converges.

Proposition 5.3.1. If the series

$$\bar{V}(\tau, x, 1) = V_0(\tau, x) + \sum_{n=1}^{+\infty} \frac{V_n(\tau, x)}{n!} \tag{5.3.18}$$

is convergent, where $V_n(\tau, x)$ is defined by Equation (5.3.16), it is indeed an exact solution of the PDE system (5.3.1).

Proof. For brevity, we define the vector

$$\mathbf{V_n} = \{V_0(\tau, x), V_1(\tau, x), V_2(\tau, x), \cdots, V_n(\tau, x)\}. \tag{5.3.19}$$

Differentiating the deformation equations (5.3.4) n times with respect to the embedding parameters p and setting $p = 0$, we have the so-called nth-order deformation equation

$$\mathcal{L}[\frac{V_n(\tau, x)}{n!} - \mathbf{1}_{n>1}\frac{V_{n-1}(\tau, x)}{(n-1)!}] = -\frac{1}{(n-1)!}R_n(\mathbf{V_{n-1}}, \tau, x), \tag{5.3.20}$$

where

$$R_n(\mathbf{V_{n-1}}, \tau, x) = \frac{\partial^{n-1}}{\partial p^{n-1}}[(\mathcal{L} - \mathcal{A})\bar{V}(\tau, x, p)]|_{p=0}. \tag{5.3.21}$$

After summing up all the terms, we have

$$\begin{aligned}
&-\sum_{n=1}^{+\infty} \frac{1}{(n-1)!}R_n(\mathbf{V_{n-1}}, \tau, x) \\
&= \sum_{n=1}^{+\infty} \mathcal{L}[\frac{V_n(\tau, x)}{n!} - \mathbf{1}_{n>1}\frac{V_{n-1}(\tau, x)}{(n-1)!}] \\
&= \mathcal{L}\sum_{n=1}^{+\infty}[\frac{V_n(\tau, x)}{n!} - \mathbf{1}_{n>1}\frac{V_{n-1}(\tau, x)}{(n-1)!}] \\
&= \mathcal{L}(\lim_{n\to+\infty} \frac{V_n(\tau, x)}{n!}).
\end{aligned}$$

If the series (5.3.18) converges, we have

$$\lim_{n\to+\infty} \frac{V_n(\tau, x)}{n!} = 0.$$

As a result, we obtain

$$\sum_{n=1}^{+\infty} \frac{1}{(n-1)!} R_n(\mathbf{V_{n-1}}, \tau, x) = 0. \tag{5.3.22}$$

On the other hand, according to Equation (5.3.21), we have

$$\begin{aligned} &\sum_{n=1}^{+\infty} \frac{1}{(n-1)!} R_n(\mathbf{V_{n-1}}, \tau, x) \\ &= \sum_{n=1}^{+\infty} \frac{1}{(n-1)!} \frac{\partial^{n-1}}{\partial p^{n-1}} [(\mathcal{L} - \mathcal{A}) \bar{V}(\tau, x, p)]|_{p=0} \\ &= \sum_{n=0}^{+\infty} \frac{1}{n!} \frac{\partial^n}{\partial p^n} [(\mathcal{L} - \mathcal{A}) \bar{V}(\tau, x, p)]|_{p=0}. \end{aligned} \tag{5.3.23}$$

According to Equations (5.3.22) and (5.3.23), we have

$$\sum_{n=0}^{+\infty} \frac{1}{n!} \frac{\partial^n}{\partial p^n} [(\mathcal{L} - \mathcal{A}) \bar{V}(\tau, x, p)]|_{p=0} = 0. \tag{5.3.24}$$

For a general p, $\bar{V}(\tau, x, p)$ does not perfectly satisfy the nonlinear PDE (5.3.1). Let denote the residual error of Equation (5.3.1) as

$$\varepsilon(\tau, x, p) = (\mathcal{L} - \mathcal{A}) \bar{V}(\tau, x, p). \tag{5.3.25}$$

Obviously, when the residual error $\varepsilon(\tau, x, p)$ becomes zero for some p, it corresponds to an exact solution of PDE system (5.3.1). The Taylor series of the residual error $\varepsilon(\tau, x, p)$ with respect to the embedding parameter p is

$$\varepsilon(\tau, x, p) = \sum_{n=0}^{+\infty} \frac{p^n}{n!} \frac{\partial^n}{\partial p^n} \varepsilon(\tau, x, p)|_{p=0} = \sum_{n=0}^{+\infty} \frac{p^n}{n!} \frac{\partial^n}{\partial p^n} [(\mathcal{L} - \mathcal{A}) \bar{V}(\tau, x, p)]|_{p=0}. \tag{5.3.26}$$

When $p = 1$, the above expression becomes, using (5.3.24),

$$\varepsilon(\tau, x, 1) = \sum_{n=0}^{+\infty} \frac{1}{n!} \frac{\partial^n}{\partial p^n} \varepsilon(\tau, x, p)|_{p=0} \sum_{n=0}^{+\infty} \frac{1}{n!} \frac{\partial^n}{\partial p^n} [(\mathcal{L} - \mathcal{A}) \bar{V}(\tau, x, p)]|_{p=0} = 0. \tag{5.3.27}$$

It means that $\bar{V}(\tau, x, p)$ becomes an exact solution of the original PDE system (5.3.1) when p

is set to be 1. Thus, if the series

$$\bar{V}(\tau, x, 1) = V_0(\tau, x) + \sum_{n=1}^{+\infty} \frac{V_n(\tau, x)}{n!} \tag{5.3.28}$$

is convergent, it is indeed a solution of the original PDE system (5.3.1). This completes the proof. □

Remark 5.3.2. Proposition 5.3.1 ensures that the series solution (5.3.18) is indeed an exact solution to the original PDE system (5.3.1) when it is convergent. It is still an open problem to prove the convergence of such a series solution when we apply the homotopy analysis method to solve nonlinear ODEs or PDEs. However, this method has been successfully and widely used to solve a number of nonlinear ODEs and PDEs in physics, engineering and quantitative finance (Ayub et al. 2003, Liao 2003*b*, Zhu 2006, Abbasbandy & Zakaria 2008, Zhao & Wong 2012), with a rigorous proof of the convergence being left as a future mathematical challenge. Here, we follow Zhu (2006) and Zhao & Wong (2012) to use numerical experiments to demonstrate the convergence of the series.

The series solution is considered to be convergent if and only if the sequence of partial sums $\{S_n\}_{n\geq 0}$ is a Cauchy sequence. The partial sum with n terms is defined as

$$S_n = \sum_{i=0}^{n} \frac{V_i(t, x)}{i!}. \tag{5.3.29}$$

In this chapter, we try to demonstrate that the sequence of partial sums $\{S_n\}_{n\geq 0}$ is convergent through the numerical evidence in next section.

The series solution constructed through the homotopy analysis method (HAM) may be not as elegant and simple as those produced by Merton. However, its significant contribution is that it works for general utility functions, including, but not limited to the cases where Merton had solved. In addition, the infinite series solution is exact and explicit because no discretization errors are introduced. To verify our series solution, four examples with different utility functions are demonstrated in the next section.

5.4 Examples

In this section, four examples are presented to demonstrate the accuracy and versatility of the homotopy analysis method (HAM) for the nonlinear HJB equation. The accuracy of our solution approach is demonstrated through the first two examples, the classical Merton problem subject to power utility function (Example 1) and logarithmic utility function (Example 2). Our series solution is compared with Merton's solution to serve the purpose of verifying the homotopy analysis method (HAM). Then, two more utility functions are chosen to demonstrate the versatility of the HAM in Examples 3 and 4. The main reason to select them is that, up to date, there is still no analytic solution for the Merton problem presented in Examples 3 and 4. Using our explicit series solution, the corresponding optimal investment policies can be easily derived, with which some interesting economic interpretations of the optimal investment strategy can be articulated.

One thing we need to point out is that although the two utility functions U and B can be of different forms as stated in Equation (5.2.4), it is reasonable from an economic point of view that they are taken the same form as the risk aversion associated with one particular investor must be consistent. In all these following examples, utility function B is assumed to take the same form with the utility function U.

It is also pointed out that all the computations reported in this chapter were performed with Maple 17 on 64-bit quad-core Intel 2.83GHz system with 16GB of RAM.

5.4.1 Example 1: power utility

In the first example, we consider the power utility function which is defined as:

$$U(x) = \frac{x^\gamma}{\gamma}, \tag{5.4.1}$$

where $\gamma < 1$ and $\gamma \neq 0$. The Arrow-Pratt measure of relative risk aversion of such a utility is defined as

$$\delta[U(\cdot)] = -\frac{U''(x)}{U'(x)}x = 1 - \gamma. \tag{5.4.2}$$

The power utility function (5.4.1) belongs to the CRRA class as its Arrow-Pratt measure is constant. The smaller the value of γ is, the more risk-aversion the investor shows. The HJB equation arising from the Merton problem subject to such a utility function is

$$\begin{cases} V_t + rxV_x - \dfrac{\lambda^2}{2}\dfrac{V_x^2}{V_{xx}} + \dfrac{1-\gamma}{\gamma}(\dfrac{\partial V}{\partial x})^{\frac{\gamma}{\gamma-1}} e^{\frac{\rho t}{\gamma-1}} = 0, \\ V(T,x) = e^{-\rho T}\dfrac{x^\gamma}{\gamma}, \quad \forall (t,x) \in [0,T] \times [0,\infty). \end{cases} \tag{5.4.3}$$

It is well known that there is a closed-form solution derived by Merton (1969) with the value function being of the simple form

$$V(t,x) = e^{-\rho t}\frac{x^\gamma}{\gamma} b(t), \tag{5.4.4}$$

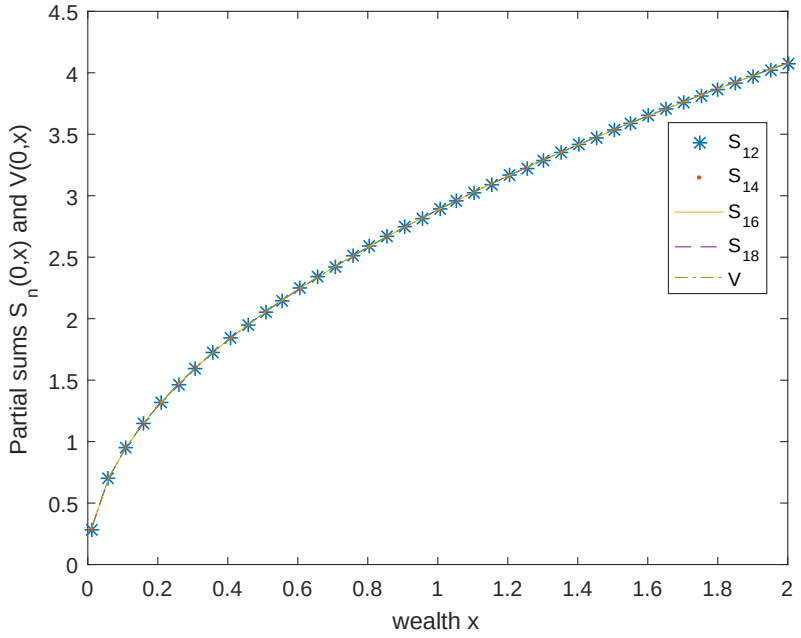
where

$$\begin{cases} b(t) = [\dfrac{1+(\nu-1)e^{-\nu(T-t)}}{\nu}]^{1-\gamma}, \\ \nu = \dfrac{1}{1-\gamma}\{\rho - \gamma[r + \dfrac{(\mu-r)^2}{2\sigma^2}(1-\gamma)]\}. \end{cases} \tag{5.4.5}$$

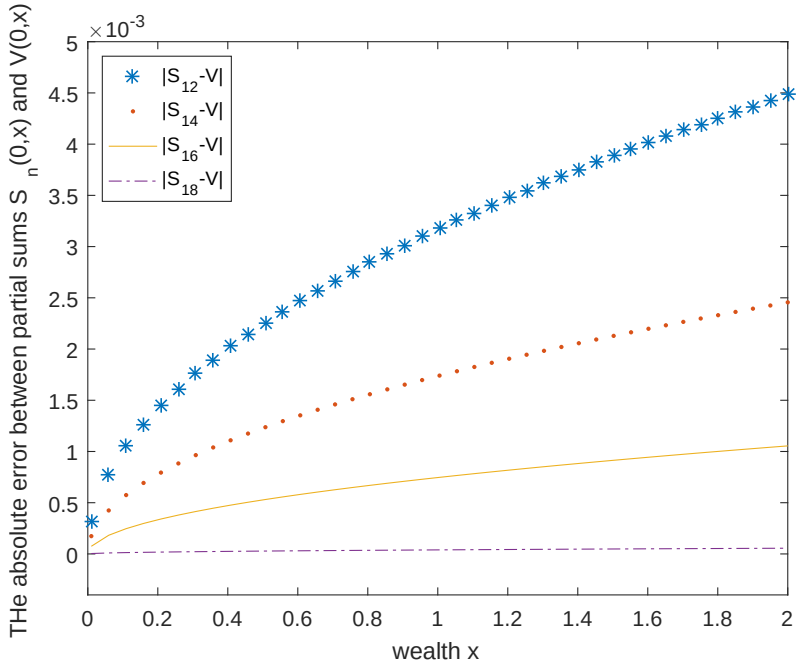
Such a simple solution will be adopted as a benchmark to compare with our series solution in order to demonstrate the accuracy of the latter. Without loss of generality, we have taken ρ to be zero. The other parameters are set as

$$\mu = 0.1, \quad r = 0.05, \quad \sigma = 0.5, \quad T = 1, \quad \gamma = 0.5. \tag{5.4.6}$$

Figure 5.1(a) displays the values obtained from the partial sums of the explicit series solution and those calculated from the analytical solution (5.4.4) at time $t = 0$. These curves are indistinguishable in Figure 5.1(a), which shows that the explicit series solution converges to the analytical solution exactly. To demonstrate the convergence more clearly, the absolute error between our series solution and the analytical solution (5.4.4) is shown in Figure 5.1(b). The absolute error diminishes gradually with the number of terms n increasing. In this example, we carried out the summation up to 18 terms, when a converged value function is obtained, inclusion of more terms in the solution resulted in a contribution in the order of 10^{-4}.



(a) Comparisons between $S_n(0,x)$ and $V(0,x)$.



(b) The absolute error between $S_n(0,x)$ and $V(0,x)$.

Figure 5.1: Convergence of solution when utility function is $\frac{x^\gamma}{\gamma}$ with $\gamma = 0.5$.

Through our symbolic calculation, the optimal investment proportion is produced as

$$u^* = -\frac{\mu - r}{x\sigma^2}\frac{\bar{V}_x(t,x,1)}{\bar{V}_{xx}(t,x,1)} = \frac{\mu - r}{(1-\gamma)\sigma^2}, \tag{5.4.7}$$

which is perfectly consistent with that calculated based on the analytical solution (5.4.4). Obviously, the optimal investment policy is independent of wealth and time for this case. It implies that the investor does not need to change the proportion invested in the risky asset. Such a constant proportion investment policy results from the fact that the investor guided by such a power utility function shows constant relative risk aversion, which implies that his attitude to the financial risk is independent of his wealth Merton (1969).

5.4.2 Example 2: logarithmic utility

In addition to power utility function, another classical one is logarithmic utility function defined as

$$U(x) = \ln x. \tag{5.4.8}$$

The Arrow-Pratt measure of relative risk aversion of such a utility function is

$$\delta[U(\cdot)] = -\frac{U''(x)}{U'(x)}x = 1, \tag{5.4.9}$$

which implies that it also belongs to the CRRA class. The investor does not change his attitude to the risk as his wealth varies from time to time.

The corresponding HJB equation in this example is written as

$$\begin{cases} V_t + rxV_x - \dfrac{\lambda^2}{2}\dfrac{V_x^2}{V_{xx}} - e^{-\rho t}(1 + \rho t + \ln V_x) = 0, \\ V(T,x) = e^{-\rho T}\ln x, \quad \forall (t,x) \in [0,T] \times [0,\infty). \end{cases} \tag{5.4.10}$$

The analytical solution to such a HJB equation has been produced by different authors (Merton 1969, Karatzas & Shreve 1998). With $\rho = 0$, the solution to the PDE system (5.4.10) is

expressed as

$$V(t,x) = (T-t+1)\ln\frac{x}{T-t+1} + [r+\frac{\lambda^2}{2}][(T+1)(T-t) - \frac{T^2-t^2}{2}], \tag{5.4.11}$$

which is chosen as the benchmark to verify our series solution. To calculate the values from the expression (5.4.11), the other parameters are taken as

$$\mu = 0.1, \quad r = 0.05, \quad \sigma = 0.5, \quad T = 1. \tag{5.4.12}$$

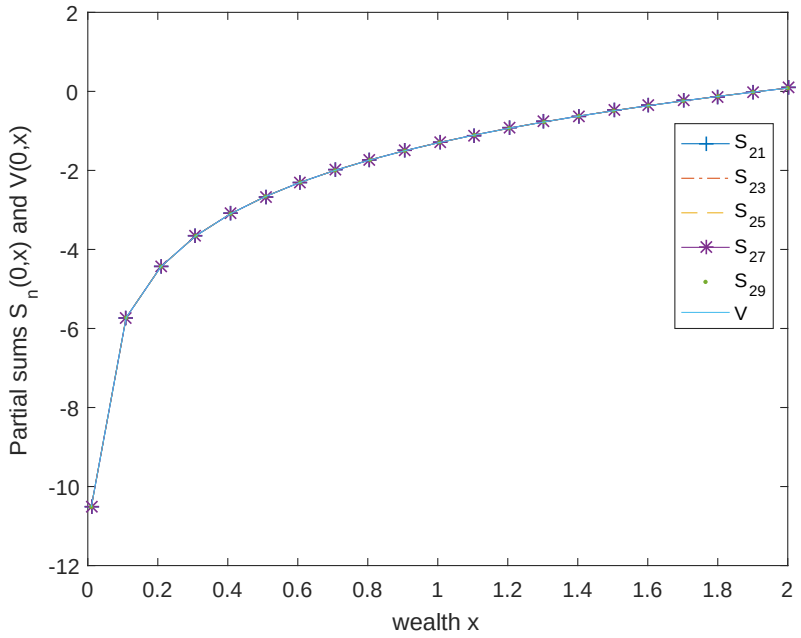
Figure 5.2(a) shows the results from the partial sums of our series solution and those from the analytical solution (5.4.11) at time $t = 0$. The convergence is demonstrated again as the number of terms n increases in Figure 5.2(b). In this case, the summation is carried out up to 29 terms so that the absolute error between our explicit series solution and the solution (5.4.11) has reached the level of 10^{-4}.

Again, through the complicated symbolic calculation, the optimal investment proportion is produced as

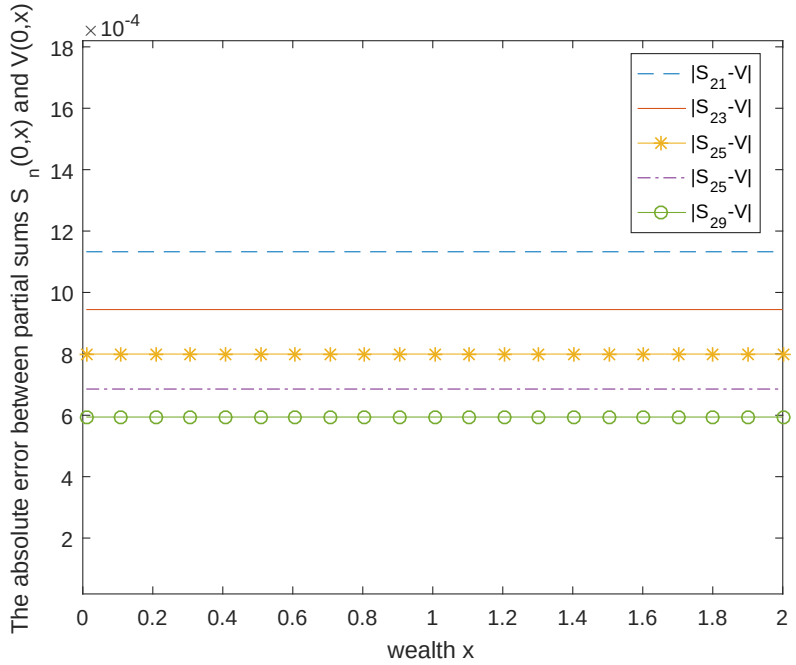
$$u^* = -\frac{\mu - r}{x\sigma^2}\frac{\bar{V}_x(t,x,1)}{\bar{V}_{xx}(t,x,1)} = \frac{\mu - r}{\sigma^2}, \tag{5.4.13}$$

which matches very well with that calculated from the analytical solution (5.4.11). The logarithmic utility function actually is considered as the limiting form of power utility function as γ tends towards zero (Merton 1969). An investor with the logarithmic utility function also shows constant relative risk-aversion and his attitude to the risk does not change with his wealth. Consequently, the optimal investment policy for such an investor is also constant proportion invested in the risky asset.

Both Example 1 and Example 2 are classical cases where many researchers have done some great work. Some analytical solutions have been obtained by different authors with specific methods (Merton 1969, Karatzas & Shreve 1998). In this chapter, such analytical solutions are used as a verification of our series solution and a test to the number of reasonable terms needed to produce numerical values with a desirable accuracy, such as 10^{-4}. From the experiments



(a) Comparisons between $S_n(0,x)$ and $V(0,x)$.



(b) The absolute error between $S_n(0,x)$ and $V(0,x)$.

Figure 5.2: Convergence of solution when utility function is $\ln x$.

above, one can clearly observe that the absolute error between our explicit series solution and the analytical solution would decrease to 10^{-4} as n increases, although the number of terms they need differs with different utility functions. While it only needs 18 terms for the series solution in Example 1, it takes 29 terms to achieve almost the same accuracy in Example 2. Clearly, the number of terms needed to achieve a particular level of accuracy depends on a specific form of the utility function, as one would expect.

In addition, it is worthwhile to point out that the utility functions in Example 1 and Example 2 share a common feature that both of them belong to the CRRA class. A very important character of an investor with such type of utility functions is that they do not change their attitude towards the risk when their total wealth moves up and down. As a result, their optimal investment policy is characterized by a constant proportion invested in the risky asset.

5.4.3 Example 3: exponential utility

In both of the previous two examples, the Arrow-Pratt measure of relative risk aversion for the utility function is constant, which implies that both of the utility functions belong to the CRRA class. In this example,we consider an exponential utility function defined as

$$U(x) = -\frac{e^{-\eta x}}{\eta}, \eta > 0, \tag{5.4.14}$$

to demonstrate that our solution approach could deal with the cases with non-CRRA utility function successfully. As shown in Merton (1969), the Arrow-Pratt measure of relative risk aversion for this case is:

$$\delta[U(\cdot)] = -\frac{U''(x)}{U'(x)}x = \eta x, \tag{5.4.15}$$

which implies that the investor shows more risk-aversion as his wealth x becomes larger. As a result, we expect that this will lead to a different investment behavior when the investor tries to maximize his expected utility.

Prior to our series solution presented in this chapter, an analytical solution can only be obtained in some special cases. For examples, the time horizon is assumed to be infinite (Merton

1969) or the utility function is only defined on the terminal wealth without consumption (Henderson 2005). We apply our homotopy analysis method to solve the Merton problem defined on the finite horizon and the consumption is still taken into account.The HJB equation can be written as

$$\begin{cases} V_t + rxV_x - \dfrac{\lambda^2}{2}\dfrac{V_x^2}{V_{xx}} - \dfrac{V_x}{\eta} + \dfrac{\ln V_x}{\eta}V_x + \dfrac{\rho t}{\eta}V_x = 0, \\ V(T,x) = -e^{-\rho T}\dfrac{e^{-\eta x}}{\eta}, \quad \forall (t,x) \in [0,T] \times [0,\infty). \end{cases} \tag{5.4.16}$$

Unlike the two previous examples where the analytical solutions have been obtained in literature, there is still no analytical solution which can be used as the benchmark. Without a benchmark, we show the absolute error between different partial sums numerically in order to demonstrate that our series solution is convergent. The parameters in this example are set as

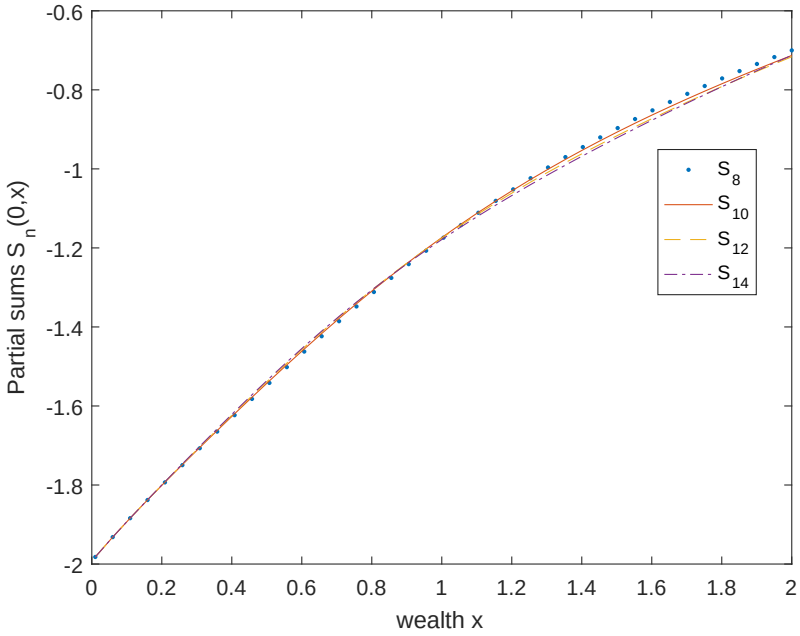
$$\mu = 0.1, \quad r = 0.05, \quad \sigma = 0.5, \quad T = 1, \quad \eta = 1. \tag{5.4.17}$$

Figure 5.3(a) displays the numerical results from some partial sums S_n at time $t = 0$. It is observed that these curves corresponding to different n values converge more and more into one as n increases. The absolute error between different partial sums is presented in Figure 5.3(b). When n reaches 14, the absolute error has arrived at the level of 10^{-3} already, which indicates a good numerical convergence.

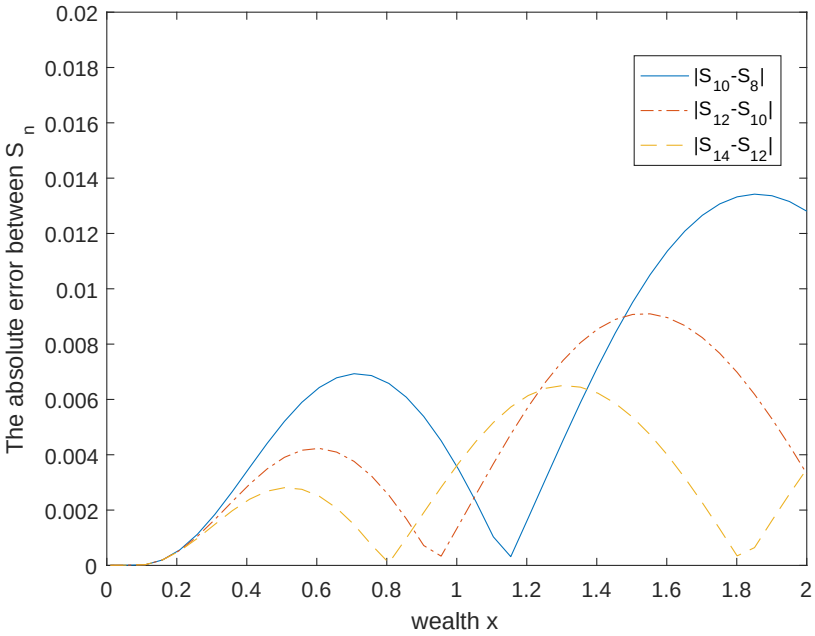
With the constructed series solution, the optimal investment policy can now be derived as

$$u^*(t,x) = -\frac{\mu - r}{x\sigma^2}\frac{\bar{V}_x(t,x,1)}{\bar{V}_{xx}(t,x,1)}, \tag{5.4.18}$$

which depends on wealth and time, instead of being constant in both two previous examples. Here we demonstrate how the optimal investment proportion changes with wealth and time in Figure 5.4(a). It is very interesting to observe from Figure 5.4(a) that the optimal investment proportion is a monotonically decreasing function of wealth; the wealthier an investor is, the lower proportion of his wealth he wants to invest in the risky asset. The investor allocates almost all his wealth on the risk-free asset when his wealth tends towards infinity. When the wealth goes to the other extreme, i.e. approaching zero, another interesting phenomenon



(a) Partial sums S_n.



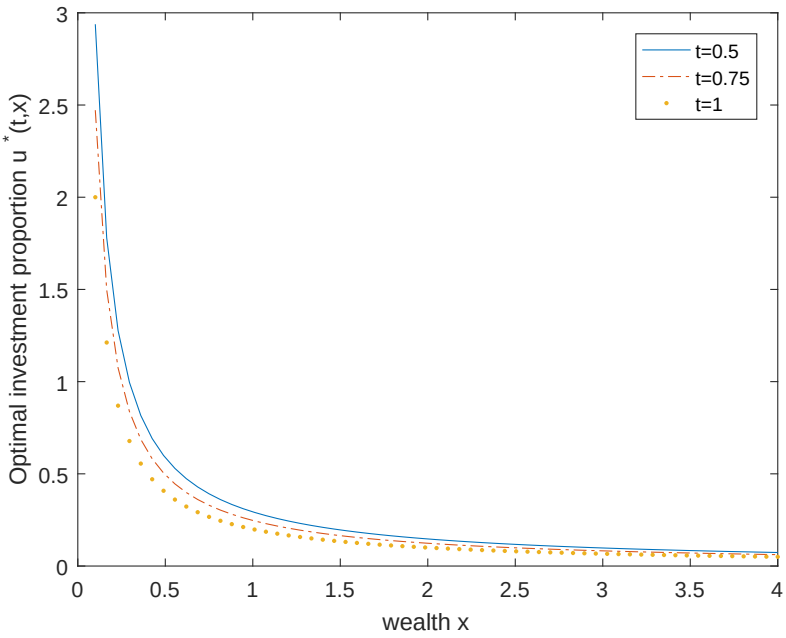
(b) The absolute error between S_n and S_{n-2} .

Figure 5.3: Convergence of solution when utility function is $U(x) = -\frac{e^{-\eta x}}{\eta}$ with $\eta = 1$.

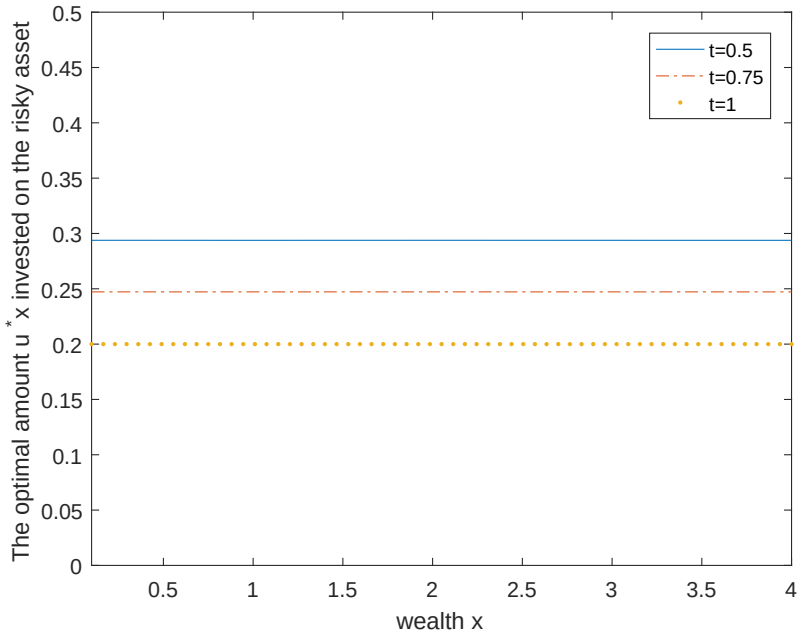
occurs. From Figure 5.4(a), one can clearly observe that the optimal investment proportion has already reached 1 when his wealth is sufficiently small and the proportion even becomes larger than 1, heading towards infinity when his wealth shrinks down to zero. In contrast to the cases in previous examples with a CRRA utility function, the optimal investment policy in this example is significantly different.

All these phenomena can be interpreted from an economic point of view. As shown in (5.4.15), the investor's relative risk aversion is a linear function of his wealth. On one hand, when his wealth increases, his relative risk aversion goes up simultaneously. As a result, he would allocate more money on the risk-free asset because he loathes risk more now. On the other hand, his risk aversion almost vanishes when his wealth shrinks towards zero. In this case, he prefers risky asset to risk-free asset. When his wealth is sufficiently small, the investment proportion becomes greater than 1, which implies that he has allocated more money than his net wealth on the risky asset by borrowing at the interest rate r. In other words, in order to maximize his expected utility, the investor would borrow as much as possible to invest on risky asset when his net wealth is zero. This rather weird behavior is permitted in Merton's model because both shorting the risky asset and unlimited borrowing from a bank at interest rate r are allowed.

Another interesting observation is that, in Figure 5.4(b), the amount invested in the risky asset ux is independent of the wealth but depends on time, unlike the cases presented in the first two examples in which the proportion of wealth (not the amount) invested in the risky asset is not only independent of the wealth itself but also independent of time. This observation is consistent with what Merton had observed when he considered the infinite-horizon case with the exponential utility function (Merton 1969). However, there are still some differences with Merton's infinite case. The significant one is that the amount invested in the risky asset changes with time in our finite-horizon case, which is actually the time effect; while such an effect is absent in Merton's infinite case.



(a) Optimal investment proportion u^*.



(b) Optimal investment amount u^*x.

Figure 5.4: The optimal investment policy when utility function is $U(x) = -\frac{e^{-\eta x}}{\eta}$ with $\eta = 1$.

5.4.4 Example 4: mixed power utility

All the utility functions presented in the previous examples can be labeled as the classical ones, on which extensive research has been done. Recently, Fouque et al. (2015) presented a new utility function of mixed form to characterize an investor whose relative risk aversion varies with his wealth. We demonstrate that the homotopy analysis method can deal with the nonlinear HJB equation with such a new utility function as well. Again, demonstrating the versatility of our solution approach through another form of utility function is crucially important because there should be no restrictions on the form of utility function in reality; investors with different risk aversion may choose different utility functions when they try to maximize their expected utility.

The utility function adopted in Example 4 is the mixture of two power utility functions taking the form

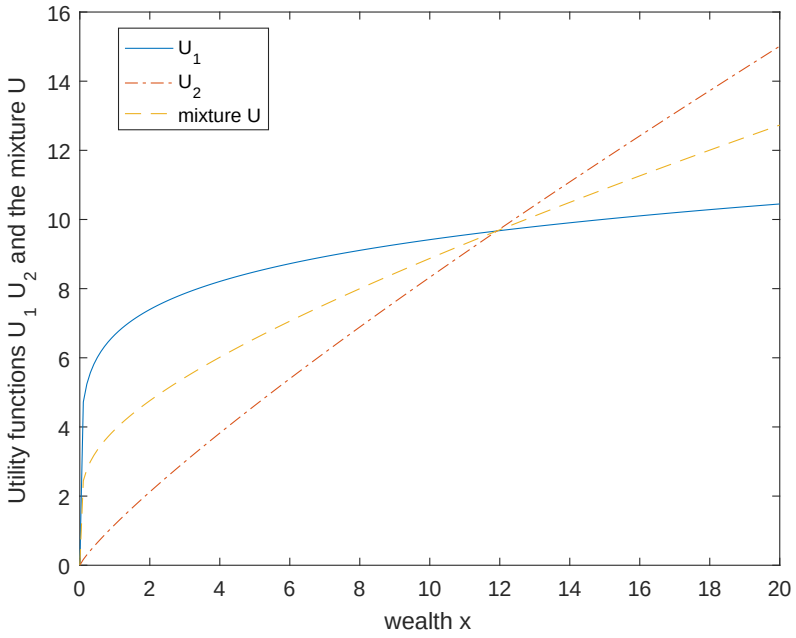
$$U(x) = kU_1 + (1-k)U_2 = k\frac{x^{\gamma_1}}{\gamma_1} + (1-k)\frac{x^{\gamma_2}}{\gamma_2}, k \in [0,1], \gamma_1 < \gamma_2 < 1, \gamma_{1,2} \neq 0. \tag{5.4.19}$$

The Arrow-Pratt measure of the mixed power utility function can be expressed as:

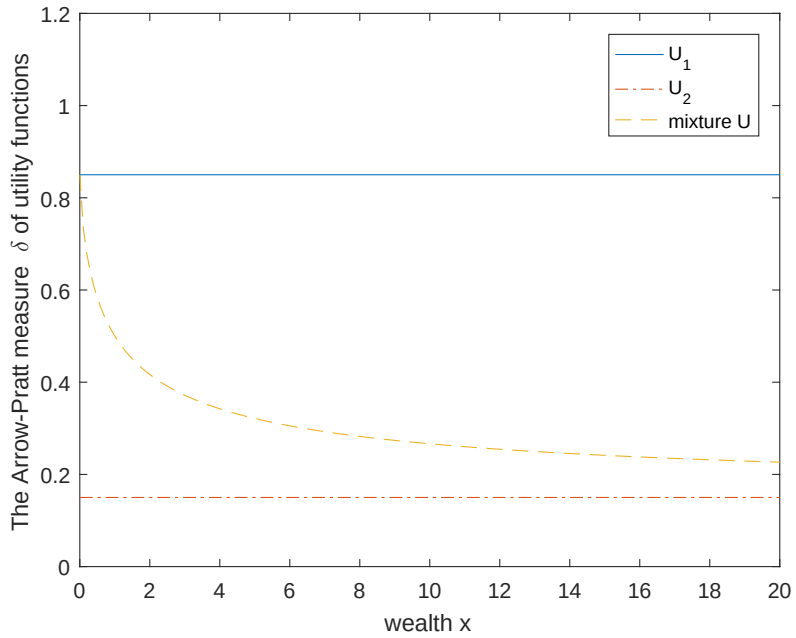
$$\delta[U(\cdot)] = \frac{U''(x)}{U'(x)}x = \frac{k(1-\gamma_1) + (1-k)(1-\gamma_2)x^{\gamma_2-\gamma_1}}{k + (1-k)x^{\gamma_2-\gamma_1}}. \tag{5.4.20}$$

It is interesting to note that although the new utility function (5.4.19) is just a simple sum of two power utility functions studied in Example 1, such a linear combination of power utility functions has led to a totally nonlinear result; the solution is by no means a simple sum of the results obtained from Example 1. Of course, this is basically due to the fact that the system we are dealing with is a fully nonlinear one. From the expression (5.4.20), the Arrow-Pratt measure now becomes a function of wealth, instead of being a constant in the case of single power utility function. On the other hand, the mixed power utility function degenerates to a single power utility function U_1 (or U_2) when we set $k = 1$ (or $k = 0$). One can appreciate more of the complexity associated with the mixture of two simple power utilities through their

visualization shown in Figure 5.5(a) (a graphic display of U_1, U_2 and the mixture U) and Figure 5.5(b) (the corresponding Arrow-Pratt measure of relative risk aversion) with a set of parameters. Clearly, it can be observed in Figure 5.5(b) that both U_1 and U_2 are CRRA utility



(a) Utility function U_1, U_2 and the mixture U



(b) Arrow-Pratt risk aversion $\delta = -\frac{U''}{U'}x$

Figure 5.5: Mixed power utilities with $k = 0.5, \gamma_1 = 0.15, \gamma_2 = 0.85$.

functions, whereas the mixture of them displays a wealth-varying nature of risk aversion.

For the utility function of such a mixed form, there is no explicit solution even when we only consider to maximize the utility from the terminal wealth, which is actually a simplified Merton problem (Fouque et al. 2015). However, our homotopy analysis method can still yield an explicit series solution. In this case, the objective function can be mathematically stated as

$$\max_{u(\cdot)} e^{-\rho T} U(X_T). \tag{5.4.21}$$

The corresponding HJB equation in Example 4 becomes

$$\begin{cases} V_t + rxV_x - \frac{\lambda^2}{2}\frac{V_x^2}{V_{xx}} = 0, \\ V(T,x) = e^{-\rho T}[k\frac{x^{\gamma_1}}{\gamma_1} + (1-k)\frac{x^{\gamma_2}}{\gamma_2}], \quad \forall (t,x) \in [0,T] \times [0,\infty). \end{cases} \tag{5.4.22}$$

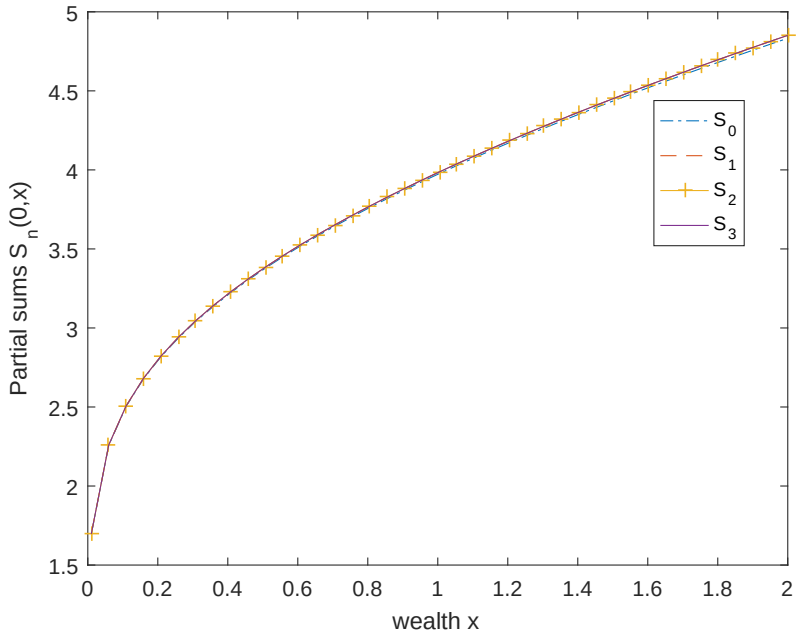
Without loss of generality, in the presentation of the following results, we have set ρ to zero. The other parameters are defined as follows:

$$\mu = 0.1, \quad r = 0.05, \quad \sigma = 0.5, \quad T = 1. \tag{5.4.23}$$

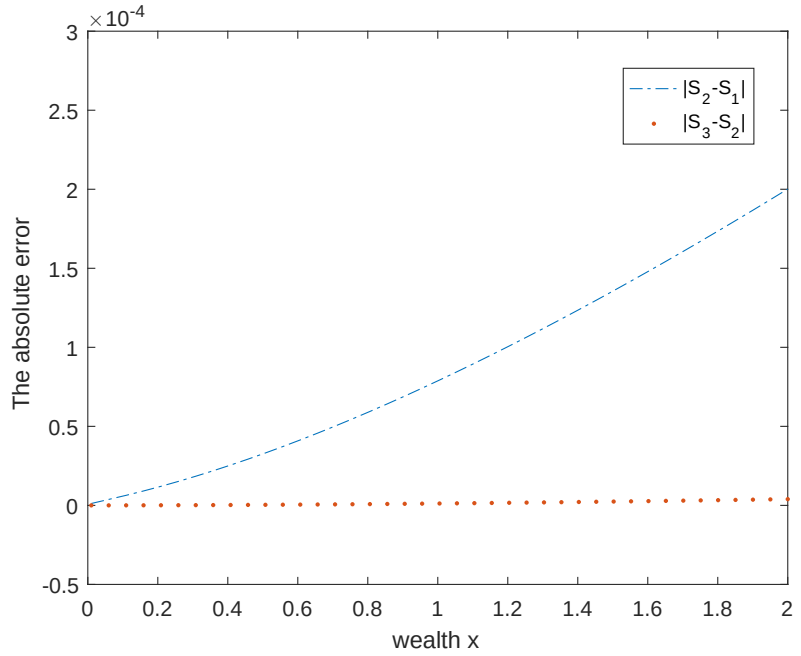
Numerical results are presented in following figures to demonstrate the convergence of the partial sums S_n at time $t = 0$.

Amazingly, the results obtained after summing up two terms and those with three terms are hardly distinguishable as shown in Figure 5.6(a). In other words, the series (5.3.18) is numerically convergent through our numerical results. From Figure 5.6(b), it is clear that the absolute difference $|S_3 - S_2|$ and $|S_2 - S_1|$ has already reached the level of 10^{-4}, which implies that a converged series solution accurate to the 4th decimal place has been obtained with meagerly two terms. We owe such a surprisingly fast convergence to the simplicity of our objective function (5.4.21) because only the utility from the terminal wealth X_T is considered in this example.

Unlike all the previous examples, our series solution exhibits an excellent convergence for this example with merely two terms! Thus, the explicit form of the truncated solution in this



(a) Partial sums $S_n(0, x)$.



(b) The absolute error between S_n and S_{n-1}.

Figure 5.6: Convergence of solution when utility function $U(x) = k\frac{x^{\gamma_1}}{\gamma_1} + (1-k)\frac{x^{\gamma_2}}{\gamma_2}$ with $k = 0.5, \gamma_1 = 0.15, \gamma_2 = 0.85$.

example could be even used as an approximation formula for the solution being used in hedging practice. The first two terms obtained in Maple 17 are

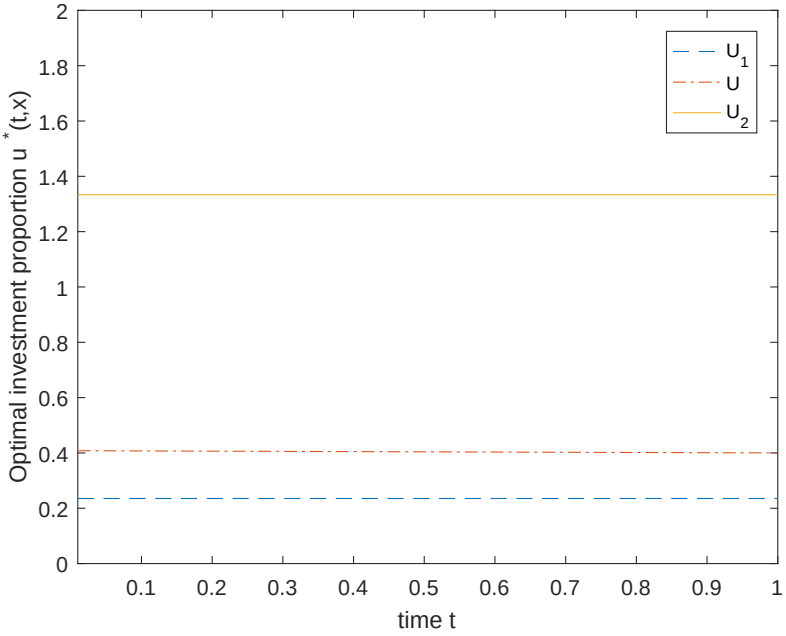
$$V_0(\tau, x) = \frac{k}{\gamma_1}(xe^{r\tau})^{\gamma_1} + \frac{1-k}{\gamma_2}(xe^{r\tau})^{\gamma_2}, \tag{5.4.24}$$

$$V_1(\tau, x) = \frac{C\tau[(1-k)(xe^{r\tau})^{\gamma_2} + k(xe^{r\tau})^{\gamma_1}]^2}{k(1-\gamma_1)(xe^{r\tau})^{\gamma_1} + (1-k)(1-\gamma_2)(xe^{r\tau})^{\gamma_2}}, \tag{5.4.25}$$

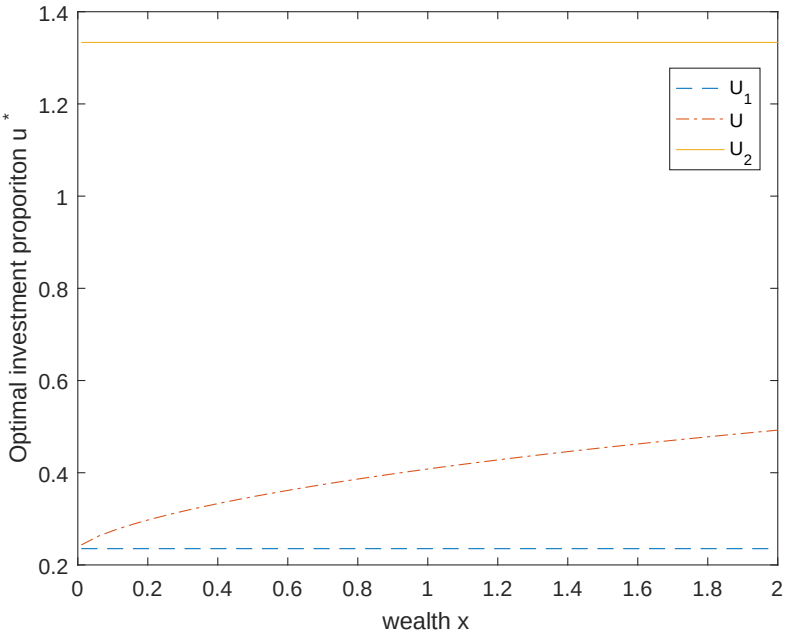
where $C = \frac{\lambda^2}{2}$. As a result, $S_1(\tau, x) = V_0(\tau, x) + V_1(\tau, x)$ can be regarded as a good approximation solution for the value function $V(\tau, x)$.

With the value function being approximated by $S_1(\tau, x)$, the optimal investment policy can be numerically calculated. The numerical results implies that the optimal investment policy also depends on wealth and time, instead of being constant. Here we demonstrate how the optimal investment policy changes with wealth and time in Figure 5.7. The variation of the optimal investment proportions associated with two power utility functions U_1, U_2 and their mixture U are displayed in Figure 5.7(a) with a fixed wealth. One can clearly observe that all of them are kept constant no matter how time varies. The optimal investment proportion associated with the mixed power utility function U is actually sandwiched between those associated with two power utility functions U_1 and U_2. This can be easily explained from the Arrow-Pratt measure for the mixed power utility function being actually bounded by those associated with the two power utility functions as shown in Figure 5.5(b). The more risk-aversion the investor shows, the less he invests in the risky asset. Therefore, the optimal investment proportion for U_2 is on the top; while that for U_1 is in the bottom.

On the other hand, for a fixed time t, the variation of the optimal investment policy in wealth x direction is plotted in Figure 5.7(b). As is pointed out in Example 1, the optimal investment proportions for U_1 and U_2 are still constant. However, the optimal investment proportion for the mixed utility function U is now a monotonically increasing function of wealth, instead of being constant. In other words, the investor allocates higher proportion of his wealth in risky asset as his wealth becomes larger, which is an absolutely different investment strategy with those in all previous examples. The reason for such an increasing optimal investment proportion with



(a) The variation with time.



(b) The variation with wealth.

Figure 5.7: The optimal investment proportion for mixed utility function.

wealth is that the relative risk aversion of an investor taking such a mixed utility is a decreasing function of wealth, as shown in Figure 5.5(b).

5.5 Conclusions

In this chapter, the fully nonlinear HJB equation that arises from the Merton problem subject to some general utility functions is solved with the homotopy analysis method (HAM). The nonlinear PDE is decomposed into an infinite series of linear PDEs, which can be solved analytically, leading to an explicit series solution.

Four examples are presented to demonstrate the accuracy and versatility of our solution approach. Through these examples, convergent solutions with a desirable accuracy can be achieved with enough terms being included in the series. Of course, the exact number of terms varies with the form of the utility function taken for each case. For some cases, it is amazingly fast to achieve an accuracy to the 4th decimal place with meagerly two terms! Furthermore, the fact that we have presented an analytic solution in series form for the cases where no analytic solutions have been reported in the literature alone suffices to demonstrate the versatility of our solution approach.

Chapter 6

A closed-form analytical solution to the HJB equation for the Merton problem defined on a finite horizon with exponential utility function

6.1 Introduction

Optimal investment and consumption problem (also referred to as the *Merton problem*) is a well-known and classic topic in mathematical finance. The key task of dynamic portfolio selection problem is to obtain the optimal investment and consumption policy for an investor to maximize his expected utility from intermediate consumption and terminal wealth. Merton (1969, 1971) formulated it as a stochastic optimal control problem in his seminal papers and solved it for the first time under the assumptions that the stock price follows a geometric Brownian motion and the utility function is constant relative risk aversion (CRRA). The optimal investment and consumption policies were both explicitly derived by Merton through solving the fully nonlinear HJB equation on a finite horizon. Exponential utility function, which shows constant absolute risk aversion (CARA), was also mentioned by Merton in the landmark papers and an

analytical solution corresponding to such a utility function was only obtained on an infinite horizon instead of a finite horizon. In fact, solving the Merton problem on a finite horizon is more useful because the investment horizon can never be infinite in reality. However, when exponential utility function is defined over both consumption and terminal wealth, finding a closed-form analytical solution on a finite horizon becomes a much harder problem, which remains unsolved in the past few decades.

The seminal work (Merton 1969, 1971) has established the framework for dynamic portfolio selection problem, although Merton's original model is a very simple one where both the return rate and volatility of the stock price are assumed to be constant. Since then, there is abundant literature on the extension of the classic Merton problem from different aspects which include, but are not limited to, adding stochastic risk premium, stochastic volatility, transaction costs, etc. While some authors focused on achieving closed-form analytical solutions, others resort to numerical solutions (Kushner 1990, Barles & Souganidis 1991, Fleming & Soner 2006) if finding a closed-form solution becomes formidably difficult. Since the focus of this chapter is to show a newly discovered analytical closed-form solution, we mainly focus our literature review on those papers in which closed-form solutions are presented.

With closed-form solutions as a common feature, the literature can be grouped into three different categories.

Papers in the first category adopt a basic assumption that the utility function belongs to the CRRA class. Zariphopoulou (1999) extended the Merton problem to the case in which the coefficients in the underlying diffusion process are arbitrary nonlinear functions and utility function is defined on both intermediate consumption and terminal wealth. Brennan & Xia (2002) adopted a stochastic model to characterize the term structure, while Wachter (2002) placed her focus on a model in which the risk premium follows an Ornstein-Uhlenbeck process. Using Malliavin derivatives, Detemple et al. (2005) explicitly solved the case where stochastic interest rate and investment constraints are taken into account. A further important extension with a multi-dimensional model was presented in Liu (2007), in which the so-called "quadratic" asset returns are introduced as a very general way to incorporate features such as stochastic

interest rate, stochastic return rate as well as stochastic volatility all under one umbrella.

In order to preserve analytic tractability, an important assumption shared by the papers in the second category is that consumption is not allowed and utility function is defined over terminal wealth only. Typical examples in this category include Kim & Omberg (1996) in which risk premium follows an Ornstein-Uhlenbeck process, Tehranchi (2004) where a random factor is introduced into the utility function to allow the terminal wealth be subject to another risk factor that is not necessarily correlated with the underlying, Henderson (2005) where an imperfectly judgeable stochastic income is added to the classic Merton problem to demonstrate how such an income affects the optimal allocation of wealth between the risky and risk-free assets.

In the last category, the investment horizon in the Merton problem is assumed to be infinite, which really has simplified the HJB equation to a nonlinear ordinary differential equation (ODE) and considerably facilitated the solution process. Major contributions are Merton (1971) where geometric Brownian motion hypothesis is replaced by an alternative price mechanism, Liu (2004) where transaction costs are taken into consideration and Chacko & Viceira (2005) where the investor has a recursive preference over intermediate consumption with unit elasticity of infratemporal substitution.

It should be remarked that analytic tractability is a result of the corresponding assumption made in each of these three categories. Consequently, removing all three assumptions listed above and, in the mean time, still obtaining a closed-form solution constitute a substantially difficult problem. Tackling such a difficulty is the objective and the key contribution of the current chapter, in which we show a new closed-form solution for the Merton problem defined on a finite horizon with consumption and terminal wealth both being included under a non-CRRA utility function[1].

[1]Exponential utility function, which is obviously non-CRRA, has been discussed many times in the literature. However, closed-form solutions are only available in the case where the investment horizon is assumed to be infinite (Merton 1969, Liu 2004) or where the utility function is only defined on the terminal wealth without consumption (Henderson 2005, Zeng & Taksar 2013). The importance of studying portfolio selection problems under exponential utility is further manifested by a very recently published paper (Xing 2017), in which some stability issues associated with the portfolio maximization problem are examined, although Xing's study still restricts the utility function on the terminal wealth only.

In order to achieve such an objective, we choose to focus on the classic Merton problem first instead of the generalized Merton problem. The main contribution of this chapter is that we present two different methods, with which closed-form analytical solutions are obtained for the Merton problem with exponential utility function being defined over both consumption and terminal wealth on a finite horizon. We also demonstrate that these two solutions obtained with different methods are equivalent to each other although they appear to be of different form. The solution approach adopted here to obtain the closed-form solution for such a problem may shed some light to the solution of some similar problems based on more sophisticated models in the future. In addition, we show a verification theorem that demonstrates the value function is indeed the solution to the original optimal stochastic control problem. With the availability of the newly derived closed-form solution, we are also able to carry out some economic and mathematical discussions on the optimal investment and consumption policy associated with this problem.

The rest of the chapter is organized as follows. In Section 6.2, the Merton problem and the corresponding HJB equation are briefly reviewed for the convenience of the readers. In Section 6.3, two methods are presented to solve the corresponding HJB equation with exponential utility function on a finite horizon. In Section 6.4, some economic interpretations of the optimal investment and consumption policies are discussed, respectively, while concluding remarks are provided in the last section.

6.2 The Merton problem and the HJB equation

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbf{P})$ be a complete filtered probability space. $\mathbf{P}$ denotes the physical or real-world measure and the filtration $(\mathcal{F}_t)_{t\geq 0}$ represents the history of the market. For simplicity, throughout this chapter, we assume a frictionless market and no dividends.

6.2.1 The Merton problem

We consider a financial market with two assets being traded continuously on a finite horizon $[0, T]$. One asset is a risk-free bond, whose price $\{P(t), t \geq 0\}$ evolves according to the differential equation

$$dP(t) = rP(t)dt, \qquad t \in [0, T], \tag{6.2.1}$$

with r being the risk-free interest rate. The other one is a risky asset with its price being modeled as a diffusion process $S(t)$ satisfying the following stochastic differential equation (SDE)

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \qquad t \in [0, T], \tag{6.2.2}$$

where μ is the drift rate, σ is the volatility coefficient and $W(t)$ is a standard Brownian motion.

An investor starts with a known initial wealth x_0 and the wealth at time t is denoted as $X(t)$. At any time t, prior to T, the investor needs to make a decision on how much to be consumed and, in the mean time, how much to be invested in the risky asset, in order to maximize his expected utility from the accumulated consumption and the terminal wealth. The consumption rate per unit time at time t is denoted as $c(t)$ and the investment proportion $u(t)$ represents the fraction of total wealth that is invested in the risky asset at time t. $1 - u(t)$ is thus the remaining left in form of the risk-free bond within the framework of this two-asset model. While the investment proportion on the risky asset $u(t)$ is allowed to be negative, which is financially interpreted as short selling, the remainder invested in the risk-free asset, $1 - u(t)$, may also become negative, which corresponds to unconstrained borrowing at the interest rate r.

With all these being taken into consideration, wealth process $X(t)$ follows the SDE:

$$dX(t) = u(t)X(t)\frac{dS(t)}{S(t)} + r[1 - u(t)]X(t)dt - c(t)dt, \tag{6.2.3}$$

that is

$$dX(t) = \{[r + u(t)(\mu - r)]X(t) - c(t)\}dt + X(t)u(t)\sigma dW(t). \tag{6.2.4}$$

The objective of the Merton problem is to obtain the optimal investment and consumption

policies, i.e. to determine $u(t)$ and $c(t)$, such that the expected utility from intermediate consumption and terminal wealth is maximized. Mathematically, such an objective functional is stated as:

$$\max_{(u,c)\in\mathcal{A}} \quad \mathbf{E}[\int_0^T e^{-\rho s}U(c(s))ds + e^{-\rho T}B(X_T)], \tag{6.2.5}$$

where $\mathbf{E}$ is the expectation operator; U is a function measuring the utility from consumption $c(t)$ and B is also a function (also referred to as *bequest function*) measuring the utility from the terminal wealth X_T.[2] ρ is the subjective discount rate or the subjective rate of time preference.[3] $\mathcal{A}$ is the class of all admissible investment and consumption strategy defined as follows:

Definition 6.2.1. An investment and consumption strategy pair (π, c) is admissible, if the following conditions are satisfied:

1. The investment strategy $\pi(t)$ is an $\mathbf{R}$ valued, $\mathcal{F}_t$ adapted process such that

$$\mathbf{E}\int_0^T |u(t)X(t)|^2 dt < \infty, \tag{6.2.6}$$

   where $\|\cdot\|$ is the Euclidean norm in $\mathbf{R^m}$;

2. The consumption strategy $c(t)$ is an $\mathbf{R}^+$ valued, $\mathcal{F}_t$ adapted process such that

$$\mathbf{E}\int_0^T |c(t)|dt < \infty, . \tag{6.2.7}$$

3. The corresponding objective function is finite, i.e.

$$|\mathbf{E}[\int_0^T e^{-\rho s}U(c(s))ds + e^{-\rho T}U(X(T)]| < \infty, \tag{6.2.8}$$

In summary, the Merton problem has now been reformulated as an optimal stochastic control problem with the objective functional (6.2.5), driven by the dynamics of the wealth (6.2.4). The

[2] In general, utility functions U and B are assumed to be of the same form.

[3] The fact that the subjective discount rate ρ is possibly different from the risk-free interest rate r results from the concept of the time preference in economics. In other words, each individual investor may choose a discount rate tailored for his own situation.

solution of such an optimal stochastic control problem is discussed in the next subsection.

6.2.2 The HJB equation

To seek a solution of the optimal stochastic control problem formulated in the previous subsection, we can adopt the dynamic programming method since wealth process $X(t)$ is obviously Markovian. The fundamental idea of the dynamic programming is to consider a family of optimal control problems with different initial times and states and establish relationships among these problems in the family, so that they can be solved with the aid of the so-called HJB equation.

Let $(t,x) \in [0,T) \times \mathbf{R}^+$ and consider the following control system over $[t,T]$

$$\begin{cases} dX(s) = \{[r + u(s)(\mu - r)]X(s) - c(s)\}ds + X(s)u(s)\sigma dW(s), \\ X(t) = x, \end{cases} \tag{6.2.9}$$

with the cost functional being defined as

$$\mathbf{J}(c(\cdot), u(\cdot); t, x) = \mathbf{E}[\int_t^T e^{-\rho s}U(c(s))ds + e^{-\rho T}B(X(T))]. \tag{6.2.10}$$

The value function is defined as

$$V(t,x) = \max_{(u,c)\in\Pi_1[t,T]\times\Pi_2[t,T]} \mathbf{J}(c(\cdot), u(\cdot); t, x). \tag{6.2.11}$$

According to the dynamic programming method (Yong & Zhou 1999), the value function $V(t,x)$ satisfies the HJB equation:

$$\begin{cases} \max\limits_{(u,c)\in\mathbf{R^2}} \phi(u, c; t, x) = 0, \\ V(T,x) = e^{-\rho T}B(x), \quad (t,x) \in [0,T] \times [0,\infty), \end{cases} \tag{6.2.12}$$

where

$$\phi(u,c;t,x)=\frac{\partial V}{\partial t}+\{[u(\mu-r)+r]x-c\}\frac{\partial V}{\partial x}+\frac{1}{2}x^2\sigma^2u^2\frac{\partial^2V}{\partial x^2}+e^{-\rho t}U(c). \tag{6.2.13}$$

The first-order necessary condition for a regular interior maximum of Equation (6.2.12) is

$$\begin{cases}\frac{\partial\phi}{\partial c}=-\frac{\partial V}{\partial x}+e^{-\rho t}\frac{\partial U}{\partial c}=0,\\ \frac{\partial\phi}{\partial u}=(\mu-r)x\frac{\partial V}{\partial x}+ux^2\sigma^2\frac{\partial^2V}{\partial x^2}=0.\end{cases} \tag{6.2.14}$$

The optimal pair (u^*,c^*) can be obtained in terms of the value function $V(t,x)$ as

$$\begin{cases}u^*=-\frac{\mu-r}{x\sigma^2}\frac{V_x}{V_{xx}},\\ c^*=(\frac{\partial U}{\partial c})^{-1}(e^{\rho t}V_x),\end{cases} \tag{6.2.15}$$

where $V_{xx}:=\frac{\partial^2V}{\partial x^2}$, $V_x:=\frac{\partial V}{\partial x}$ and $(\frac{\partial U}{\partial c})^{-1}$ is the inverse function of $\frac{\partial U}{\partial c}$[4].

On the other hand, the sufficient second-order condition for a regular interior maximum of Equation (6.2.12) is

$$\begin{cases}\phi_{uu}<0,\quad \phi_{cc}<0,\\ \det\begin{pmatrix}\phi_{uu} & \phi_{uc}\\ \phi_{uc} & \phi_{cc}\end{pmatrix}>0,\end{cases} \tag{6.2.16}$$

where $\phi_{uu}:=\frac{\partial^2\phi}{\partial u^2}$, $\phi_{cc}:=\frac{\partial^2\phi}{\partial c^2}$, and $\phi_{uc}:=\frac{\partial^2\phi}{\partial u\partial c}$. It is easy to verify that the mixed-derivative ϕ_{uc} is zero. The only thing we need to check is the sign of ϕ_{cc} and ϕ_{uu} in order to ensure the obtained value function $V(t,x)$ is indeed a solution to the original HJB equation (6.2.12).

Upon substituting the optimal policies (u^*,c^*) into the HJB equation (6.2.12), we obtain another PDE system

$$\begin{cases}V_t+rxV_x-\frac{(\mu-r)^2}{2\sigma^2}\frac{V_x^2}{V_{xx}}-(\frac{\partial U}{\partial c})^{-1}(e^{\rho t}V_x)V_x+e^{-\rho t}U(c^*)=0,\\ V(T,x)=e^{-\rho T}B(x),\quad (t,x)\in[0,T]\times[0,\infty),\end{cases} \tag{6.2.17}$$

[4]It is assumed that utility function U is smooth. Since utility function is concave, we have $\frac{\partial^2U}{\partial c^2}<0$. As a result, $\frac{\partial U}{\partial c}$ is a monotonic function and its inverse function $(\frac{\partial U}{\partial c})^{-1}$ exists.

where $V_t := \frac{\partial V}{\partial t}$. Although the maximization operator in the HJB equation (6.2.12) has been eliminated, the new PDE system (6.2.17) still remains nonlinear, which poses fundamental challenges for closed-form analytical solutions being discovered.

6.2.3 Utility function

Utility function, which is used to characterize an investor's preference, plays a vital role in economics. In expected utility theory, there exist various utility functions to describe different kinds of investors with diversified preferences. Among these numerous utility functions, power utility function $U(x) = \frac{x^\gamma}{\gamma}(\gamma < 1, \gamma \neq 0)$ and logarithmic utility function $U(x) = \ln x$ are critically important, not only for their simplicity and convenience in mathematics, but also for the perfect economic interpretations. To measure investors' risk-aversion, Pratt (1964) presented a functional measure for utility function. The Arrow-Pratt relative risk-aversion of power utility function and logarithmic utility function are both constant as follows:

$$\delta[U(\cdot)] = -\frac{U''(x)}{U'(x)}x = \begin{cases} 1-\gamma, & U(x) = \frac{x^\gamma}{\gamma}, \\ 1, & U(x) = \ln x. \end{cases}$$

Consequently, both power utility function and logarithmic utility function belong to the constant relative risk-aversion (CRRA) class. Due to the nice property of the CRRA class, the Merton problem with such a utility function has been solved perfectly by Merton (1969) for both the finite-horizon and the infinite-horizon cases.

In addition to the CRRA class, exponential utility function, defined as

$$U(x) = -\frac{e^{-\eta x}}{\eta}, \tag{6.2.18}$$

also needs to be explored. The Arrow-Pratt measure of relative risk-aversion for exponential utility function is

$$\delta[U(\cdot)] = -\frac{U''(x)}{U'(x)}x = \eta x, \tag{6.2.19}$$

which means the investor's relative risk-aversion is a monotonically increasing function of his wealth. The more wealth the investor has, the more risk-aversion he shows. Obviously, exponential utility function is a non-CRRA one.

Merton (1969) managed to solve the HJB equation with exponential utility function with an assumption that the investment horizon is infinite. However, the finite-horizon case remains unsolved due to the high nonlinearity of the HJB equation. Recently, Zhu & Ma (2018) applied the homotopy analysis method (HAM) to the finite-horizon case and obtained an explicit series solution. By observing the numerical results calculated from their truncated series solution, they found that the optimal investment amount is a monotonically function of time in the finite-horizon case, instead of constant in the infinite-horizon case. However, their series solution is still inadequate to explain the monotonicity mathematically and a closed-form solution with a sufficient degree of smoothness and thus differentiability is expected. After some failed attempts, we have finally managed to obtain such a closed-form solution through two distinct approaches: an indirect method and a direct method, which are demonstrated in next section in details.

6.3 The closed-form analytical solutions

In this section, we present two distinct approaches to derive closed-form analytical solutions: an indirect method and a direct method. Although the analytical solutions obtained with these two different methods appear to be of different forms, they are proved to be equivalent with each other finally.

6.3.1 Indirect method

In the indirect method, we first review another class of parameterized utility functions, the hyperbolic absolute risk-aversion (HARA) class, which is mathematically defined as

$$U(x) = \frac{1-\gamma}{\eta\gamma}(\alpha + \frac{\eta}{1-\gamma}x)^{\gamma}, \quad \eta > 0, \quad \alpha + \frac{\eta}{1-\gamma}x > 0. \tag{6.3.1}$$

The corresponding Arrow-Pratt relative risk-aversion is shown as

$$\delta[U(\cdot)] = -\frac{U''(x)}{U'(x)}x = \frac{x}{\frac{\alpha}{\eta} + \frac{x}{1-\gamma}}. \tag{6.3.2}$$

The HARA class has many famous members, including the CRRA class. Obviously, when $\alpha = 0$, it degenerates to the CRRA class. Another interesting thing is that exponential utility function can be considered as a limiting-from of the HARA class given $\alpha = 1$, i.e.

$$\lim_{\gamma\to-\infty} U(x) = \lim_{\gamma\to-\infty} \frac{1}{\eta}(\frac{1}{\gamma} - 1)[1 + \frac{1}{\frac{(1-\gamma)}{\eta x}}]^{\frac{(1-\gamma)}{\eta x}\gamma\frac{\eta x}{(1-\gamma)}} = -\frac{1}{\eta}e^{-\eta x}. \tag{6.3.3}$$

Actually, Merton (1971) had ever considered the optimal investment and consumption problem with the HARA utility function on a finite horizon and derived an analytical solution when the bequest function B was zero. In that case, the investor was just maximizing his utility from intermediate consumption since the utility from terminal wealth was ignored. On the other hand, when the bequest function is also taking the form as

$$B(x) = \frac{1-\gamma}{\eta\gamma}(\alpha + \frac{\eta}{1-\gamma}x)^{\gamma}, \tag{6.3.4}$$

the PDE system (6.2.17) becomes

$$\begin{cases} V_t + rxV_x - \frac{(\mu-r)^2}{2\sigma^2}\frac{V_x^2}{V_{xx}} + \frac{(1-\gamma)\alpha}{\eta}V_x + e^{-\rho t}\frac{(1-\gamma)^2}{\eta\gamma}(e^{\rho t}V_x)^{\frac{\gamma}{\gamma-1}} = 0, \\ V(T,x) = e^{-\rho T}\frac{1-\gamma}{\eta\gamma}(\alpha + \frac{\eta}{1-\gamma}x)^{\gamma}, \quad (t,x) \in [0,T]\times[0,\infty). \end{cases} \tag{6.3.5}$$

For this case where both the utility from the intermediate consumption and the utility from the terminal wealth are taken into account, Merton only hinted, in one of the footnotes, that an analytical solution could be also obtained with bequest function being of some special forms. However, the details of how to derive such an analytical solution have not been presented in his paper. Although Merton's insight for the possibility of the existence of an analytical solution for the case where the bequest function is non-zero is greatly appreciated, how such a solution could be constructed remains a challenge. Our research starts with such an interesting question.

Question 1. Can we derive an analytical solution to the Merton problem subject to the HARA utility function with a non-zero bequest function, i.e. solving the PDE system (6.3.5) analytically?[5]

As part of the contribution to the literature, we have proved the following theorem, which also provides us an affirmative answer to Question 1.

Theorem 6.3.1. Given $\eta > 0, \gamma < 0$, the value function governed by the PDE system (6.3.5) is expressed as

$$V(t,x) = -e^{-\rho t}[a_1(t)x + b_1(t)]^\gamma, \tag{6.3.6}$$

with

$$\begin{cases} a_1(t) = e^{-A_1(T-t)\frac{1-\gamma}{\gamma}}[\frac{1}{\eta}(1-\gamma)(-\gamma)^{\frac{1}{\gamma-1}} + \frac{B_1}{A_1}(1-e^{A_1(T-t)})]^{\frac{1-\gamma}{\gamma}}, \\ b_1(t) = e^{-\int_t^T A_2(s)ds}[\alpha(\frac{\gamma-1}{\eta\gamma})^{\frac{1}{\gamma}} - \int_t^T B_2(s)e^{\int_s^T A_2(u)du}dz], \end{cases} \tag{6.3.7}$$

where

$$\begin{cases} C = \frac{(\mu-r)^2}{2\sigma^2}, \\ A_1 = \frac{\rho}{1-\gamma} - \frac{r\gamma}{1-\gamma} - \frac{\gamma C}{(1-\gamma)^2}, \\ B_1 = \frac{1-\gamma}{\eta\gamma}(-\gamma)^{\frac{\gamma}{1-\gamma}}, \\ A_2(t) = \frac{\rho}{\gamma} + \frac{C}{\gamma-1} + \frac{1}{\eta}(\frac{1-\gamma}{\gamma})^2[-a_1(t)\gamma]^{\frac{\gamma}{\gamma-1}}, \\ B_2(t) = \frac{\alpha(\gamma-1)}{\eta}a_1(t). \end{cases} \tag{6.3.8}$$

Furthermore, it is also the solution to the original HJB equation (6.2.12) when both the utility function U and the bequest function B are of the HARA form (6.3.1).

Proof. We assume that a trial solution to the PDE system (6.3.5) is of the form as (6.3.6). After substituting this trial solution into the PDE system, we have an ODE system

$$\begin{cases} \dot{a}_1 = \frac{1}{\gamma}(\rho + \frac{\gamma C}{\gamma-1} - \gamma r)a_1 + \frac{1}{\eta}(\frac{1-\gamma}{\gamma})^2(-\gamma)^{\frac{\gamma}{\gamma-1}}a_1^{\frac{2\gamma-1}{\gamma-1}}, \\ \dot{b}_1 = \frac{1}{\gamma}[\rho + \frac{\gamma C}{\gamma-1} + \frac{(1-\gamma)^2}{\eta\gamma}(-a_1\gamma)^{\frac{\gamma}{\gamma-1}}]b_1 + \frac{\alpha(\gamma-1)}{\eta}a_1, \end{cases} \tag{6.3.9}$$

with $a_1(T) = \frac{\eta}{1-\gamma}(\frac{\gamma-1}{\eta\gamma})^{\frac{1}{\gamma}}$ and $b_1(T) = \alpha(\frac{\gamma-1}{\eta\gamma})^{\frac{1}{\gamma}}$.

[5] Cox & Huang (1989) provided an answer to this question by applying the Harrison-Kreps-Pliska martingale methodology and obtained an explicit solution (fc. Example 3.4 in their paper). However, their solution is too complicated to analyze and it is also hard to extend their solution to the exponential case.

Using a nonlinear transformation $h(t) = a_1^{\frac{\gamma}{1-\gamma}}$, we can obtain

$$\dot{h} = A_1 h + B_1, \tag{6.3.10}$$

with $h(T) = \frac{1}{\eta}(\frac{\gamma-1}{\gamma})^{\frac{1}{1-\gamma}}(\frac{1}{1-\gamma})^{\frac{\gamma}{1-\gamma}}$ and

$$\begin{cases} A_1 = \frac{\rho}{1-\gamma} - \frac{r\gamma}{1-\gamma} - \frac{\gamma C}{(1-\gamma)^2}, \\ B_1 = \frac{1-\gamma}{\eta\gamma}(-\gamma)^{\frac{\gamma}{1-\gamma}}. \end{cases} \tag{6.3.11}$$

The solution to the ODE system (6.3.10) can be easily shown as

$$h(t) = e^{-A_1(T-t)}[h(T) + \frac{B_1}{A_1}(1 - e^{A_1(T-t)})]. \tag{6.3.12}$$

As a result, we obtain

$$a_1(t) = h(t)^{\frac{1-\gamma}{\gamma}} = e^{-A_1(T-t)\frac{1-\gamma}{\gamma}}[\frac{1}{\eta}(\frac{\gamma-1}{\gamma})^{\frac{1}{1-\gamma}}(\frac{1}{1-\gamma})^{\frac{\gamma}{1-\gamma}} + \frac{B_1}{A_1}(1 - e^{A_1(T-t)})]^{\frac{1-\gamma}{\gamma}}. \tag{6.3.13}$$

Since the ODE system governing $b_2(t)$ is a linear one, it is also easy to obtain the solution as

$$b_1(t) = e^{-\int_t^T A_2(s)ds}[b_1(T) - \int_t^T B_2(s)e^{\int_s^T A_2(u)du}ds], \tag{6.3.14}$$

where

$$\begin{cases} A_2(t) = \frac{\rho}{\gamma} + \frac{C}{\gamma-1} + \frac{1}{\eta}(\frac{1-\gamma}{\gamma})^2[-a_1(t)\gamma]^{\frac{\gamma}{\gamma-1}}, \\ B_2(t) = \frac{\alpha(\gamma-1)}{\eta}a_1(t). \end{cases} \tag{6.3.15}$$

Consequently, we have obtained the analytical solution (6.3.6) to the PDE system (6.3.5).

Finally, in order to ensure that the value function (6.3.6) is also the solution to the original HJB equation (6.2.12), it is necessary to verify the second-order condition (6.2.16), i.e.

determining the sign of the ϕ_{cc} and ϕ_{uu}. Given $\gamma < 0, \eta > 0$, we have

$$\begin{cases} \phi_{uu} = -\sigma^2 x^2 e^{-\rho t}(\gamma - 1)\gamma a_1^2 (a_1 x + b_1)^{\gamma - 2} < 0, \\ \phi_{cc} = -e^{-\rho t}\eta(\alpha + \frac{\eta x}{1-\gamma})^{\gamma - 2} < 0. \end{cases} \tag{6.3.16}$$

As a result, the value function (6.3.6) is indeed a solution to the HJB equation (6.2.12) when both the utility function U and the bequest function B are of the HARA form. This completes the proof. □

This theorem gives us an affirmative answer to the first question. In addition, the link between the HARA utility function and exponential utility function has already been demonstrated in Equation (6.3.3) through a limit process. Naturally, one wonders whether or not the solution of the Merton problem with exponential utility function can be obtained by taking the limit with respect to the parameter γ in the solution of the Merton problem with the HARA utility function.

Question 2. Can we solve the Merton problem with exponential utility function through taking the limit with respect to the parameter γ in the case with the HARA utility function?

After some tedious calculations, we finally also provide an affirmative answer to Question 2. Despite some non-trivial procedures are involved when we take the complicated limit, such a successful limit process would yield an analytical solution, which is summarized in the next theorem.

Theorem 6.3.2. As the parameter γ tends towards $-\infty$, the limiting-form of the solution (6.3.6) to the PDE system (6.3.5) is

$$V(t, x) = -e^{-\rho t} e^{\frac{\bar{a}_1(t)x + \bar{b}_1(t)}{\alpha}}, \tag{6.3.17}$$

with

$$\begin{cases} \bar{a}_1(t) = -\eta g(t), \\ \bar{b}_1(t) = -e^{-h_1(t)}\{\alpha[h_2(t)H(t) + \ln \eta] + \int_t^T e^{h_1(s)}[G(s) - \alpha g(s)H(s)]ds\}, \end{cases} \tag{6.3.18}$$

where

$$\begin{cases} g(t) = \frac{r}{1+(r-1)e^{-r(T-t)}}, \\ h_1(t) = \int_t^T g(s)ds, \\ h_2(t) = 1 + \int_t^T g(s)e^{h_1(s)}ds, \\ f(t) = \frac{r-\rho-C}{r}g(t)[T-t-\frac{1-e^{-r(T-t)}}{r}] - 1, \\ F(t) = (C+\rho-r)(T-t) + f(t) + \ln \eta g(t), \\ D(t) = \rho + C - 2g(t) + g(t)[\ln \eta g(t) - F(t)], \\ H(t) = \int_t^T D(s)ds, \\ G(t) = \alpha g(t)[1+F(t)]. \end{cases} \tag{6.3.19}$$

Furthermore, given $\alpha = 1$, the value function (6.3.17) is indeed the solution to the HJB equation (6.2.12) when utility function is of exponential form (6.3.3).

The proof is cumbersome and we leave the details in Appendix B.1. Although the solution obtained in Theorem 6.3.2 is of complicated structure, it is indeed analytical and in a closed form.

Of course, the completion of discovering a closed-form solution for the Merton problem with exponential utility function through taking the limit of the solution to the same problem but with the HARA utility function, which we refer to as an indirect method, has aroused our curiosity of whether or not there exists a simple and direct method to solve the Merton problem with exponential utility function on a finite horizon. Fortunately, such a simple and direct method has also been discovered and is presented in the next subsection.

6.3.2 Direct method

In the above subsection, we have demonstrated an indirect method to solve the Merton problem with exponential utility function on a finite horizon. However, the solution (6.3.17) obtained in Theorem 6.3.2 is too complex to analyze although it is indeed analytical and in a closed form. Another question comes to our mind naturally.

Question 3. Can we provide a simple and direct method to obtain a closed-form analytical solution to the Merton problem with exponential utility function, instead of adopting the

indirect method presented in the previous subsection?

As one of the main contributions of this chapter, such a simple and direct method which leads to a closed-form analytical solution and thus answer to Question 3 are presented here.

Theorem 6.3.3. The value function

$$V(t,x) = -e^{-\rho t} b_2(t) \frac{e^{a_2(t)x}}{\eta}, \tag{6.3.20}$$

where

$$\begin{cases} a_2(t) = -\eta g(t), \\ b_2(t) = \frac{1}{g(t)} e^{-\frac{C+\rho-r}{r-1+e^{r(T-t)}}[\frac{e^{r(T-t)}-1}{r}+(r-1)(T-t)]}, \end{cases} \tag{6.3.21}$$

is an exact solution of the PDE system

$$\begin{cases} V_t + rxV_x - C\frac{V_x^2}{V_{xx}} - \frac{V_x}{\eta} + \frac{\ln V_x}{\eta} V_x + \frac{\rho t}{\eta} V_x = 0, \\ V(T,x) = -e^{-\rho T}\frac{e^{-\eta x}}{\eta}, \quad (t,x) \in [0,T] \times [0,\infty). \end{cases} \tag{6.3.22}$$

Furthermore, it is also the solution to the original HJB equation (6.2.12) when both the utility function U and the bequest function B are of exponential form as (6.3.3).

Proof. Define $I(t,x) = e^{\rho t} V(t,x)$. The PDE system (6.3.22) can then be rewritten as

$$\begin{cases} I_t + rxI_x - C\frac{I_x^2}{I_{xx}} - \frac{I_x}{\eta} + \frac{\ln I_x}{\eta} I_x - \rho I = 0, \\ I(T,x) = -\frac{e^{-\eta x}}{\eta}, \quad \forall (t,x) \in [0,T) \times [0,\infty), \end{cases} \tag{6.3.23}$$

where $I_t := \frac{\partial I}{\partial t}$, $I_x := \frac{\partial I}{\partial x}$, and $I_{xx} := \frac{\partial^2 I}{\partial x^2}$.

It is assumed that an ansate to the new PDE system (6.3.23) is taking the form as

$$I(t,x) = -b_2(t)\frac{e^{a_2(t)x}}{\eta}. \tag{6.3.24}$$

By substituting this trial solution into the PDE system (6.3.23), we obtain that functions $a_2(t)$

and $b_2(t)$ satisfy the nonlinear ODE system

$$\begin{cases} \dot{a_2} = -ra_2 - \frac{a_2^2}{\eta}, \\ \dot{b_2} = (C + \frac{a_2}{\eta} + \rho)b_2 - \frac{a_2 b_2}{\eta} \ln \frac{-a_2 b_2}{\eta}, \end{cases} \tag{6.3.25}$$

with $a_2(T) = -\eta$ and $b_2(T) = 1$. By introducing a transformation $y(t) = \frac{1}{a_2(t)}$, we come to a linear ODE system

$$\dot{y} = ry + \frac{1}{\eta}, \tag{6.3.26}$$

with $y(T) = -\frac{1}{\eta}$. The solution can be easily obtained as

$$y(t) = -\frac{1 + e^{-r(T-t)}(r-1)}{\eta r}. \tag{6.3.27}$$

As a result, we obtain

$$a_2(t) = -\eta g(t), \tag{6.3.28}$$

where $g(t) = \frac{r}{1+e^{-r(T-t)}(r-1)}$.

Having obtained function $a_2(t)$, we now turn to the other nonlinear ODE system

$$\dot{b_2} = (C - g + \rho)b_2 + g b_2 \ln (g b_2), \tag{6.3.29}$$

with $b_2(T) = 1$. After a nonlinear transformation $z(t) = \ln(b_2(t))$ is introduced, the ODE system becomes

$$\dot{z} = g(t)z + A_3(t), \tag{6.3.30}$$

with $z(T) = 0$ and $A_3(t) = C + \rho - g(t) + g(t) \ln g(t)$.

By applying the method of variation of parameters, we find the solution to the ODE system

(6.3.30) as

$$\begin{aligned} z(t) &= -\frac{r}{r-1+e^{r(T-t)}}\int_t^T A_3(s)\frac{r-1+e^{r(T-s)}}{r}ds \\ &= -\ln g(t) - \frac{C+\rho-r}{r-1+e^{r(T-t)}}[\frac{e^{r(T-t)}-1}{r}+(r-1)(T-t)]. \end{aligned} \tag{6.3.31}$$

Consequently, a closed-form solution to the PDE system (6.3.22) has been finally obtained through the back substitutions of functions $a_2(t)$ and $b_2(t)$ into the expression (6.3.20).

Furthermore, to make sure that the value function (6.3.20) is also a solution to the original HJB equation (6.2.12), the second-order condition (6.2.16) needs to be checked. It is easy to verify that

$$\begin{cases} \phi_{uu} = -\sigma^2 x^2 e^{-\rho t} a_2^2 b_2 \frac{e^{-a_2 x}}{\eta} < 0, \\ \phi_{cc} = -e^{-\rho t}\eta e^{-\rho c} < 0. \end{cases} \tag{6.3.32}$$

Therefore, the value function (6.3.20) is indeed a solution to the original HJB equation (6.2.12) when both the utility function U and the bequest function B are of exponential form as (6.3.3). This completes the proof. □

It should be remarked that not only is this direct method simpler than the indirect one presented in the previous subsection, the analytical solution (6.3.20) obtained through the direct method is also of a much more elegant form than that obtained with the indirect method.

6.3.3 The equivalence theorem and verification theorem

In the previous subsections, we have provided two different analytical solutions to the original HJB equation (6.2.12) with different methods when utility function is of exponential form. One would come up with a question about the uniqueness of the solution.

Question 4. Are these two analytical solutions, which appear to be of different forms, be equivalent with each other? If so, can we provide the rigorous mathematical proof?

According to the theory of stochastic control (Yong & Zhou 1999), the HJB equation (6.2.12) admits a unique solution. As a result, these two obtained solutions, which are of vastly different

forms, need to be proven equivalent; such an equivalence is summarized in the following theorem with the rigorous and yet cumbersome mathematical proof being left in Appendix B.2.

Theorem 6.3.4. (Equivalence) Given $\alpha = 1$, the solution (6.3.17) obtained with the indirect method in Theorem 6.3.2 is equivalent to the solution (6.3.20) obtained with the direct method in Theorem 6.3.3.

According to the analytical solution (6.3.20) obtained with the direct method, the optimal consumption rate and investment proportion are obtained in the feedback form as

$$\begin{cases} u^* = -\frac{\mu - r}{x\sigma^2}\frac{V_x}{V_{xx}} = \frac{\mu - r}{x\sigma^2 \eta g(t)}, \\ c^* = (\frac{\partial U}{\partial c})^{-1}(e^{\rho t}V_x) = g(t)x - \frac{1}{\eta}\ln\left[g(t)b_2(t)\right], \end{cases} \tag{6.3.33}$$

where functions $g(t)$ and $b_2(t)$ have been defined before.

Remark 6.3.1. As Merton (1969) pointed out in his seminal paper, we need to verify the feasibility of the optimal pairs. Since the consumption rate is non-negative, i.e. $c^* \geq 0$, for any $x > 0$ and $0 \leq t \leq T$, we must have

$$\ln\left[g(t)b_2(t)\right] \leq 0. \tag{6.3.34}$$

After some tedious calculations, the above condition is equivalent to a constraint on the parameters as $C + \rho - r \geq 0$.

As shown in Equation (6.3.33), the optimal investment and consumption are obtained in terms of the value function which is governed by the HJB equation. We also need to verify such given controls are optimal for the original stochastic optimal control problem. Since the equivalence between the analytical solutions (6.3.17) and (6.3.20) has been demonstrated in Theorem 6.3.4, we only provide the verification theorem for the value function (6.3.20) obtained in Theorem 6.3.3. Before demonstrating the verification theorem, we need a lemma about the admissibility of the optimal pair.

Lemma 6.3.1. When the parameters satisfy $C + \rho - r \geq 0$, the optimal pair (6.3.33) is admissible and we also have the family

$$\{e^{-\eta X^{u^*,c^*}(\tau)}\}_{\tau \in \mathcal{K}} \tag{6.3.35}$$

is uniformly integrable, where $\mathcal{K}$ denotes the set of all stopping times.

Proof. Corresponding to the optimal pair (6.3.33), the optimal wealth $X^*(t)$ satisfies

$$dX^*(t) = [A(t)X^*(t) + B(t)]dt + D(t)dW(t), \tag{6.3.36}$$

where $A(t) = r - g(t), B(t) = \frac{(\mu-r)^2}{\sigma^2 \eta g(t)} - \frac{1}{\eta} \ln [g(t)b_2(t)]$ and $D(t) = \frac{\mu - r}{\sigma^2 \eta g(t)}$. Since $g(t)$ is uniformly bounded on $[0, T]$, it is easy to check that all functions $A(t), B(t)$ and $D(t)$ are uniformly bounded, which guarantees that the SDE (6.3.36) admits a unique strong solution

$$X^*(t) = e^{l(t)}[X(0) + \int_0^t e^{-l(s)} B(s)ds + \int_0^t e^{-l(s)} D(s)dW(s)], \tag{6.3.37}$$

where $l(t) = \int_0^t A(s)ds < 0$ and for any $p \geq 1$, we also have an estimate

$$\mathbf{E} \max_{0 \leq s \leq T} |X^*(s)|^p \leq K_T(1 + |X(0)|^p), \tag{6.3.38}$$

where K_T is a positive constant depends on T. As a result, the optimal pair (6.3.33) satisfies the integrable condition (6.2.6) and (6.2.7). To demonstrate the corresponding objective functional is finite, we have

$$\mathbf{E}e^{-\eta c^*(t)} = \mathbf{E}e^{-\eta g(t) X^{u^*,c^*}(t) + \ln g(t) b_2(t)} \leq C_1 \mathbf{E} e^{-\eta g(t) X^{u^*,c^*}(t)} \leq C_1 \{\mathbf{E} e^{-\eta X^{u^*,c^*}(t)}\}^{g(t)}, \tag{6.3.39}$$

where the second inequality holds according to Jessen inequality. Furthermore, we have

$$\begin{aligned}
&|\mathbf{E}[\int_0^T e^{-\rho s}U(c^*(s))ds + e^{-\rho T}U(X^{u^*,c^*}(T)]|,\\
&= \frac{1}{\eta}[\int_0^T e^{-\rho s}\mathbf{E}e^{-\eta c^*}ds + e^{-\rho T}\mathbf{E}e^{-\eta X^{u^*,c^*}(T)}]\\
&\leq C_2\{\int_0^T \mathbf{E}e^{-\eta c^*}ds + \mathbf{E}e^{-\eta X^{u^*,c^*}(T)}\}\\
&\leq C_3\{\int_0^T [\mathbf{E}e^{-\eta X^{u^*,c^*}(s)}]^{g(s)}ds + \mathbf{E}e^{-\eta X^{u^*,c^*}(T)}\}.
\end{aligned}$$

The last step of the proof is to demonstrate that the family $\{e^{-\eta X^{u^*,c^*}(\tau)}\}_{\tau\in\mathcal{K}}$ is uniformly integrable, i.e.

$$\mathbf{E}e^{-\eta X^{u^*,c^*}(\tau)} < \infty. \tag{6.3.40}$$

Generally, by setting $M(t) = -\eta\int_0^t e^{-l(s)}D(s)dW(s)$, we have the following conclusion that

$$Z(t) = \exp(M(t) - \frac{1}{2}\langle M\rangle_t) \tag{6.3.41}$$

is only a local martingale, where $\langle f\rangle_t$ is the quadratic variation process of process $f(t)$. Since $D(t)$ is a deterministic function and uniformly bounded, the following *Novikov* condition can be verified as follows:

$$\mathbf{E}[e^{\frac{1}{2}\langle M\rangle_t}] = \mathbf{E}[e^{\frac{1}{2}\eta^2\int_0^T e^{-2l(s)}D^2(s)ds}] < \infty. \tag{6.3.42}$$

Consequently, $Z(t)$ is indeed a martingale. Therefore, we have

$$\mathbf{E}e^{-\eta\int_0^t e^{-l(s)}D(s)dW(s)} = e^{\frac{1}{2}\eta^2\int_0^t e^{-2l(s)}D^2(s)ds}. \tag{6.3.43}$$

To prove (6.3.40), we have

$$\begin{aligned}
&\mathbf{E}e^{-\eta X^{u^*,c^*}(\tau)} \\
&\leq C_1 \mathbf{E}e^{-\eta \int_0^\tau e^{l(\tau)-l(s)} D(s)dW(s)} \\
&\leq C_1 \big[\mathbf{E}e^{-\eta \int_0^\tau e^{-l(s)} D(s)dW(s)}\big]^{l(\tau)} \quad \text{(Jessen's inequality)} \\
&\leq C_1 \big[e^{-\eta \int_0^\tau e^{-2l(s)} D^2(s)ds}\big]^{l(\tau)} \quad \text{(Equation (6.3.43))} \\
&< \infty.
\end{aligned}$$

As a result, the corresponding objective functional is finite, which demonstrates the admissibility of the optimal pair (π^*, c^*) given by (6.3.33). This completes the proof. □

After demonstrating that the optimal pair (π^*, c^*) given by (6.3.33) is admissible, we now provide our verification theorem.

Theorem 6.3.5. (Verification) Suppose that the wealth process X_t follows the dynamics (6.2.3) and the value function $V(t,x)$ is expressed in Equation (6.3.20) in Theorem 6.3.3. Then, for any admissible control pair (u,c), we have

$$V(t,x) \geq J(u,c;t,x). \tag{6.3.44}$$

Furthermore, when the parameters satisfy $C+\rho-r \geq 0$, the admissible pair (u^*,c^*) given in the feedback form (6.3.33), is optimal, i.e.

$$V(t,x) = J(u^*,c^*;t,x). \tag{6.3.45}$$

Proof. Since $V(t,x)$ satisfies the HJB equation (7.2.9), the auxiliary function $I(t,x) = e^{\rho t}V(t,x)$ is governed by

$$\begin{cases} \max\limits_{(u,c)\in\mathbf{R^2}} \{\mathcal{L}^{u,c} I(t,x) + U(c)\} = 0, \\ I(T,x) = -\frac{e^{-\eta x}}{\eta}, \quad \forall (t,x) \in [0,T) \times [0,\infty), \end{cases} \tag{6.3.46}$$

where $\mathcal{L}^{u,c} I(t,x) = \frac{\partial I}{\partial t} + \{[u(\mu-r)+r]x - c\}\frac{\partial I}{\partial x} + \frac{1}{2}x^2\sigma^2u^2\frac{\partial^2 I}{\partial x^2} - \rho I$. Define a localizing sequence

of stopping times as follows:

$$\tau^N := T \wedge \inf\{t > 0| \quad |X(t)| \geq N\}, N = 1, 2, \cdots . \tag{6.3.47}$$

After applying Ito's formula to function $e^{-\rho s} I(s, X(s))$ on $[t, \tau^N]$, we have

$$e^{-\rho\tau^N} I(\tau^N, X(\tau^N))) = e^{-\rho t} I(t, x) + \int_t^{\tau^N} e^{-\rho s} \mathcal{L}^{u,c} I(s, X(s)) ds + \int_t^{\tau^N} \{\cdots\} dW(s). \tag{6.3.48}$$

Since function $I(s, X(s))$, and all the coefficients and derivatives in $\mathcal{L}^{u,c}[I(s, X)]$ are bounded on $s \in [0, \tau^N]$ because $I(t, x) \in \mathcal{C}^{1,2}([0, T] \times \mathbf{R})$. After applying Dynkin's formula to Equation (6.3.48) and taking conditional expectation on both sides, we obtain

$$I(t, x) = \mathbf{E}[e^{-\rho(\tau^N - t)} I(\tau^N, X(\tau^N))|X(t) = x] - \mathbf{E}[\int_t^{\tau^N} e^{-\rho(s-t)} \mathcal{L}^{u,c} I(s, X(s)) ds]. \tag{6.3.49}$$

Because $I(t, x)$ solves the HJB equation (6.3.46), we have

$$\mathcal{L}^{u,c} I(t, x) + U(c) \leq 0. \tag{6.3.50}$$

Consequently, Equation (6.3.49) becomes

$$I(t, x) \geq \mathbf{E}[\int_t^{\tau^N} e^{-\rho(s-t)} U(c(s)) ds + e^{-\rho(\tau^N - t)} I(\tau^N, X(\tau^N)))|\mathcal{F}_t], \tag{6.3.51}$$

i.e

$$V(t, x) \geq \mathbf{E}[\int_t^{\tau^N} e^{-\rho s} U(c(s)) ds + e^{-\rho\tau^N} I(\tau^N, X(\tau^N)))|\mathcal{F}_t] \tag{6.3.52}$$

From Lemma 6.3.1, we have the family $\{e^{-\eta X^{u,c}(\tau)}\}_{\tau \in \mathcal{K}}$ is uniformly integrable. As a result, we have

$$\mathbf{E}[V(\tau, X(\tau))] \leq C_1 \mathbf{E}[e^{-\eta g(\tau) X(\tau)}] \leq C_1 \mathbf{E}[e^{-\eta X(\tau)}] < \infty. \tag{6.3.53}$$

By taking limit, we come to

$$\lim_{N\to\infty} \mathbf{E}[e^{-\rho\tau^N} V(\tau^N, X(\tau^N)))|\mathcal{F}_t] = \mathbf{E}[e^{-\rho T} U(X(T))|\mathcal{F}_t] \tag{6.3.54}$$

On the other hand, it is noted that

$$0 \leq \int_t^{\tau^N} e^{-\rho s} e^{-\eta c(s)} ds \uparrow \int_t^T e^{-\int_t^s \rho(u)du} e^{-\eta c(s)} ds. \tag{6.3.55}$$

According to Condition (6.2.8), we apply monotone convergence theorem to produce

$$\lim_{N\to\infty} \mathbf{E}[\int_t^{\tau^N} e^{-\int_t^s \rho(u)du} U(c(s))ds|\mathcal{F}_t] = \mathbf{E}\int_t^T [e^{-\int_t^s \rho(u)du} U(c(s))ds|\mathcal{F}_t]. \tag{6.3.56}$$

Combining (6.3.52), (6.3.54) and (6.3.56), we come to

$$V(t,x) \geq \mathbf{E}\int_t^T [e^{-\rho s} U(c(s))ds + e^{-\rho T} U(X(T))|\mathcal{F}_t] = J(u,c;t,x) \tag{6.3.57}$$

In order to prove the equality (7.3.1), we replace $(u(\cdot), c(\cdot))$ by the optimal pair (u^*, c^*) defined by Equation (6.3.33) in the above steps. Then all the inequalities become equalities, which completes the proof. □

Before leaving this section, we would re-emphasize the rareness of closed-form analytical solutions for the Merton problem. As shown in the literature review, closed-form analytical solutions are available with at least one of the following assumption: (1) the utility function belongs to the constant relative risk aversion (CRRA) class; (2) the utility function is defined over terminal wealth only and consumption is not allowed; (3) the investment horizon is infinite. This chapter presents a closed-form analytical solution, for the first time, to the Merton problem without any one of these assumptions listed above. With the new discovered analytical solution, we can explicitly explore more economic insight about the optimal policies, which have not been analyzed before. Some preliminary results of such an exploration are reported in the next section.

6.4 Discussions

In this section, some economic interpretations are presented with the help of our analytical solution (6.3.20) obtained from the direct method. First of all, we show that the optimal policies obtained via our analytical solution on a finite horizon are consistent with those obtained by Merton (1969) on an infinite horizon, with the latter being demonstrated as a special case of the former in a limit process of allowing time horizon to become infinite. Then the optimal investment proportion and optimal consumption are discussed from both economic and mathematical aspects.

6.4.1 Consistency between the finite and infinite cases

In the literature, Merton (1969) managed to produce an analytical solution when utility function is of exponential form with an assumption that investment horizon is infinite. The assumption has simplified the HJB equation to be a nonlinear ODE and made it possible to obtain an analytical solution in that case. However, this assumption is impractical in reality and only holds in a perfect mathematic model. Here, our newly derived analytical solution would now enable us to explore the economic insight associated with the Merton problem defined on a finite horizon.

According to the first-order condition (6.2.15), the optimal investment proportion and consumption rate are obtained in the feedback form as

$$\begin{cases} u^* = \frac{\mu - r}{x\sigma^2 \eta g(t;T)}, \\ c^* = g(t;T)x - \dfrac{1}{\eta}\ln\left[g(t;T)b_2(t;T)\right], \end{cases} \tag{6.4.1}$$

with the time horizon T being added as a parameter. By taking the limit of T approaching infinity, we have

$$\begin{cases} \lim\limits_{T\to\infty} g(t;T) = \lim\limits_{T\to\infty} \frac{r}{1+(r-1)e^{-r(T-t)}} = r, \\ \lim\limits_{T\to\infty} b_2(t;T) = \lim\limits_{T\to\infty} \frac{1}{g(t;T)} e^{-\frac{C+\rho-r}{r-1+e^{r(T-t)}}\left[\frac{e^{r(T-t)}-1}{r}+(r-1)(T-t)\right]} = \frac{1}{r}e^{-\frac{C+\rho-r}{r}}. \end{cases} \tag{6.4.2}$$

Accordingly, the optimal policies become

$$\begin{cases} \lim\limits_{T\to\infty} u^*(t;T) = \frac{\mu-r}{x\sigma^2\eta r}, \\ \lim\limits_{T\to\infty} c^*(t;T) = rx + \frac{C+\rho-r}{\eta r}, \end{cases} \tag{6.4.3}$$

which are in perfect agreement with the two formulae derived by Merton (1969) for the infinite-horizon case. Clearly, the infinite-horizon case is indeed a degenerated case of the finite horizon one.

However, what is really interesting would be the differences between the finite-horizon case and the Merton's infinite-horizon case. In the following, we would analyze the optimal policies in the finite-horizon case and present some interesting economic interpretations and discussions, according to the analytical solution (6.3.20).

6.4.2 Optimal investment proportion u^*

When utility function belongs to the CRRA class, Merton (1969) had come to a conclusion that the optimal investment proportion u^* is constant. Since exponential utility adopt in this chapter is not a member of the CRRA class, the corresponding optimal polices are expected to be different from those in Merton's case with the CRRA utility function.

Instead of being constant, the optimal investment proportion u^* now is inversely proportional to current wealth x as shown in Equation (6.4.1). Mathematically, it implies that the optimal investment proportion u^* would be affected by the current wealth. From economical aspects, that is because the risk aversion of the investor would also be affected by his wealth as shown in Equation (6.2.19). The more wealth the investor has, the more risk aversion he shows. As wealth becomes extremely large, the risk aversion of the investor also goes up to a very high level. Accordingly, the optimal investment proportion u^* approaches zero, which implies that all his wealth is allocated on risk-free asset to avoid the risk.

From Equation (6.4.1), it is obvious that the sign of u^* depends on the sign of risk premium $\mu - r$. Further discussions about the optimal investment proportion u^* can be divided into three cases based on the sign of the proportion u^* on the risky asset and the sign of the proportion

$1 - u^*$ on the risk-free asset.

Case 1. $u^* > 1$

In this case, the investment proportion on risk-free asset $1 - u^*$ is negative. The investor would borrow money at the interest rate r to realize a negative investment on risk-free asset and invest all his wealth, including money collected by shorting the risk-free asset, on the risky asset. This behavior appears to be weird but reasonable from both mathematical and economic aspects. According to Equation (6.4.1), there exist two situations that would make u^* be greater than 1: a large enough risk premium $\mu - r$ and a small enough current wealth x with a positive risk premium $\mu - r$. For the situation where risk premium $\mu - r$ is large enough, the investor would like to make profits by investing on the risky asset although they may take the corresponding risk. Such a high risk premium gives the investor a great motivation to overlook the risk. For the situation where risk premium $\mu - r$ is positive but not large enough, a sufficiently small wealth x can also result in u^* being greater than 1. We have to come back to the risk aversion of the investor as shown in Equation (6.2.19), from which it is observed that the risk aversion almost vanishes when x shrinks to zero. In other words, the investor does not care the risk too much when his wealth x is very small. In this situation, the investor would also like to borrow as much as possible and leave all his wealth on the risky asset to earn a small but positive risk premium. As a result, either a sufficiently large risk premium $\mu - r$ or a sufficiently small wealth x with a positive risk premium can result in u^* being greater than 1, which implies that the investor would heavily invest on the risky asset instead of investing on both the risky and risk-free assets in order to maximize his total utility.

Case 2. $0 \leq u^* \leq 1$

In this case, both the investment proportion on the risky asset u^* and on the risk-free asset $1 - u^*$ are non-negative. In other words, the investor allocates his wealth on both the risky and risk-free assets, respectively. A non-negative u^* corresponds to a non-negative risk premium $\mu - r$, i.e. the expected return on the risky asset is not less than that on the risk-free asset. Consequently, one should allocate some wealth on the risky asset in order to earn such a positive

risk premium. However, risk premium $\mu - r$ is not large enough to result in u^* being greater than 1 just as in Case 1. Therefore, there is not enough incentive to motivate the investor to allocate all his wealth on the risky asset. The investor would also like to make some investment on the risk-free asset in order to manage the risk. Contrary to the case above, in which one needs to even short the risk-free asset and invest all his wealth on the risky asset, it is optimal in this case to invest on both the risky and risk-free assets in order to trade off between returns and risk.

Case 3. $u^* < 0$

In this case, a negative u^* implies that the risk premium $\mu - r$ is also negative. In other words, the expected return through an investment on the risk-free asset is expected to exceed that through an investment on the risky asset. Furthermore, there is no risk associated with the risk-free asset and thus the investor would naturally prefer the risk-free asset to the risky one. Such a preference can be realized by shorting the risky asset and then allocating all his wealth on the risk-free asset. In this case, the investor would put all his wealth on one side again just as in Case 1. The difference is that the investor chooses the risk-free asset as the optimal asset instead of the risky asset in Case 1.

The optimal investment proportion u^* shown in Equation (6.4.1) has enabled us to not only carry out the above discussions on the economic insight of the investment proportion with different cases, but also explore the difference of the optimal investment policies between the finite-horizon and the infinite-horizon cases. Although u^* is inversely proportional to the wealth x in both the finite-horizon and the infinite-horizon cases, there still exists a significant difference: the coefficient of proportionality is constant in Merton's infinite-horizon case, while it is a time-dependent one in the finite-horizon case. Unlike the infinite-horizon case, in which the optimal amount invested in the risky asset, u^*x, maintains a constant, the optimal investment amount for the finite-horizon case now varies with time as Equation (6.4.1) can be rewritten as $u^*x = \dfrac{\mu - r}{\sigma^2 \eta g(t;T)}$, where $g(t;T)$ is a monotonically increasing function of time t.

To further demonstrate how the optimal policies change with time t in the finite-horizon case, the optimal investment proportion u^* and the optimal investment amount u^*x are plotted,

respectively, in Figure 6.1 with parameters being defined as follows:

$$\mu = 0.1, r = 0.05, \sigma = 0.5, \eta = 1, T = 1, \rho = 0.$$

Since $x = 0$ is a singular point for the optimal investment proportion u^*, we only figure out the curves on the domain $[0.02, 1]$.

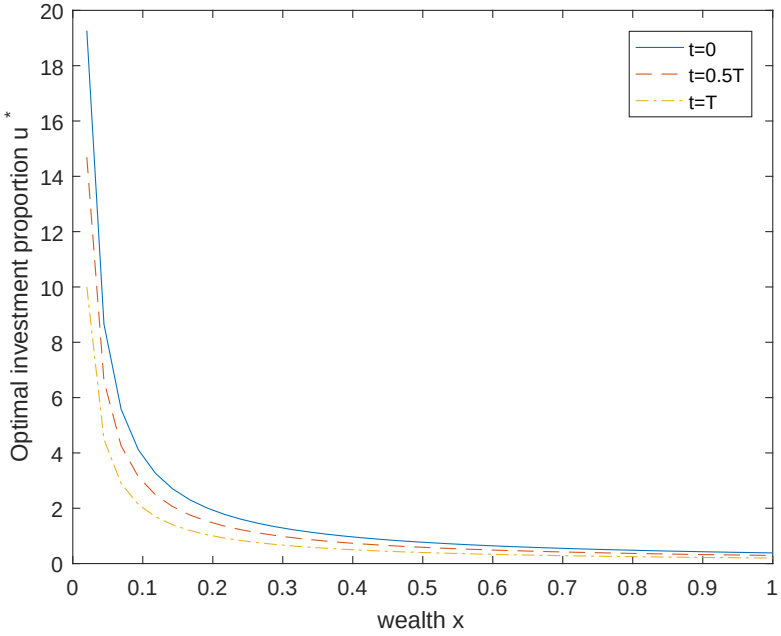
From Figure 6.1(a), although the optimal investment proportion u^* is inversely proportional to the wealth, the curvature of the curves are different at the different times because the coefficient of proportionality is a function of time t. From Figure 6.1(b), one can also observe that the optimal investment amount u^*x does not change with wealth x once time t is fixed. However, the optimal investment amount u^*x changes with time t monotonically as shown in Figure 6.1(b). In other words, the optimal amount invested in the risky asset is reduced as time goes by; while it is a constant from Equation (6.4.3) in the Merton's infinite-horizon case.

Actually some results of Figure 6.1 have been pointed out by Zhu & Ma (2018), who applied the homotopy analysis method (HAM) to the Merton problem with exponential utility function on a finite horizon and obtained a solution of an infinite series form. They also found that the optimal investment amount u^*x is a decreasing function of time t based on the numerical results from their truncated series solution. However, their series solution can not explain the monotonicity mathematically. Here our closed-form analytical solution has a great advantage over numerically-produced results. The monotonicity of the optimal investment amount u^*x can be easily stated in the following proposition.

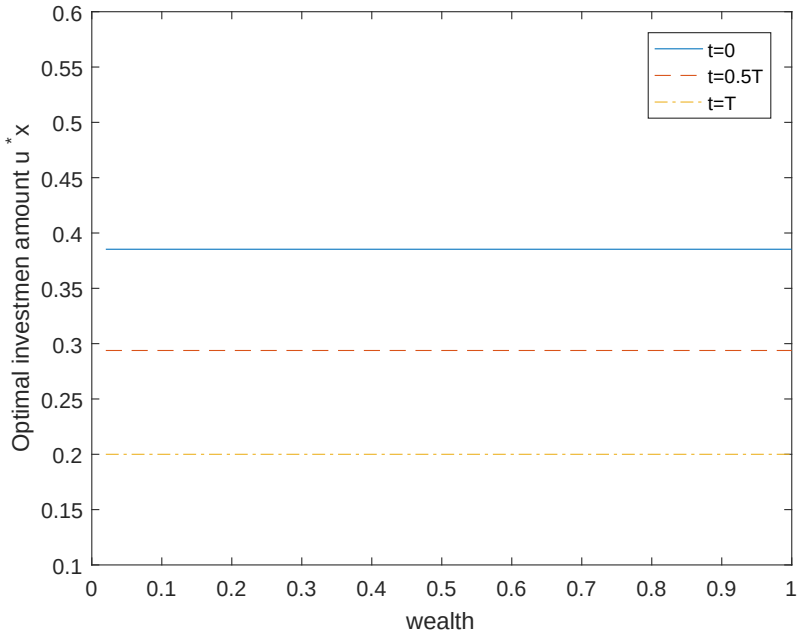
Proposition 6.4.1. For an investor with exponential utility function on a finite horizon, the optimal investment amount u^*x is a decreasing function of time t when risk premium $\mu - r$ is positive; while it becomes an increasing function of time t when risk premium is negative.

Proof. It is easy to obtain

$$\frac{\partial}{\partial t}(u^*x) = -\frac{\mu - r}{\sigma^2 \eta g^2(t;T)} \frac{\partial g(t;T)}{\partial t} = -\frac{\mu - r}{\sigma^2 \eta g^2(t;T)} \frac{r^2(1-r)e^{-r(T-t)}}{[1 + (r-1)e^{-r(T-t)}]^2}. \tag{6.4.4}$$



(a) Optimal investment proportion u^*.



(b) Optimal investment amount u^*x .

Figure 6.1: Optimal investment proportion and amount.

Obviously, when risk premium $\mu - r$ is positive, $\frac{\partial}{\partial t}(u^*x)$ is negative, which implies that the optimal investment amount is a decreasing function of time. When $\mu - r$ is negative, $\frac{\partial}{\partial t}(u^*x)$ becomes positive and the optimal investment amount now becomes an increasing function of time. □

Remark 6.4.1. According to Proposition 6.4.1, the optimal investment amount is a function of time t in the finite case; while it is constant in Merton's infinite case from Equation (6.4.3). This is a significant difference between the finite-horizon case and the infinite-horizon case with the same exponential utility function.

Since the Merton problem in this chapter is defined on a finite horizon, we are interested in the question how the time horizon T affects the optimal investment proportion u^*, i.e., the *horizon effect* of the optimal investment proportion u^*. Such an examination of the horizon effect is motivated by the classic work of Samuelson (1969) and Merton (1969). They concluded that an investor with the CRRA utility should choose a constant investment proportion, regardless of investment horizon. However, more and more later empirical evidences suggest that the horizon does paly an important role in practice. For example, Brennan et al. (1997) compared the optimal portfolio proportions of an investor with a long horizon and those of an investor with a short horizon and found that they are significantly different. Barberis (2000) found that there are enough empirical evidences to support that an investor with a longer horizon would allocate more to stocks. Kim & Omberg (1996) and Wachter (2002) also theoretically reached some similar conclusions from their models, respectively. In the following proposition, we demonstrate the horizon effect of optimal investment proportion in our model.

Proposition 6.4.2. For an investor with exponential utility function on a finite horizon, the optimal investment proportion u^* is an increasing function of horizon T if risk premium $\mu - r$ is positive; while it becomes a decreasing function of horizon T if risk premium $\mu - r$ is negative.

Proof. To analyze the horizon effect of the optimal investment proportion, we only need to

explore the derivative of u^* with respect to horizon T. Given $\mu - r > 0$, we have

$$\frac{\partial u^*}{\partial T} = -\frac{\mu - r}{x\sigma^2\eta g(t;T)^2}\frac{\partial g(t;T)}{\partial T} = \frac{(\mu - r)(1-r)e^{-r(T-t)}}{x\sigma^2\eta} > 0. \tag{6.4.5}$$

As a result, the optimal investment proportion u^* is an increasing function of horizon T. Similarly, when risk premium $\mu - r$ is negative, we have $\dfrac{\partial u^*}{\partial T} < 0$, which indicates that the optimal investment proportion u^* is a decreasing function of horizon T. This completes the proof. □

Remark 6.4.2. A remarkably significant difference between our model and Merton (1969) lies in the utility function. In Merton's finite-horizon case where utility function is a member of the CRRA class, he concluded that the optimal investment proportion is independent of horizon time T, i.e. the horizon effect does not appear; while the time horizon effect does really exist in our model where utility function is of exponential form.

6.4.3 Optimal consumption rate c^*

After discussing the optimal investment proportion u^* above, we now come to the other key component of the Merton problem, the optimal consumption rate c^*, which is in the feedback form as shown in Equation (6.4.1). Obviously, it is an affine function of wealth x and is consist of two components. The first part corresponds to the investor's wealth; while the second part is a constant, which represents the basic consumption.

First of all, at any given time t, we consider the sensitivity of the optimal consumption rate c^* with respect to wealth x, which is also called *marginal propensity to consume* (MP) in economics. From Equation (6.4.1), the MP is defined as

$$\frac{\partial c^*}{\partial x} = g(t) = \frac{r}{1 + (r-1)e^{-r(T-t)}} > 0. \tag{6.4.6}$$

In other words, an increase in wealth x allows the investor to afford more consumption, which is known as *positive wealth effect* economically. In addition, $g(t)$ is also an increasing function of time, which indicates that the MP increases as time t approaches horizon T.

In addition to the MP, there is also another important economic concept, *average propensity to consume* (ACP), which is measured by the consumption-wealth ratio $\frac{c^*}{x}$. Merton (1969) observed that such a ratio is constant when utility function belongs to the CRRA class, which means that the investor would consume a constant proportion of his wealth at any time t. When utility function is of exponential form, the optimal consumption-wealth ratio in the finite-horizon case becomes

$$\frac{c^*}{x} = \frac{r}{1+(r-1)e^{-r(T-t)}} + \frac{1}{x\eta}\frac{C+\rho-r}{e^{r(T-t)}+r-1}[\frac{e^{r(T-t)}-1}{r} + (r-1)(T-t)]; \tag{6.4.7}$$

while its counterpart in the Merton's infinite-horizon case is

$$\frac{c^*}{x} = r + \frac{C+\rho-r}{\eta r}\frac{1}{x}. \tag{6.4.8}$$

Proposition 6.4.3. For the Merton problem defined on a finite horizon with exponential utility function, at any given time t prior to T,

1. when $C+\rho-r > 0$, the optimal consumption-wealth ratio is a decreasing function of wealth x;
2. when $C+\rho-r = 0$, the optimal consumption-wealth ratio is independent of wealth x;
3. when $C+\rho-r < 0$, the optimal consumption-wealth ratio is an increasing function of wealth x.

Furthermore, as wealth x tends toward infinity, we have

$$\lim_{x\to\infty}\frac{c^*}{x} = g(t). \tag{6.4.9}$$

Proof. According to the Taylor's Theorem, we have

$$e^{r(T-t)} = 1 + r(T-t) + \frac{r^2(T-t)^2}{2}e^{\xi}, \xi \in [0, r(T-t)]. \tag{6.4.10}$$

Consequently, we obtain

$$\frac{e^{r(T-t)}-1}{r}+(r-1)(T-t)=r(T-t)+\frac{r(T-t)^2}{2}e^{\xi}>0, \tag{6.4.11}$$

and

$$e^{r(T-t)}+r-1=r(T-t+1)+\frac{r^2(T-t)^2}{2}e^{\xi}>0. \tag{6.4.12}$$

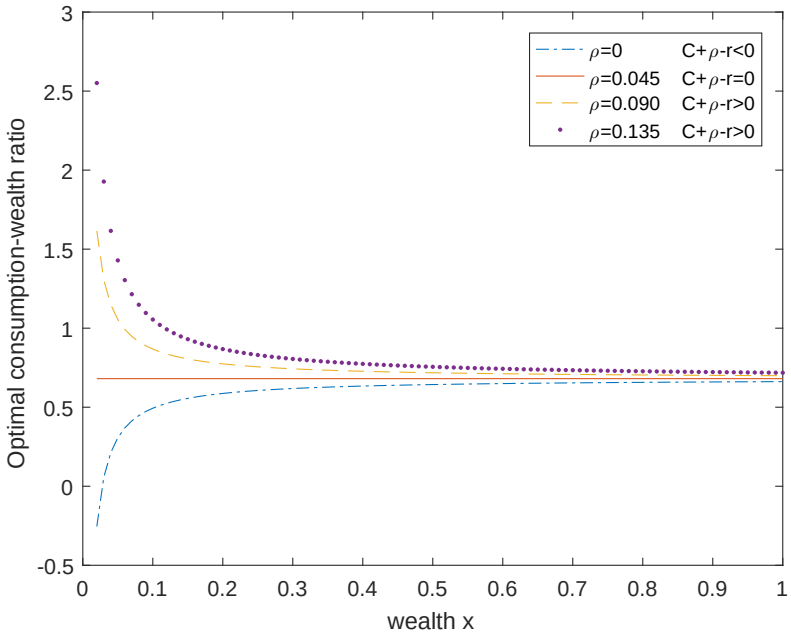
According to Equations (6.4.11), (6.4.12) and (6.4.7), the monotonicity of consumption-wealth ratio with respect to wealth x depends on the sign of $C+\rho-r$. Such a ratio is a decreasing function of wealth x if $C+\rho-r>0$; while it becomes an increasing function of wealth if $C+\rho-r<0$. Furthermore, it is easy to take the limit with respect to x to obtain Equation (6.4.9). This completes the proof. □

Remark 6.4.3. When utility function belongs to the CRRA class, the optimal consumption-wealth ratio is constant, which implies that the MP is equal to the ACP (Merton,1969). From Equation (6.4.7), the optimal consumption-wealth ratio for exponential utility function is no longer constant. For any fixed time t, the monotonicity of the ratio with respect to wealth depends on the sign of $C+\rho-r$ mathematically. It is also interesting to note that the ACP converges to the MP as wealth x approaches infinity from Equation (6.4.9) given fixed other parameters.

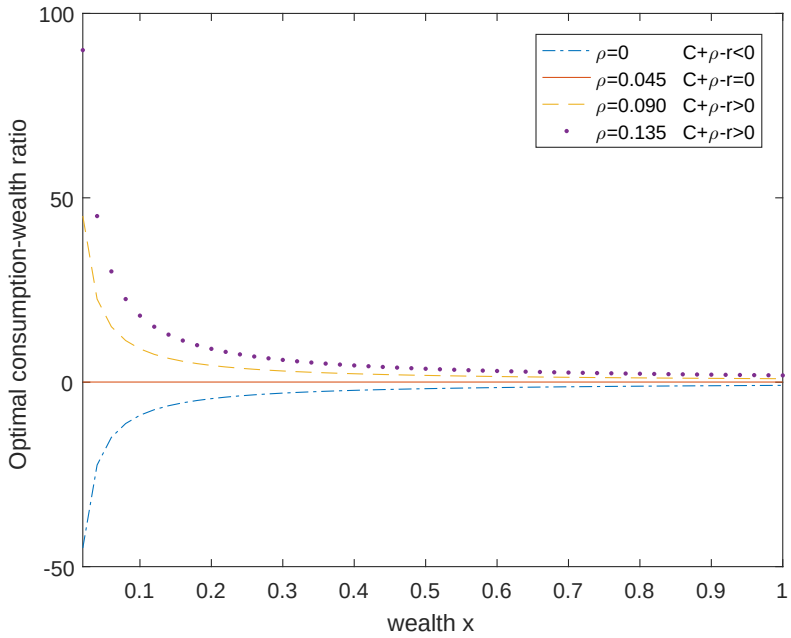
Although there are three scenarios to discuss mathematically as shown in Proposition 6.4.3, we only need to explain the first two scenarios, because the last one has been discarded by Remark 6.3.1 for its corresponding consumption rate may be negative. Some figures are plotted to show how the optimal consumption-wealth ratio changes with wealth x at the fixed time $t=0.5$, which demonstrates the results in Proposition 6.4.3 more clearly. In Figure 6.2, the parameters are set as follows:

$$\mu=0.1, r=0.05, \sigma=0.5, \eta=1, T=1.$$

Again, we just show the behavior over $[0.02, 1]$ for $x=0$ is a singular point. In order to make



(a) Finite horizon case.



(b) Infinite horizon case.

Figure 6.2: Optimal consumption-wealth ratio $\frac{c^*}{x}$ varies with wealth x with $t = 0.5$.

comparisons, we also plot the optimal consumption-wealth ratio in the Merton's infinite case. From Figure 6.2, it is noted that the optimal consumption-wealth ratio may be negative when $C+\rho-r<0$ in both finite and infinite horizon cases, which implies that discarding the case where $C+\rho-r<0$ is reasonable. When $C+\rho-r$ is non-negative, the optimal consumption-wealth ratio is a non-increasing function of wealth in both finite-horizon and infinite-horizon cases.

As shown in Equation (6.4.6), the MP is positive and the investor would consume more as his wealth increases. In other words, there will be a positive ΔC for an increased wealth ΔX. However, the increased wealth ΔX may not be consumed totally. He would allocate part of the increased wealth on investment in order to accumulate more wealth so that he can consume in the future. Economically, consumption and investment are actually considered as two substitute goods for an investor. When the increased amount in wealth ΔX is fixed, there would be a *substitution effect* on how to allocate this increased amount: the more is allocated for consumption, the less would be left for investment. Actually, the consumption-wealth ratio provides a mechanism to examine how an investor trades off between consumption and investment according to his own preference (Wachter 2002). It has been clearly manifested in Proposition 6.4.3 how to trade off between consumption and investment along the wealth direction in our model. For the first two scenarios where $C+\rho-r$ is non-negative, mathematically, we have

$$\frac{c^*}{x}=g(t)+\frac{1}{x\eta}\frac{C+\rho-r}{e^{r(T-t)}+r-1}[\frac{e^{r(T-t)}-1}{r}+(r-1)(T-t)]\geq g(t)=\frac{\partial c^*}{\partial x}, \tag{6.4.13}$$

according to Equations (6.4.11) and (6.4.12). Such an inequality indicates that the MP is less than the ACP, which implies that the former would drive the latter to move down. That is the reason why the optimal consumption-wealth ratio, or the ACP, is a non-increasing function of wealth x when $C+\rho-r\geq 0$.

Clearly, for a fixed interest rate r and a fixed Sharped ratio $\frac{\mu-r}{\sigma}$, there is a critical subjective rate of time preference, $\rho_c=r-C$, beyond which the monotonicity of the optimal consumption-wealth ratio has changed. There appears to be an interesting paradox in the first

scenario of Proposition 6.4.3, where a sufficiently large subjective rate of time preference implies that the more wealth an investor has, the "proportionally" less consumption he is willing to commit[6]. This would appear to be at odds with the definition of the subjective rate of time preference, as a larger subjective rate of time preference implies that one should consume more now than in the future. After a short period of confusion, a perfect explanation to this appear-to-be conflicting observation rests with the optimal process displayed in Equation (6.2.5). For a large subjective rate of time preference ρ, one may indeed "proportionally" consume less during the entire consumption process, leaving enough money to invest so that there would be a sufficiently large terminal wealth to be added to the entire optimization process. Of course, the balance between consumption and investment is nonlinearly governed by Equation (6.2.5).

Finally, let us now examine the behavior of the optimal consumption rate along the time direction, which depends on the region of the parameter space that the given parameters lie in. This is summarized in the following proposition.

Proposition 6.4.4. Given a set of parameters C, r, x, η, ρ, T and $C + \rho - r \geq 0$, let $p(t)$ be a function defined as

$$p(t) = -e^{r(t-T)} + \frac{r}{1-r}(t-T) + \frac{r^2 x\eta}{(1-r)(C+\rho-r)} + \frac{1-2r}{(1-r)^2}. \tag{6.4.14}$$

1. If the "discriminant " function $p(t)$ satisfies that $p(0) \geq 0$, the optimal consumption rate is an increasing function of time for $0 \leq t \leq T$.

2. If $p(0) < 0$ and $p(T) > 0$, let $\bar{t}$ be the root of the equation $p(t) = 0$. Then the optimal consumption rate is a decreasing function of time for $0 < t < \bar{t}$; while it is an increasing function of time for $\bar{t} < t < T$.

3. If $p(T) \leq 0$, the optimal consumption rate is a decreasing function of time for $0 \leq t \leq T$.

Proof. To explore how the optimal consumption rate changes over time t, we take the first-order

[6]The reason we say "proportionally less" is because the absolute total consumption still increases as x becomes larger. But, it is the ratio of $\frac{c^*}{x}$ that is decreasing as x increases, or the amount spent on consumption is proportionally less.

derivative with respect to time t

$$\frac{\partial c^*}{\partial t} = \frac{(C+\rho-r)(1-r)^2 e^{r(T-t)}}{\eta[r-1+e^{r(T-t)}]^2} p(t), \tag{6.4.15}$$

where $p(t)$ is defined in Equation (6.4.14). Furthermore, we can easily show that

$$p'(t) = -re^{r(t-T)} + \frac{r}{1-r} \geq p'(T) = \frac{r^2}{1-r} > 0.$$

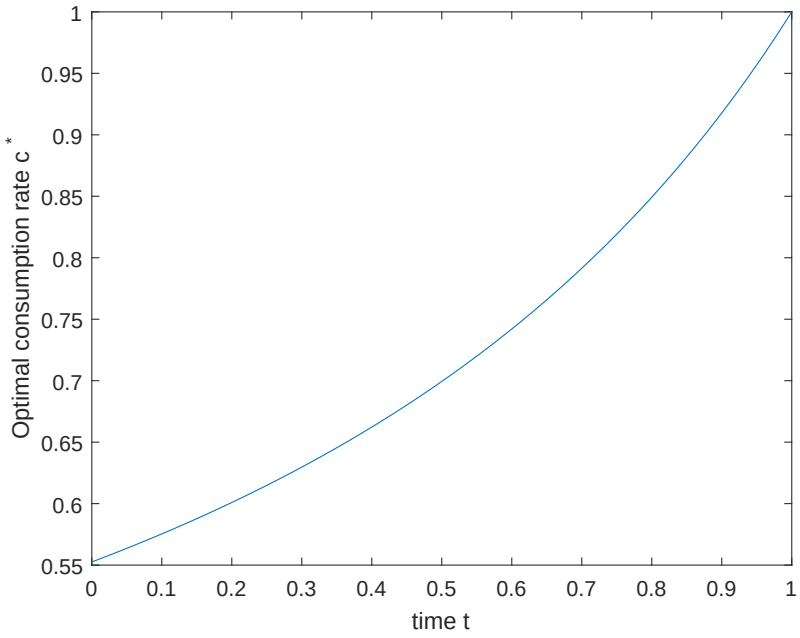
Thus, $p(t)$ is an increasing function of time t.

As a result of the monotonicity of function $p(t)$, the parameter space is divided into three regions according to the sign of $p(0)$ and $p(T)$: 1) if $p(0) \leq 0$, $p(t)$ is non-negative for $0 \leq t \leq T$; 2) if $p(0) < 0, p(T) > 0$, let $\bar{t}$ be the root of the equation $p(t) = 0$. Then $p(t)$ is negative for $0 < t < \bar{t}$ and positive for $\bar{t} < t < T$; 3) if $p(T) \geq 0$, $p(t)$ is non-positive for $0 \leq t \leq T$. From Equation (6.4.15), the sign of $\frac{\partial c^*}{\partial t}$ is determined by the sign of function $p(t)$, which completes the proof. □

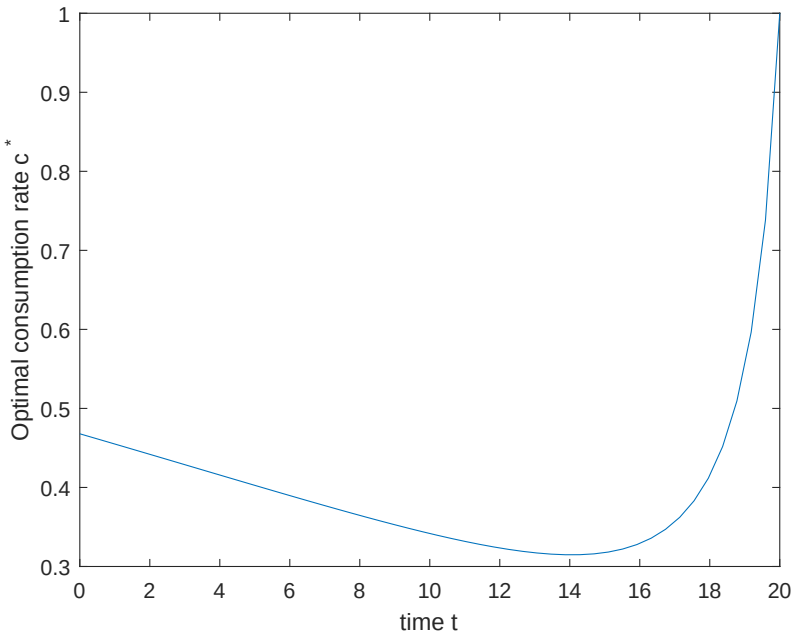
To demonstrate the influence of parameters on the monotonicity of the optimal consumption rate as summarized in Proposition 6.4.4, let's take the horizon effect of the optimal consumption rate as an example. That is, we explore how the investment horizon T affects the monotonicity of the optimal consumption rate along the time direction when other parameters are fixed. In Figure 6.3, we plot out how the optimal consumption rate changes with different values of T where the other parameters are set as follows

$$\mu = 0.1, r = 0.05, \sigma = 0.5, \eta = 1, \rho = 0.09, x = 1.$$

From Figure 6.3, it is obvious that the allocation of the optimal consumption rate along the time direction is significantly different when the time horizon T takes different values, which clearly demonstrates the horizon effect. In Figure 6.3(a) where $T = 1$, the optimal consumption rate is an increasing function of time t, which implies that these parameters lie in the first region stated in Proposition 6.4.4. On the other hand, in Figure 6.3(b) where $T = 20$,



(a) T=1.



(b) T=20.

Figure 6.3: Optimal consumption rate c^* varies with time t.

the optimal consumption rate is a decreasing function for $0 < t < \bar{t}$ and an increasing function for $\bar{t} < t < T$, which is obviously the second case stated in Proposition 6.4.4. The reason for these different behaviors with different time horizon T is that for a long horizon T, the investor would allocate more money on investment to accumulate wealth in his early life and thus the optimal consumption rate would decrease in his early life. When he enters his late life or approaches retirement, consumption would become more important than investment and the optimal consumption rate starts increasing. For a short horizon T, the optimal consumption rate would always increase with time because the horizon is so short that the investor does not have enough time for the accumulation of his wealth; it is therefore more reasonable for him to just keep increasing the optimal consumption rate from the very beginning.

While the first two cases stated in Proposition 6.4.4 have been indeed numerically realized, it is interesting for us to notice that the third case is never observed with the given parameters above, no matter how many different values of T we have tried on. Later on, we realized that there is actually a nice mathematical explanation to this. According to Proposition 6.4.4, the third case appears only when

$$p(T) = -1 + \frac{r^2 x\eta}{(1-r)(C+\rho-r)} + \frac{1-2r}{(1-r)^2} = \frac{r^2 x\eta}{(1-r)(C+\rho-r)} - \frac{r^2}{(1-r)^2} \tag{6.4.16}$$

is non-positive. Obviously, the sign of Equation (6.4.16) is already independent of T. As shown in Equation (6.4.16), when the value x is sufficiently small, $p(T)$ could be negative. To numerically display the third case stated in Proposition 6.4.4, we chose a small value of wealth $x = 0.02$ and took the remaining parameters as:

$$\mu = 0.1, r = 0.05, \sigma = 0.475, \eta = 1, \rho = 0.09, T = 1.$$

The results are displayed in Figure 6.4, which clearly shows that the optimal consumption rate is decreasing with time. Now it is even more interesting to note that the behaviors displayed in Figure 6.4 and Figure 6.3(a) are totally different when only the value of wealth x is changed

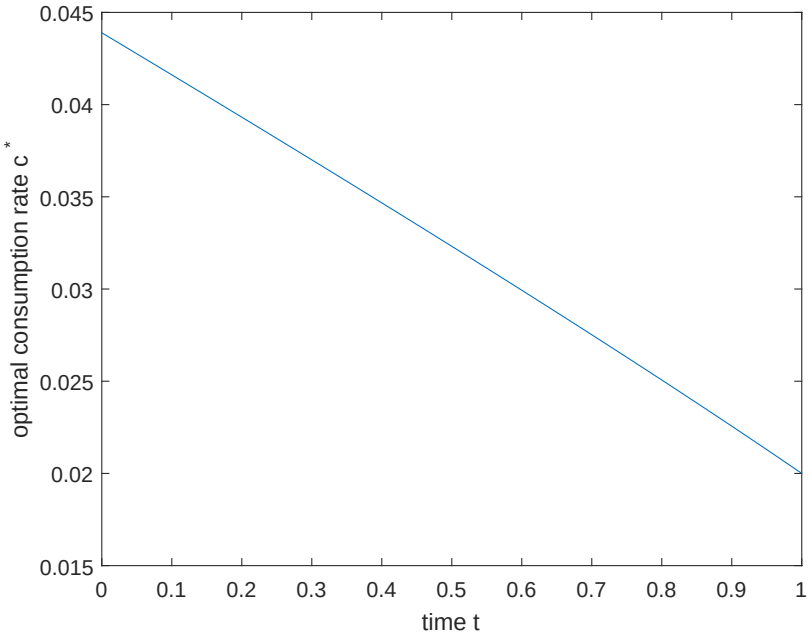


Figure 6.4: Optimal consumption-wealth ratio with $x = 0.02$.

from $x = 1$ to $x = 0.02$; while the other parameters, including time horizon, are the same. An explanation for the behavior of the consumption rate in Figure 6.4 is that the investor has so little wealth that there is not enough motivation to invest. As a result, he would prefer consumption to investment. The choice of more consumption and less investment would lead to a further decrease of the total wealth and the less available wealth would lead to a further reduction of consumption. Therefore, such a spiral-down of consumption is manifested in Figure 6.4 in the form of a mathematically decreasing function.

6.5 Conclusions

In this chapter, a closed-form analytical solution to the Merton problem with exponential utility function being defined over both intermediate consumption and terminal wealth on a finite horizon is added to the literature for the first time. This is achieved through two different approaches with one being an indirect method and the other one being a simple and direct method. The solutions obtained with different methods appear to be of different forms but finally have been demonstrated to be actually equivalent.

Utilizing the newly obtained solution, we are able to provide some interesting economic interpretations. On one hand, the obtained result for the finite-horizon case is consistent with that of the infinite-horizon case, with the latter being demonstrated as a special case of the former. On the other hand, there are some significant differences between the finite-horizon case and the infinite-horizon case, which can be summarized as

- while the optimal investment amount u^*x in Merton's infinite-horizon case is constant, it varies monotonically with time t in the finite-horizon case and the monotonicity depends on the sign of risk premium;
- while there is no *horizon effect* in Merton's infinite-horizon case, the *horizon effect* of optimal policies is clearly in presence in the finite-horizon case, which has been explained both from mathematical and economic viewpoints.

Finally, comparing with the other closed-form solutions available in the literature for the finite-horizon case, the optimal policies can also be quite different if utility functions are of different forms. When utility function belongs to the CRRA class, the optimal investment proportion u^* and the optimal consumption-wealth ratio $\frac{c^*}{x}$ are both constant, whereas they are functions of wealth x and time t when utility function is of exponential form, as newly presented in this chapter.

Chapter 7

A monotone numerical scheme for the HJB equation arising from the Merton problem

7.1 Introduction

One of the well-known topics in mathematical finance is the continuous-time portfolio optimization problem (also referred to as the *Merton problem*). In a financial market consisting of a bond that grows at a risk-free interest rate and a risky asset that is modelled as a geometric Brownian motion with constant drift rate and volatility, an investor needs to make a decision on how to allocate his wealth in order to maximize his utility from terminal wealth. Merton (1969) formulated it as a stochastic optimal control problem, which could be solved by the dynamic programming method with a fully nonlinear Hamilton-Jacobi-Bellman (HJB) equation.

In the literature, only when utility function is of some special forms, can an analytical solution to the HJB equation be obtained. With utility function being of power or logarithm form, Merton (1969) managed to produce an analytical solution for both of them belonging to the constant relative risk aversion (CRRA) class. Merton (1971) also obtained an analytical solution when utility function was a member of the hyperbolic absolute risk aversion (HARA)

class. However, in addition to these special classes, varieties of utility functions have been presented to characterize diversified investors in the expected utility theory. It is necessary to explore how to solve the HJB equation with general utility functions which include but are not limited to these special forms.

Economically, the investor characterized by a CRRA utility function would not change his attitude to the risk because his relative risk aversion is constant no matter how much money he has. However, more and more empirical evidences show that the investor's attitude to the risk may change with his wealth as Brunnermeier & Nagel (2008) and Liu et al. (2014) suggest. A mixed utility function presented by Fouque et al. (2015) is obviously not a member of the CRRA class for his relative risk aversion being not constant. Due to its complicated form, there is still no analytical solution to the HJB equation with such a mixed utility function and the optimal strategy is not obtained, either. It is natural to ask how to solve the HJB equation with such a mixed power utility function and how to interpret its corresponding optimal strategy.

Of course, analytical solution is preferred when we try to solve the HJB equation. In the literature, Merton (1969, 1971) successfully produced analytical solutions with utility function being of a member of the CRRA and HARA class. Karatzas et al. (1987), Karatzas & Shreve (1998) tried to solve the Merton problem with very general conditions on the utility function. They successfully applied the Harrison-Kreps-Pliska martingale method to solve the Merton problem with the HARA utility function. Unfortunately, when utility function is of mixed power form presented by Fouque et al. (2015) , analytical solution is still unavailable so far. Recently, Zhu & Ma (2018) adopted the homotopy analysis method (HAM) to produce an explicit series solution to the Merton problem with the mixed power utility function. However, their explicit series solution has to be truncated when numerical results are computed for it is of series form.

As an alternative, an efficient, convergent and stable numerical scheme, which can deal with the HJB equation with general utility functions, is presented in this chapter for these cases where analytical solution is unavailable. Unlike numerical schemes designed for linear PDE where the convergence and stability are easily analyzed, it is difficult to guarantee the

convergence of a numerical scheme for the fully nonlinear PDE. In many cases, classical solution to the fully nonlinear HJB equation does not exist and the solution that we are looking for is a viscosity one (Crandall et al. 1992). In order to guarantee convergence to the viscosity solution, the numerical scheme must be consistent, stable, and monotone (Barles & Souganidis 1991). A positive coefficient method is usually adopted to construct a monotone scheme (Kushner 1990, Forsyth & Labahn 2007, Barles & Jakobsen 2007). Typically, forward or backward difference for the first-order derivative term is used to make sure that the positive coefficient condition is satisfied. One disadvantage of such a truncation is that the truncation error is only first order, which has been pointed out by Wang & Forsyth (2008). In the same paper, they also presented a numerical scheme using central difference as much as possible to overcome such a disadvantage.

The main contribution of this chapter is that a monotone numerical scheme, which is able to solve the Merton problem with general utility functions, is presented with proper boundary conditions. In our numerical scheme, the positive coefficient condition always holds, which ensures that our scheme is monotone. In order to obtain a high convergence order, we follow Wang & Forsyth (2008) and adopt central difference as much as possible when the positive coefficient condition is satisfied. In the time direction, fully implicit method is applied, which results in a series of nonlinear algebraic equations that are solved by an iterative method. Based on the converged numerical results, more discussions about the optimal strategy can be provided. Another contribution of this chapter is that our numerical results can be used to verify the convergence of the explicit series solution which is obtained by Zhu & Ma (2018) based on the homotopy analysis method (HAM).

The rest of this chapter is organized as follows. In Section 7.2, the Merton problem is reviewed to give a complete background reference to the financial context of the HJB equation. In Section 7.3, a monotone numerical scheme is presented. In Section 7.4, the convergence of our numerical scheme is proved by demonstrating the stability, consistency, and monotonicity. In Section 7.5, the convergence of our iteration method for the nonlinear algebraic equation is demonstrated. In Section 7.6, three numerical examples are presented. In the first two

examples, analytical solutions have been obtained in the literature, which can be used as a benchmark to verify our numerical results. In the last example, a mixed power utility function is taken into account. Our numerical results in the last example also verify an explicit series solution obtained with the homotopy analysis method. Some conclusions are presented in the last section.

7.2 Merton problem and the HJB equation

7.2.1 The Merton problem

Consider a financial market where two assets are traded continuously on a finite horizon $[0, T]$. One asset is a risk-free bond, whose price $\{P(t), t \geq 0\}$ evolves according to the ordinary differential equation (ODE)

$$dP(t) = rP(t)dt, \qquad t \in [0, T], \tag{7.2.1}$$

with r being the risk-free interest rate. The other one is a risky asset with its price being modeled as a geometric Brownian motion

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \qquad t \in [0, T], \tag{7.2.2}$$

where μ is the drift rate, σ is the volatility, and $W(t)$ is a standard Brownian motion.

An investor starts with a known initial wealth x_0 and his total wealth at time t is denoted as $X(t)$. At any time t, prior to T, the investor needs to make a decision on how to allocate his wealth in order to maximize his expected utility from the terminal wealth $X(T)$. The proportion of total wealth invested in the risky asset at time t is denoted as $u(t)$ and the remaining fraction $1 - u(t)$ is thus left in form of the risk-free bond. The investment proportion on the risky asset $u(t)$ may be negative, which corresponds to short selling. The remaining proportion $1 - u(t)$ may also become negative, which can be realized by borrowing at the interest rate r. As a

result, the total wealth $X(t)$ is governed by the following SDE:

$$dX(t) = [r + u(t)(\mu - r)]X(t)dt + X(t)u(t)\sigma dW(t). \tag{7.2.3}$$

The objective of the Merton problem is to find the optimal investment strategy, $u^*(t)$ such that the expected utility of the terminal wealth, $X(T)$, is maximized. Mathematically, such an objective functional is stated as

$$\max_{u(\cdot)} \quad \mathbf{E}U(X(T; u(\cdot))), \tag{7.2.4}$$

where $\mathbf{E}$ is the expectation operator; $U(x)$ is the utility function defined on the terminal wealth. In addition, the fact that the wealth process can not be negative in practice leads to a constraint being imposed on the optimization

$$X(t) \geq 0, \quad t \in [0, T]. \tag{7.2.5}$$

In a brief summary, the Merton problem has been reformulated as a stochastic optimal control problem with an objective functional (7.2.4), driven by the dynamics (7.2.3), and the constraint (7.2.5).

7.2.2 The HJB equation

It is noted that the total wealth $X(t)$ is obviously a Markovian process, we can apply the dynamic programming method to solve such a stochastic optimal control problem presented in the previous section. The basic idea of the dynamic programming is to consider a family of stochastic optimal control problems with different initial times and states, to establish relationships among these problems and finally to solve all of them.

Let $(t,x) \in [0,T) \times \mathbf{R}^+$ and consider the following control system over $[t,T]$

$$\begin{cases} dX(s) = [r + u(s)(\mu - r)]X(s)ds + X(s)u(s)\sigma dW(s), \\ X(t) = x, \end{cases} \tag{7.2.6}$$

with the same constraint (7.2.5). The cost functional is

$$\mathbf{J}(u(\cdot); t, x) = \mathbf{E}[U(X^{t,x;u(\cdot)}(T))]. \tag{7.2.7}$$

where $X^{t,x;u(\cdot)}(t)$ denotes the solution to the SDE system (7.2.6) with a given strategy $u(\cdot)$. The value function is then defined as

$$V(t,x) = \max_{u(\cdot)} \mathbf{J}(u(\cdot); t, x). \tag{7.2.8}$$

According to the dynamic programming method (Yong & Zhou 1999), the value function $V(t,x)$ satisfies the HJB equation:

$$\begin{cases} \max\limits_{u \in \mathbf{R}} \phi(u; t, x) = 0, \\ V(T,x) = U(x), \quad \forall (t,x) \in [0,T] \times [0,\infty), \end{cases} \tag{7.2.9}$$

where

$$\phi(u; t, x) = \frac{\partial V}{\partial t} + [u(\mu - r) + r]x\frac{\partial V}{\partial x} + \frac{1}{2}x^2\sigma^2u^2\frac{\partial^2 V}{\partial x^2}. \tag{7.2.10}$$

Obviously, the PDE system (7.2.9) is highly nonlinear and it is difficult to obtain an analytical solution for a general utility function $U(x)$ unless it belongs to some special families, such as the CRRA or HARA class. In the following, we would present a monotone numerical scheme for the solution of such a nonlinear PDE system with $U(x)$ being general utility functions.

7.3 Numerical scheme for the HJB equation

In this section, a monotone numerical scheme for the nonlinear PDE system (7.2.9) with general utility functions $U(x)$ is presented. For convenience purposes, we introduce time reversal $\tau = T - t$ and rearrange the PDE system (7.2.9) as

$$\begin{cases} \dfrac{\partial V}{\partial \tau} = \max\limits_{u \in \mathbf{R}} \{\mathcal{L}^u V\}, \\ V(0, x) = U(x), \end{cases} \tag{7.3.1}$$

where

$$\begin{cases} \mathcal{L}^u V = a(x, u)\dfrac{\partial^2 V}{\partial x^2} + b(x, u)\dfrac{\partial V}{\partial x}, \\ a(x, u) = \dfrac{1}{2}x^2\sigma^2 u^2, \\ b(x, u) = [r + u(\mu - r)]x. \end{cases} \tag{7.3.2}$$

The domain of the PDE system (7.3.1) is $\Omega = \{(\tau, x) \in [0, T] \times [0, \infty)\}$ that is obviously unbounded.

7.3.1 Boundary conditions

In fact, there is only a terminal condition for the original HJB equation (7.2.9). In order to establish a properly-closed PDE system, boundary conditions are necessary. Here we provide some proper boundary conditions for the HJB equation arising from the Merton problem for the first time.

At the boundary $x = 0$, the PDE (7.3.1) degenerates to

$$\frac{\partial V}{\partial \tau} = 0. \tag{7.3.3}$$

The appropriate boundary condition can be obtained with the degenerate equation above.

At $x \to \infty$, we normally use financial reasoning to impose an appropriate boundary condition. Since utility function is an increasing function of x, there are only two kinds of behaviors as x tends toward infinity: converging to a constant or blowing up. For example, exponential

utility function $U(x) = -\frac{1}{\eta}e^{-\eta x}$ with $\eta > 0$ and power utility function $U(x) = \frac{x^\gamma}{\gamma}$ with $\gamma < 0$ would tend towards zero as x approaches infinity. In this case, we assume the value function has similar asymptotic behavior to the utility function and impose the Dirichlet boundary condition at $x \to \infty$ as

$$\lim_{x\to\infty} V(\tau, x) = \lim_{x\to\infty} U(x) = 0. \tag{7.3.4}$$

There are also some utility functions that blow up when x tends to infinity. To better articulate the establishment of the appropriate boundary condition for these utility functions, we need more precise descriptions about the behavior of utility function at $x \to \infty$. As a result, we form and prove the following proposition.

Proposition 7.3.1. Assume that a smooth utility $U(x)$ has a polynomial growth rate and its growth order is α. In other words,

$$\lim_{x\to\infty} \frac{U(x)}{x^\alpha} = C, \tag{7.3.5}$$

where C is a positive constant. Then α must lie in $[0, 1]$.

Proof. Since utility function is an increasing function of wealth, $U'(x) \geq 0$ always holds for any x. According to L'Hospital rules, we have $\lim_{x\to\infty} \frac{U'(x)}{\alpha x^{\alpha-1}} = C > 0$. Therefore, we obtain $\alpha \geq 0$. In addition, the utility function is a concave function of x, which means that $U''(x) \leq 0$ for any x. Again, according to L'Hospital rules, we have $\lim_{x\to\infty} \frac{U''(x)}{\alpha(\alpha-1)x^{\alpha-2}} = C > 0$. Consequently, we have $\alpha \leq 1$. This completes the proof. □

In a brief summary, when utility function converges to constant as $x \to \infty$, such a constant can be imposed as the Dirichlet boundary for the value function $V(t, x)$ on $x \to \infty$. When utility function goes up to infinity at a polynomial growth rate α, it is natural to assume that the value function $V(\tau, x)$ also has the same growth order as an asymptotic boundary condition, i.e.

$$\lim_{x\to\infty} \frac{V(\tau, x)}{x^\alpha} = f(\tau) > 0, \quad \tau \in [0, T], \tag{7.3.6}$$

which can also be written as

$$\lim_{x\to\infty} \frac{V(\tau, x)}{U(x)} = f(\tau), \tag{7.3.7}$$

In order to avoid determining function $f(\tau)$, we impose a modified asymptotic boundary condition as

$$\lim_{x\to\infty}\frac{\partial}{\partial x}[\frac{V(\tau,x)}{U(x)}]=\lim_{x\to\infty}\frac{V_xU-VU_x}{U^2}=0, \tag{7.3.8}$$

which leads to a Robin boundary condition on $x\to\infty$

$$\lim_{x\to\infty}V_x(\tau,x)=\frac{U_x}{U}V(\tau,x). \tag{7.3.9}$$

For computational purposes, the discretization should be imposed on a bounded domain. The infinite domain in the wealth x direction is truncated as the localized finite domain $[0,X_{\max}]$. As a result, the boundary condition at $x\to\infty$ would be imposed on the boundary $x=X_{\max}$, which would introduce some errors. However, as pointed out by Barles et al. (1995), we can expect the errors incurred by imposing approximate boundary condition at $x=X_{\max}$ to be small in areas of interest if $X_{\max}$ is selected sufficiently large. In other words, by extending the computational domain, it is possible to make the near-field error arbitrarily small. In the following, we assume that the original problem has been localized on a finite domain $[0,X_{\max}]$. Upon imposing the proper boundary conditions which guarantee the uniqueness of the solution on the finite domain, our monotone numerical scheme could produce the converged numerical results.

7.3.2 Discretization

We discretize the HJB equation (7.3.1) over a finite gird defined as

$$\begin{aligned}x_i &= (i-1)\cdot\Delta x, i=1,\cdots,M;\\ \tau_n &= (n-1)\cdot\Delta\tau, n=1,\cdots,N;\end{aligned}$$

where M and N are the number of nodes in x and τ directions, $\Delta x=\frac{X_{\max}}{M-1}$, and $\Delta\tau=\frac{T}{N-1}$. The values at the grid points thus are denoted as $V_i^n=V(\tau_n,x_i)$. The operator $\mathcal{L}^uV$ defined

in Equation (7.3.1) can be discretized as:

$$(\mathcal{L}^u_{\Delta x} V^{n+1})_i = \alpha_i^{n+1} V_{i-1}^{n+1} + \beta_i^{n+1} V_{i+1}^{n+1} - (\alpha_i^{n+1} + \beta_i^{n+1}) V_i^{n+1}, 1 < i < M, \tag{7.3.10}$$

where $\alpha_i^{n+1} = \alpha_i^{n+1}(u^{n+1})$ and $\beta_i^{n+1} = \beta_i^{n+1}(u^{n+1})$; that is, the coefficients of discretized equation are functions of the optimal control u^{n+1}.

First of all, we apply central difference to the second-order derivative and forward or backward difference to the first-order derivative to discretize the operator $\mathcal{L}^u V$. The choice of forward or backward difference depends on the sign of function $b(x_i, u^n)$. Such a scheme is referred to as *forward/backward difference only*. The coefficients in such a scheme are obtained

$$\alpha^n_{i,\text{forward/backward}} = \frac{a(x_i, u^n)}{\Delta x^2} + \max\{-\frac{b(x_i, u^n)}{\Delta x}, 0\}, \tag{7.3.11}$$

$$\beta^n_{i,\text{forward/backward}} = \frac{a(x_i, u^n)}{\Delta x^2} + \max\{\frac{b(x_i, u^n)}{\Delta x}, 0\}. \tag{7.3.12}$$

Condition 7.3.1. (Positive coefficient condition)

$$\alpha_i^n \geq 0, \quad \beta_i^n \geq 0, \qquad i = 1, \cdots, M. \tag{7.3.13}$$

It is obvious that the coefficients (7.3.11)-(7.3.12) satisfies the positive coefficient condition (7.3.13), which is very important when we demonstrate the monotonicity of our numerical scheme (Barles 1997, Forsyth & Labahn 2007).

Since the first-order derivative is discretized using forward or backward difference only, the truncation error such a scheme is only first order. In order to improve the rate of convergence and accuracy, we now provide a modified scheme. Central difference is first applied to all the derivatives and the coefficients are as

$$\alpha^n_{i,\text{central}} = \frac{a(x_i, u^n)}{\Delta x^2} - \frac{b(x_i, u^n)}{2\Delta x}, \tag{7.3.14}$$

$$\beta^n_{i,\text{central}} = \frac{a(x_i, u^n)}{\Delta x^2} + \frac{b(x_i, u^n)}{2\Delta x}. \tag{7.3.15}$$

According to Equation (7.3.2), function $a(x,u)$ is positive while function $b(x,u)$ is not always. Consequently, the coefficients (7.3.14)-(7.3.15) generated by central difference do not always satisfy the positive coefficient condition. When α_i or β_i is negative, oscillations may appear in the numerical solution. To avoid the possible oscillation, we adopt forward or backward difference for the first-order derivative on the nodes when central difference does not generate positive coefficients.

The criteria of the difference method in the modified scheme has changed. On each node, we first apply central difference to all the derivatives and check the sign of the coefficients. If the positive coefficient condition is violated, we adopt forward or backward difference to first-derivative instead to guarantee the new coefficient is positive on this node. In other words, when $\alpha^n_{i,\text{central}} < 0$, we apply forward difference to $\frac{\partial V}{\partial x}$ and obtain

$$\alpha^n_{i,\text{forward}} = \frac{a(x_i,u^n)}{\Delta x^2}, \tag{7.3.16}$$

$$\beta^n_{i,\text{forward}} = \frac{a(x_i,u^n)}{\Delta x^2} + \frac{b(x_i,u^n)}{\Delta x}. \tag{7.3.17}$$

When $\beta^n_{i,\text{central}} < 0$, we adopt backward difference to the first order $\frac{\partial V}{\partial x}$ and have

$$\alpha^n_{i,\text{backward}} = \frac{a(x_i,u^n)}{\Delta x^2} - \frac{b(x_i,u^n)}{\Delta x}, \tag{7.3.18}$$

$$\beta^n_{i,\text{backward}} = \frac{a(x_i,u^n)}{\Delta x^2}. \tag{7.3.19}$$

Otherwise, the coefficients are obtained as (7.3.14)-(7.3.15). Such a modified scheme is referred to as *central difference as much as possible*.

Now the PDE (7.3.1) can be discretized using fully implicit scheme as

$$\frac{V_i^{n+1}-V_i^n}{\Delta\tau} = \max_{u^{n+1}\in\mathbf{R}}\{(\mathcal{L}^{u^{n+1}}_{\Delta x}V^{n+1})_i\} = \max_{u^{n+1}\in\mathbf{R}}\{\alpha_i^{n+1}V_{i-1}^{n+1} + \beta_i^{n+1}V_{i+1}^{n+1} - (\alpha_i^{n+1}+\beta_i^{n+1})V_i^{n+1}\}, \tag{7.3.20}$$

which is a highly nonlinear algebraic equation. This algebraic equation can be written in a matrix form as follows:

$$[I - \Delta\tau A^{n+1}(u^{n+1})]V^{n+1} = V^n + F^n, \tag{7.3.21}$$

where $u_i^{n+1} = \arg\max_{u^{n+1}\in\mathbf{R}} \{(\mathcal{L}_{\Delta x}^{u^{n+1}} V^{n+1})_i\}$, $A^{n+1}(u^{n+1})$ is a tridiagonal matrix as follows:

$$A^{n+1}(u^{n+1}) = \begin{pmatrix} -(\alpha_1^{n+1}+\beta_1^{n+1}) & \beta_1^{n+1} & & & \\ \alpha_2^{n+1} & -(\alpha_2^{n+1}+\beta_2^{n+1}) & \beta_2^{n+1} & & \\ & \ddots & \ddots & \ddots & \\ & \alpha_{M-2}^{n+1} & -(\alpha_{M-2}^{n+1}+\beta_{M-2}^{n+1}) & \beta_{M-2}^{n+1} & \\ & & \alpha_{M-1}^{n+1} & -(\alpha_{M-1}^{n+1}+\beta_{M-1}^{n+1}) \end{pmatrix}.$$

The boundary condition at $x = 0$ defined by Equation (7.3.3) is enforced by setting $\alpha_1 = 0$ and $\beta_1 = 0$. We also modify the coefficients to deal with the boundary condition at $x = X_{\max}$. When a Dirichlet condition is given at $x = X_{\max}$, the size of matrix A is $(M-1)\times(M-1)$ and the vector F^n is defined as

$$\mathrm{F}^n = \Delta\tau \begin{pmatrix} 0 \\ \mathbf{0} \\ -\beta_{M-1}^n V_M^n \end{pmatrix}.$$

When a Robin condition defined by Equation (7.3.9) is given at $x = X_{\max}$, it is discretized and the corresponding coefficients are added to the linear system (7.3.21) as the last row and in that case, the vector F^n could be removed.

For convenience, we rewrite Equation (7.3.20) as a nonlinear algebraic equation

$$G_i^{n+1}(\Delta\tau, \Delta x, V_i^{n+1}, V_{i-1}^{n+1}, V_{i+1}^{n+1}, V_i^n) = 0, \tag{7.3.22}$$

where

$$G_i^{n+1} = \frac{V_i^{n+1} - V_i^n}{\Delta\tau} - \max_{u^{n+1}\in\mathbf{R}} \{(\mathcal{L}_{\Delta x}^{u^{n+1}} V^{n+1})_i\}. \tag{7.3.23}$$

The solution obtained from the nonlinear algebraic equation (7.3.22) is approaching the solution of the HJB equation (7.3.1).

Upon obtaining the value function $V(\tau, x)$ numerically, it is easy to produce the optimal

investment proportion u^* according to the first-order condition for a regular interior maximum of Equation (7.2.10)

$$u^* = -\frac{\mu - r}{\sigma^2 x}\frac{V_x}{V_{xx}}, \tag{7.3.24}$$

where $V_x := \frac{\partial}{\partial x}V(\tau, x)$ and $V_{xx} := \frac{\partial^2}{\partial x^2}V(\tau, x)$. Central finite difference is applied to approximate the optimal investment proportion $u^*(x_i, \tau_n)$ as

$$u_i^{*,n} = -\frac{\mu - r}{\sigma^2 x_i}\frac{V_{i+1} - V_{i-1}}{V_{i+1} - 2V_i + V_{i-1}}\frac{\Delta x}{2}. \tag{7.3.25}$$

7.4 Convergence to the viscosity solution

When we present a numerical scheme for a nonlinear PDE, it is critically important to ensure that the numerical scheme really converges to the viscosity solution. Some examples have been given by Pooley et al. (2003) that seemingly reasonable discretizations of nonlinear option pricing PDE were unstable or converged to the incorrect solution. As pointed out by Barles & Rouy (1998), a numerical scheme converges to the viscosity solution if it is stable, consistent, and monotone. In the following, we follows this idea to demonstrate the convergence of our numerical scheme.

Lemma 7.4.1. (l_∞ stability) If the positive coefficient condition (7.3.13) holds, then the discrete scheme (7.3.20) is stable., i.e.

$$\| V^{n+1} \|_\infty \leq \max\{\| V^0 \|_\infty, C_2, C_3\}, \tag{7.4.1}$$

where $C_2 = \max_n |V_1^n|$, $C_3 = \max_n |V_M^n|$ and V_1^n and V_M^n are the given Dirichlet boundary conditions.

Proof. It follows from Equation (7.3.20) that

$$\begin{aligned}[1 + \Delta\tau(\alpha_i^{n+1} + \beta_i^{n+1})]|V_i^{n+1}| &\leq |V_i^n| + \Delta\tau\alpha_i^{n+1}|V_{i-1}^{n+1}| + \Delta\tau\beta_i^{n+1}|V_{i+1}^{n+1}| \\ &\leq \|V^n\|_\infty + \Delta\tau\alpha_i^{n+1}\|V^{n+1}\|_\infty + \Delta\tau\beta_i^{n+1}\|V^{n+1}\|_\infty, \end{aligned} \tag{7.4.2}$$

where $\alpha_i^{n+1} = \alpha_i^{n+1}(u^{n+1})$, $\beta_i^{n+1} = \beta_i^{n+1}(u^{n+1})$ and $u^{n+1} = \arg\max_{u^{n+1}\in\mathbf{R}}\{(\mathcal{L}_{\Delta x}^{u^{n+1}} V^{n+1})_i\}$.

If $\| V^{n+1} \|_\infty = |V_i^{n+1}|$, $1 < i < M$, then (7.4.2) becomes

$$[1 + \Delta\tau(\alpha_i^{n+1} + \beta_i^{n+1})]\|V^{n+1}\|_\infty \le \|V^n\|_\infty + \Delta\tau\alpha_i^{n+1}||V^{n+1}\|_\infty + \Delta\tau\beta_i^{n+1}\|V^{n+1}\|_\infty, \quad (7.4.3)$$

which leads to

$$\|V^{n+1}\|_\infty \le \|V^n\|_\infty. \quad (7.4.4)$$

If $i = 1$ or $i = M$, then

$$\|V^{n+1}\|_\infty = |V_1^{n+1}| \quad \text{or} \quad \|V^{n+1}\|_\infty = |V_M^{n+1}|. \quad (7.4.5)$$

Combining Equations (7.4.4) and (7.4.5), we have

$$\| V^{n+1} \|_\infty \le \max\{\| V^n \|_\infty, |V_1^{n+1}|, |V_M^{n+1}|\}. \quad (7.4.6)$$

After iterative substitutions, we come to Equation (7.4.1), which indicates the l_∞ stability of the discrete scheme (7.3.20). □

Lemma 7.4.2. (Consistency) The discrete scheme (7.3.20) is consistent, i.e. for any smooth function ϕ, with $\phi_i^n = \phi(\tau_n, x_i)$, we have

$$\lim_{\Delta\tau\to 0,\Delta x\to 0} |(\phi_\tau - \max_{u\in\mathbf{R}}\{\mathcal{L}^u\phi\})_i^{n+1} - G_i^{n+1}(\Delta\tau, \Delta x, \phi_i^{n+1}, \phi_{i-1}^{n+1}, \phi_{i+1}^{n+1}, \phi_i^n)| = 0. \quad (7.4.7)$$

Proof. Suppose $\phi(\tau, x)$ is a smooth test function with bounded derivatives of all order with respect to its variables and denote $\phi_i^n = \phi(\tau_n, x_i)$. After applying Taylor series expansions, we have

$$|\{\mathcal{L}^u\phi\}_i^{n+1} - \{\mathcal{L}_{\Delta x}^u\phi\}_i^{n+1}| = \mathcal{O}(\Delta x). \quad (7.4.8)$$

Then we obtain an estimate as

$$
\begin{aligned}
&|(\phi_\tau - \max_{u\in\mathbf{R}}\{\mathcal{L}^u\phi\})_i^{n+1} - G_i^{n+1}(\Delta\tau, \Delta x, \phi_i^{n+1}, \phi_{i-1}^{n+1}, \phi_{i+1}^{n+1}, \phi_i^n)| \\
&= |(\phi_\tau)_i^{n+1} - \max_{u\in\mathbf{R}}\{\mathcal{L}^u\phi\}_i^{n+1} - [\frac{\phi_i^{n+1}-\phi_i^n}{\Delta\tau} - \max_{u^{n+1}\in\mathbf{R}}\{(\mathcal{L}_{\Delta x}^{u^{n+1}}\phi^{n+1})_i\}]| \\
&\le |(\phi_\tau)_i^{n+1} - \frac{\phi_i^{n+1}-\phi_i^n}{\Delta\tau}| + \max_{u\in\mathbf{R}}|\{\mathcal{L}^u\phi\}_i^{n+1} - \{\mathcal{L}_{\Delta x}^u\phi^{n+1}\}_i^{n+1}| \\
&= \mathcal{O}(\Delta\tau) + \mathcal{O}(\Delta x),
\end{aligned}
$$

where the inequality follows

$$
|\max_x X(x) - \max_y Y(y)| \le \max_x\{|X(x) - Y(x)|\}. \tag{7.4.9}
$$

Then the consistency of the discrete scheme (7.3.20) is demonstrated. □

Definition 7.4.1. (Monotonicity) The discrete scheme (7.3.20) is monotone if for all $\epsilon \ge 0$, we have

$$
G_i^{n+1}(\Delta\tau, \Delta x, V_i^{n+1}, V_{i-1}^{n+1} + \epsilon, V_{i+1}^{n+1} + \epsilon, V_i^n + \epsilon) \le G_i^{n+1}(\Delta\tau, \Delta x, V_i^{n+1}, V_{i-1}^{n+1}, V_{i+1}^{n+1}, V_i^n). \tag{7.4.10}
$$

Lemma 7.4.3. (Monotonicity) If the positive coefficient condition (7.3.13) holds, then the discrete scheme (7.3.20) is monotone.

Proof. When $i = 1$ or $i = M$, the monotonicity of the scheme is obvious. When $1 < i < M$, we have

$$
\begin{aligned}
&G_i^{n+1}(\Delta\tau, \Delta x, V_i^{n+1}, V_{i-1}^{n+1}, V_{i+1}^{n+1}, V_i^n) \\
&= \frac{V_i^{n+1} - V_i^n}{\Delta\tau} + \min_{u^{n+1}\in\mathbf{R}}\{(\alpha_i^{n+1} + \beta_i^{n+1})V_i^{n+1} - \alpha_i^{n+1}V_{i-1}^{n+1} - \beta_i^{n+1}V_{i+1}^{n+1}\}.
\end{aligned}
$$

For any $\epsilon \geq 0$, we have

$$
\begin{aligned}
& G_i^{n+1}(\Delta\tau, \Delta x, V_i^{n+1}, V_{i-1}^{n+1} + \epsilon, V_{i+1}^{n+1}, V_i^n) - G_i^{n+1}(\Delta\tau, V_i^{n+1}, V_{i-1}^{n+1}, V_{i+1}^{n+1}, V_i^n) \\
&= \min_{u^{n+1}\in\mathbf{R}} \{(\alpha_i^{n+1} + \beta_i^{n+1})V_i^{n+1} - \alpha_i^{n+1}V_{i-1}^{n+1} - \beta_i^{n+1}V_{i+1}^{n+1} - \alpha_i^{n+1}\epsilon\}, \\
&\quad - \min_{u^{n+1}\in\mathbf{R}} \{(\alpha_i^{n+1} + \beta_i^{n+1})V_i^{n+1} - \alpha_i^{n+1}V_{i-1}^{n+1} - \beta_i^{n+1}V_{i+1}^{n+1}\}, \\
&\leq \max_{u^{n+1}\in\mathbf{R}} \{-\epsilon\alpha_i^{n+1}\}, \\
&\leq -\epsilon \min_{u^{n+1}\in\mathbf{R}} \{\alpha_i^{n+1}\} \\
&\leq 0,
\end{aligned}
$$

where the first inequality follows

$$
\min_x X(x) - \min_y Y(y) \leq \max_x \{X(x) - Y(x)\}, \tag{7.4.11}
$$

and the last inequality holds due to the fact that $\beta_i^{n+1}(u) \geq 0$ for the positive coefficient condition (7.3.13) holds. Similarly, we have also

$$
\begin{aligned}
G_i^{n+1}(\Delta\tau, \Delta x, V_i^{n+1}, V_{i-1}^{n+1}, V_{i+1}^{n+1} + \epsilon, V_i^n) &\leq G_i^{n+1}(\Delta\tau, \Delta x, V_i^{n+1}, V_{i-1}^{n+1}, V_{i+1}^{n+1}, V_i^n), \\
G_i^{n+1}(\Delta\tau, \Delta x, V_i^{n+1}, V_{i-1}^{n+1}, V_{i+1}^{n+1}, V_i^n + \epsilon) &\leq G_i^{n+1}(\Delta\tau, \Delta x, V_i^{n+1}, V_{i-1}^{n+1}, V_{i+1}^{n+1}, V_i^n).
\end{aligned}
$$

As a result, the monotonicity of the discrete scheme (7.3.20) has been demonstrated. □

Theorem 7.4.1. (Convergence) If the positive coefficient condition (7.3.13) holds, then the discrete scheme (7.3.20) converges to the viscosity solution of the HJB equation (7.3.1).

Proof. Since it has been shown that the discrete scheme (7.3.22) is l_∞ stable, consistent, and monotone, It really converges to the viscosity solution of the HJB equation (7.3.1) following the results of Barles (1997). □

It is also noted that the following property of the matrix $[I - \Delta\tau A^n(u)]$, which is very important when we demonstrate the convergence of our iterates for the nonlinear algebraic

equations in the next section.

Property 7.4.1. (M-matrix) If the positive coefficient condition (7.3.13) holds, then $[I - \Delta\tau A^n(u)]$ is an M-matrix for any $u \in \mathbf{R}$.

Proof. Since the positive coefficient condition holds, $\alpha_i^n(u)$ and $\beta_i^n(u)$ are both positive for any $u \in \mathbf{R}$. Hence, matrix $[I - \Delta\tau A^n(u)]$ has positive diagonals, nonpositive off diagonal, and is diagonally dominant. As a result, it is an M-matrix. □

7.5 Solutions of the nonlinear algebraic equations

After demonstrating the convergence of our numerical scheme, it is still not a practical one and there still exists another obstacle we need to overcome. Solving the nonlinear algebraic equation (7.3.21) at each time step is our another problem, for we adopt fully implicit scheme in time direction. In this chapter, we adopt a popular iterative algorithm to deal with the nonlinear algebraic equations (7.3.21). The details are shown in Algorithm 3.

Algorithm 3 The iterative scheme for the nonlinear algebraic equations (7.3.21)

Require:

Let $\hat{V}^0 := V^n$;

$\hat{V}^k := (V^{n+1})^k$;

Ensure:

1: **while** iteration error is greater than tolerance **do**

2: $k = k + 1$;

3: Solve

$$[I - \Delta\tau A^{n+1}(u^k, \hat{V}^k)]\hat{V}^{k+1} = V^n + F^n,$$

where $u_i^k = \arg\max_{u \in \mathbf{R}} \{\alpha_i(u)\hat{V}_{i-1}^k + \beta_i(u)\hat{V}_{i+1}^k - (\alpha_i(u) + \beta_i(u))\hat{V}_i^k\}$;

4: error $= \| \hat{V}^{k+1} - \hat{V}^k \|_{l_\infty}$;

5: **end while**

Now we would demonstrate that such an iterative scheme is convergent. First of all, after some manipulations of the Algorithm 3, we have

$$[I - \Delta\tau A^{n+1}(u^k)](\hat{V}^{k+1} - \hat{V}^k) = \Delta\tau[A^{n+1}(u^k)\hat{V}^k - A^{n+1}(u^{k-1})\hat{V}^k]. \tag{7.5.1}$$

Before proving the convergence of Algorithm 3, we need such a lemma:

Lemma 7.5.1. In Algorithm 3, the optimal control is determined by

$$u_i^k = \arg\max_{u\in\mathbf{R}}\{\alpha_i^{n+1}(u)\hat{V}_{i-1}^k + \beta_i^{n+1}(u)\hat{V}_{i+1}^k - (\alpha_i^{n+1}(u) + \beta_i^{n+1}(u))\hat{V}_i^k\}, \tag{7.5.2}$$

then every element of the right hand of equation (7.5.1) is nonnegative, that is ,

$$[A^{n+1}(u^k)\hat{V}^k - A^{n+1}(u^{k-1})\hat{V}^k]_i \geq 0. \tag{7.5.3}$$

Proof. According to the selection of u^k in Equation (7.5.2), we have

$$A^{n+1}(u^k)\hat{V}^k = \max_{u\in\mathbf{R}}\{A^{n+1}(u)\hat{V}^k\}, \tag{7.5.4}$$

for given $\hat{V}^k$. As a result, any other choice of u would lead to the inequality (7.5.3). □

With Lemma 7.4.1 in hand, it is easy to demonstrate that the iterative scheme converges.

Theorem 7.5.1. (Convergence of the iterative scheme) Provided that Property 7.4.1 holds, i.e. $[I - \Delta\tau A^n(u)]$ is an M-matrix, then the iterative scheme described in Algorithm 3 converges to the unique solution of the nonlinear algebraic equation for any initial iterate $\hat{V}^0$. Moreover, the iterates $\{\hat{V}^k\}$ converge monotonically.

Proof. First of all, it is proved that the iterates $\{\hat{V}^k\}$ are bounded. For $1 < i < M$, we have

$$[1 + \Delta\tau(\alpha_i^k + \beta_i^k)]\hat{V}_i^{k+1} = V_i^n + \Delta\tau\beta_i^k\hat{V}_{i-1}^{k+1} + \Delta\tau\alpha_i^k\hat{V}_{i-1}^{k+1}, \tag{7.5.5}$$

where $\alpha_i^k = \alpha_i^{n+1}(u^k)$ and $\beta_i n + 1^k = \beta_i(u^k)$. Consequently, we obtain

$$[1 + \Delta\tau(\alpha_i^k + \beta_i^k)]|\hat{V}_i^{k+1}| \leq \Delta\tau(\alpha_i^k + \beta_i^k) \parallel \hat{V}^{k+1} \parallel_\infty + \parallel V^n \parallel_\infty . \tag{7.5.6}$$

Similar to the proof of Lemma 7.5.1 we come to a conclusion that

$$\parallel \hat{V}^{k+1} \parallel_\infty \leq \max\{\parallel V^n \parallel_\infty, |\hat{V}_1^{k+1}|, |\hat{V}_M^{k+1}|\} = \max\{\parallel V^n \parallel_\infty, |V_1^n|, |V_M^n|\}, \tag{7.5.7}$$

where V_1^n and V_M^n are given Dirichlet boundary conditions at each time step τ_n. Consequently, the iterates $\{\hat{V}^k\}_{k=0}^{\infty}$ is bounded independent of iteration k

The next step is to prove that these iterates are non-decreasing sequences. After some simple calculation, we obtain

$$[I - \Delta\tau A^{n+1}(u^k)](\hat{V}^{k+1} - \hat{V}^k) = \Delta\tau[A^{n+1}(u^k) - A^{n+1}(u^{k-1})]\hat{V}^k. \tag{7.5.8}$$

According to Lemma 7.5.1, every element of the right side of Equation (7.5.8) is nonnegative. In addition, according to the Property 7.4.1 that $[I - \Delta\tau A^n(u)]$ is an M-matrix, we have $[I - \Delta\tau A^n(u)]^{-1} \geq 0$, it is obvious that the iterates $\{\hat{V}^k\}_{k=0}^{\infty}$ form a bounded non-decreasing sequence. When $\hat{V}^{k+1} = \hat{V}^k$, the iteration converges to the solution of the nonlinear algebraic equation. The uniqueness of the solution follows the Property 7.4.1 that $[I - \Delta\tau A^n(u)]$ is an M-matrix. □

7.6 Numerical examples

Since our numerical scheme can be applied to solve the HJB equation arising from the Merton problem with general utility functions, three examples are presented in this section. First of all, the utility function is considered to be of a power form (Example 1) and an exponential form (Example 2). In the literature, analytical solutions for the two examples have been obtained and they are considered as the benchmark solution to verify our numerical results. In Example 3, a new mixed power utility function, which shows non-constant relative risk aversion, is adopted. In each example, the numerical results at $\tau = T$ are reported with two difference schemes: central difference as much as possible and forward/backward difference only.

All the computations reported in this chapter were performed with Matlab R2016a on 64-bit quad-core Intel 2.83GHz system with 16GB of RAM. The parameters in the following examples are set as

$$\mu = 0.1, \quad r = 0.05, \quad \sigma = 0.5, \quad T = 1, \quad X_{\max} = 100. \tag{7.6.1}$$

7.6.1 Example 1: power utility function

In the first example, we consider the power utility function defined as

$$U(x) = \frac{x^\gamma}{\gamma}, \tag{7.6.2}$$

where $\gamma < 1$ and $\gamma \neq 0$. The Arrow-Pratt measure of relative risk aversion of such a utility is defined as

$$\delta[U(\cdot)] = -\frac{U''(x)}{U'(x)}x = 1 - \gamma. \tag{7.6.3}$$

Obviously, the power utility function (7.6.2) belongs to the CRRA class as its Arrow-Pratt measure is constant. When we implement the numerical scheme for Example 1, the parameter γ is set to be $\frac{1}{2}$ and the Robin boundary condition at $x \to \infty$ is given as

$$\lim_{x\to\infty} V_x(\tau, x) = \frac{U_x(x)}{U(x)}V(\tau, x) = \frac{\gamma}{x}V(\tau, x). \tag{7.6.4}$$

On the boundary $x = X_{\max}$, it is disctetized as $\dfrac{V_M - V_{M-1}}{\Delta x} = \dfrac{\gamma}{x_M}V_M$.

Actually, when the utility function is of power form, Zariphopoulou (1999) has derived an analytical solution to the original HJB equation (7.2.9) as follows:

$$V(\tau, x) = \frac{x^\gamma}{\gamma}e^{D\tau}, \tag{7.6.5}$$

where $D = r\gamma - \frac{\gamma}{\gamma-1}\frac{(\mu-r)^2}{2\sigma^2}$. In the following, such an analytical solution would be considered as a benchmark to verify the numerical results obtained from our monotone scheme.

The numerical results of $V(T, x)$ calculated with two difference schemes, using central difference as much as possible and using forward/backward difference only, are reported in Table 7.1 and Table 7.2 with $\gamma = \frac{1}{2}$. The results calculated from the analytical solution (7.6.5) are considered as the benchmark when we report the l_∞ errors for our numerical results. The convergence tolerance in Algorithm 3 is set to be 10^{-5} and it is the same in the following two examples. The last column of Tables 7.1 and 7.2 is the ratio of successive l_∞ errors as the

grid is refined by a factor of two and the time step sizes are reduced by a factor of four. The numerical order of convergence is then defined by

$$\text{Rate} = \log_2(\text{Ratio}). \tag{7.6.6}$$

The *iter* is the average number of the nonlinear iteration for each time step.

(M,N)	$x = 3$	$x = 4$	$x = 5$	$x = 6$	*iter*	l_∞ error	ratio
(100,10)	3.568289	4.120992	4.607631	5.047567	1.80	1.310×10^{-3}	
(200,40)	3.569273	4.121555	4.608112	5.047985	1.95	3.262×10^{-4}	4.0
(400,160)	3.569462	4.121714	4.608247	5.048103	1.9875	1.373×10^{-4}	2.4
(800,640)	3.569534	4.121769	4.608292	5.048140	1.996875	6.519×10^{-5}	2.1
Benchmark	3.569600	4.121818	4.608332	5.048176			

Table 7.1: Numerical results for value function by applying central difference as much as possible when utility function is $\frac{x^\gamma}{\gamma}$ with $\gamma = \frac{1}{2}$.

(M,N)	$x = 3$	$x = 4$	$x = 5$	$x = 6$	*iter*	l_∞ error	ratio
(100,10)	3.558002	4.111932	4.599573	5.040239	1.80	1.160×10^{-2}	
(200,40)	3.564122	4.117111	4.604144	5.044367	1.95	5.477×10^{-3}	2.1
(400,160)	3.566910	4.119506	4.606274	5.046302	1.9875	2.690×10^{-3}	2.0
(800,640)	3.568262	4.120668	4.607307	5.047241	1.996875	1.337×10^{-3}	2.0
Benchmark	3.569600	4.121818	4.608332	5.048176			

Table 7.2: Numerical results for value function by applying forward/backward difference only when utility function is $\frac{x^\gamma}{\gamma}$ with $\gamma = \frac{1}{2}$.

It is noted that the successive l_∞ errors in both Table 7.1 and Table 7.2 are approaching to zero as the grid spacing is diminished, which shows a clear convergence trend to the benchmark solution (7.6.5). In addition, both Table 7.1 and Table 7.2 show that the average number of nonlinear iteration is about two, which indicates that the nonlinear iteration converges rapidly. The results obtained using cental difference as much as possible in Table 7.1 are more accurate than those obtained using forward or backward difference only in Table 7.2 on the same grid, because the central difference has higher order of accuracy. Especially on the grid $(M, N) = (800, 640)$, the l_∞ error in Table 7.1 is of 10^{-5} level; while the counterpart in Table 7.2 is of 10^{-3} level. Moreover, the ratio in Table 7.2 is around two, which implies that the numerical order of convergence is one according to Equation (7.6.6). If the central difference is

applied to all the nodes, the ratio in Table 7.2 is expected to four. Although the ratio in Table 7.1 is greater than those in Table 7.2, it has not reached four, which implies that there must be some nodes on which the coefficients generated by the central difference is negative. To make them satisfy the positive coefficient condition, forward or backward difference is applied instead on these nodes.

Based on the analytical solution (7.6.5), the analytical form of the optimal investment proportion should be

$$u^* = -\frac{\mu - r}{x\sigma^2}\frac{V_x}{V_{xx}} = \frac{\mu - r}{\sigma^2(1-\gamma)}, \tag{7.6.7}$$

which is constant. Numerically, the optimal investment proportion $u^*(T, x)$ can also be computed according to Equation (7.3.25). Comparisons between the numerical results and the analytical form (7.6.7) are shown in Figure 7.1.

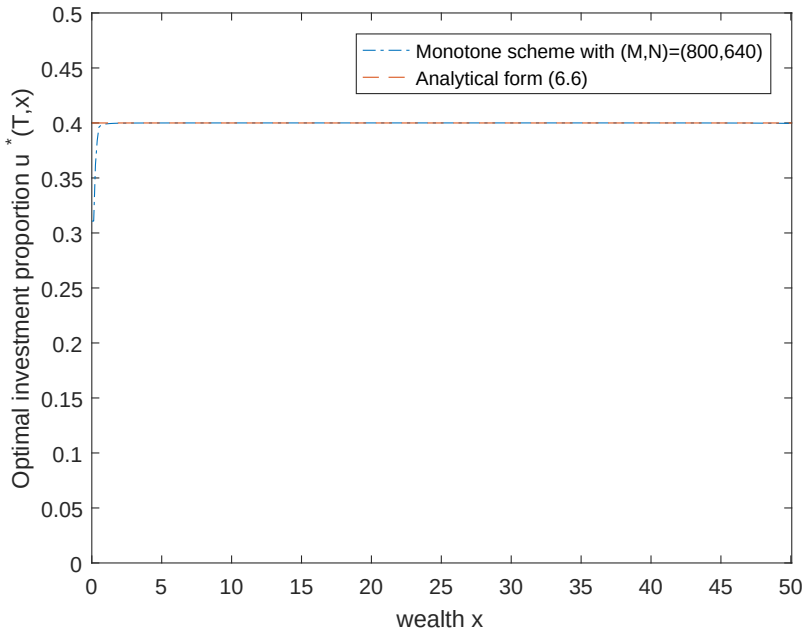


Figure 7.1: Optimal investment proportion u^* for utility $U(x) = \frac{x^\gamma}{\gamma}$ with $\gamma = \frac{1}{2}$.

It is observed in Figure 7.1 that the numerical results match with the analytical form (7.3.25) very well, which demonstrates that our numerical scheme converges to the exact solution. It is also noted that there are some computational error near the boundary $x = 0$, which is introduced by imposing the approximate boundary condition.

7.6.2 Example 2: exponential utility function

In this example, an exponential utility function defined as

$$U(x) = -\frac{e^{-\eta x}}{\eta}, \eta > 0, \tag{7.6.8}$$

is taken into account to demonstrate that our numerical scheme could deal with the cases with non-constant relative risk aversion utility function successfully. The Arrow-Pratt measure of relative risk aversion for this case is:

$$\delta[U(\cdot)] = -\frac{U''(x)}{U'(x)}x = \eta x, \tag{7.6.9}$$

which is an increasing function of x, instead of being constant.

In this case, the exponential utility function would converge to zero as x tends toward infinity. As a result, we impose a Dirichlet boundary condition as

$$\lim_{x\to\infty} V(\tau, x) = 0. \tag{7.6.10}$$

Actually, we have an alternative boundary condition based on the financial reasoning of exponential utility function. As shown in Equation (7.6.9), the risk aversion of the investor increases as his wealth grows up. When wealth x approaches towards infinity, the investor's risk aversion goes extremely and consequently, he would allocate all of his money on the risk-free asset, i.e. $u^* = 0$. As an approximation, the optimal investment proportion u^* should be zero at $x = X_{\max}$. Upon substituting $u^* = 0$ into the PDE (7.3.1), we have a degenerate equation as

$$\frac{\partial V}{\partial \tau} = rx\frac{\partial V}{\partial x}, \tag{7.6.11}$$

which can be adopted as an appropriate boundary condition at $x = X_{\max}$. In this chapter, we only adopt the Dirichlet boundary condition (7.6.10) to implement our numerical scheme.

In fact, the analytical solution to the Merton problem with exponential utility function was

derived by Ma & Zhu. (2017) Ma & Zhu. (2017) as follows

$$V(\tau, x) = -\frac{1}{\eta} e^{-C\tau - \eta x e^{r\tau}}, \tag{7.6.12}$$

where $C = \frac{(\mu - r)^2}{2\sigma^2}$.

Two discretization methods, using central difference method as much as possible and using forward/backward difference only, are implemented respectively with $\eta = 1$. The numerical results and convergence analysis are reported in Tables 7.3 and 7.4. The results calculated from the analytical solution (7.6.12) is considered as the benchmark to calculate the l_∞ error.

(M,N)	$x = 1$	$x = 2$	$x = 3$	$x = 4$	*iter*	l_∞ error	ratio
(100,10)	-0.359053	-0.128931	-0.046289	-0.016615	1.80	1.130×10^{-2}	
(200,40)	-0.353126	-0.125029	-0.044268	-0.015673	1.950	5.376×10^{-3}	2.1
(400,160)	-0.350458	-0.123289	-0.043372	-0.015258	1.9875	2.708×10^{-3}	2.0
(800,640)	-0.347739	-0.122421	-0.042927	-0.015053	1.996875	8.843×10^{-4}	3.1
Benchmark	-0.347750	-0.121536	-0.042476	-0.014845			

Table 7.3: Numerical results for value function by applying central difference as much as possible when utility function is $U(x) = -\frac{e^{-eta}}{\eta}$ with $\eta = 1$.

(M,N)	$x = 1$	$x = 2$	$x = 3$	$x = 4$	*iter*	l_∞ error	ratio
(100,10)	-0.359053	-0.128931	-0.046289	-0.016615	1.80	1.130×10^{-2}	
(200,40)	-0.353126	-0.125029	-0.044268	-0.015674	1.95	5.376×10^{-3}	2.1
(400,160)	-0.350460	-0.123289	-0.043372	-0.015258	1.9875	2.710×10^{-3}	2.0
(800,640)	-0.349121	-0.122421	-0.042927	-0.015053	1.996875	1.371×10^{-3}	2.0
Benchmark	-0.347750	-0.121536	-0.042476	-0.014845			

Table 7.4: Numerical results for value function by applying forward or backward difference only when utility function is $U(x) = -\frac{e^{-eta}}{\eta}$ with $\eta = 1$.

The numerical result in both Table 7.1 and 7.2 are converging to the benchmark solution by observing the l_∞ error. Comparing the results in both tables, we find that there is no significant difference except the results on grid $(M, N) = (800, 640)$. It means that central difference for first-derivative is adopted only on few nodes because the coefficients generated by central difference not violate the positive coefficient condition on most nodes. On grid $(M, N) = (800, 640)$, the l_∞ error is improved from level of 10^{-3} to level of 10^{-4} and the ratio of successive l_∞ error is improved from 2.0 to 3.1, which implies that the central difference

can really improve the order of accuracy and convergence if it does not violate the positive coefficient conditions. In this example, the nonlinear iteration also converges quickly with the average number of nonlinear iteration being around two in both Table 7.3 and Table 7.4.

According to the analytical solution (7.6.12), the optimal investment proportion can be expressed as

$$u^* = -\frac{\mu - r}{x\sigma^2}\frac{V_x}{V_{xx}} = \frac{\mu - r}{\sigma^2 \eta e^{r\tau} x}. \tag{7.6.13}$$

When implementing our monotone scheme, we replace the computational domain $[0, X_{\max}]$ with $[\epsilon, X_{\max}]$ where $\epsilon = 10^{-3}$ since $x = 0$ is a singular point. The comparisons between the numerical results and the analytical form (7.6.13) are shown in Figure 7.2. Most part of our numerical results are in good agreement with the results from the analytical form (7.6.13). The truncation error is relative large near the point $x = 0$ because it is a singular point. In order to reduce the truncation error near $x = 0$, we have to refine the grids by increasing the number of grid points along wealth direction. It is also pointed out that applying a non-uniform grid near the singular point $x = 0$ may be a good alternative, which we would not discuss in details here.

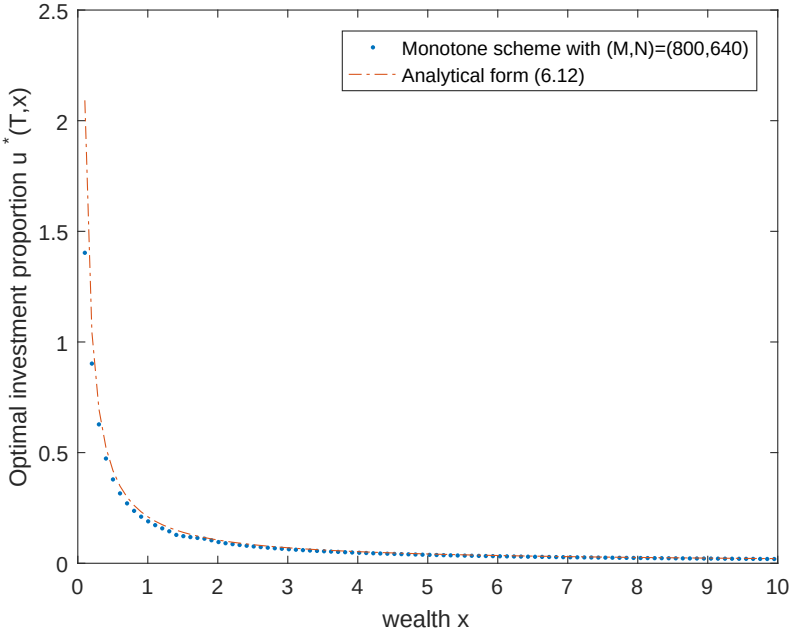


Figure 7.2: Optimal investment proportion u^* for utility function $U(x) = -\frac{e^{-eta}}{\eta}$ with $\eta = 1$..

7.6.3 Example 3: mixed power utility function

Recently, Fouque et al. (2015) presented a new utility function of mixed form to characterize an investor whose relative risk aversion varies with his wealth. The mixed utility function is taking the form as

$$U(x) = kU_1 + (1-k)U_2, k \in [0,1], \tag{7.6.14}$$

where $U_1 = \frac{x^{\gamma_1}}{\gamma_1}$ and $U_2 = \frac{x^{\gamma_2}}{\gamma_2}$. Its Arrow-Pratt measure is expressed as:

$$\delta[U(\cdot)] = \frac{U''(x)}{U'(x)}x = \frac{k(1-\gamma_1) + (1-k)(1-\gamma_2)x^{\gamma_2-\gamma_1}}{k + (1-k)x^{\gamma_2-\gamma_1}}. \tag{7.6.15}$$

Obviously, this mixed utility function is non-constant relative risk aversion with $k \in (0,1)$. Although it is just a simple sum of two power utility function in Example 1, such a linear combination results in a total nonlinear result due to the nonlinearity of the PDE.

In this example, the parameters are set as $k = 0.5$, $\gamma_1 = 0.15$, and $\gamma_2 = 0.85$. Obviously, the mixed utility function would blow up as x approaches infinity and its growth order is 0.85. As a result, the boundary condition at $x \to \infty$ is given as

$$\lim_{x\to\infty} V_x(\tau, x) = \frac{U_x(x)}{U(x)}V(\tau, x). \tag{7.6.16}$$

Since there is no analytical solution for this example in the literature, the results calculated at a very fine grid $(M, N) = (1600, 5120)$ are chosen as the benchmark solution. Numerical results computed with two difference schemes are reported in Tables 7.5 and 7.6.

(M,N)	$x = 2$	$x = 4$	$x = 6$	$x = 8$	*iter*	l_∞ error	ratio
(25,10)	4.840164	6.161236	7.255904	8.229597	1.80	2.048×10^{-2}	
(50,40)	4.850215	6.163160	7.259521	8.246478	1.950	3.602×10^{-3}	5.7
(100,160)	4.850961	6.163560	7.260065	8.249110	1.9875	9.703×10^{-4}	3.7
(200,640)	4.851130	6.163654	7.260179	8.249782	1.996875	2.978×10^{-4}	3.3
(1600,5120)	4.851182	6.163685	7.260220	8.250080			

Table 7.5: Numerical results for value function by applying central difference as much as possible for the mixed utility function (7.6.14) with $k = 0.5$, $\gamma_1 = 0.15$, and $\gamma_2 = 0.85$.

It is observed that the ratio of successive l_∞ errors in Table 7.5 is significantly greater than

(M,N)	$x=2$	$x=4$	$x=6$	$x=8$	*iter*	l_∞ error	ratio
(25,10)	4.829856	6.154625	7.250093	8.222792	1.80	2.722×10^{-2}	
(50,40)	4.845531	6.160000	7.256893	8.243977	1.95	6.032×10^{-3}	4.5
(100,160)	4.848670	6.162002	7.258780	8.247941	1.9875	2.371×10^{-3}	2.5
(200,640)	4.849994	6.162880	7.259541	8.249210	1.996875	1.047×10^{-3}	2.3
(1600,5120)	4.851041	6.163588	7.260140	8.250001			

Table 7.6: Numerical results for value function by applying forward or backward difference only for the mixed utility function (7.6.14) with $k=0.5$, $\gamma_1=0.15$, and $\gamma_2=0.85$.

two; while it is about two in Table 7.6. Obviously, the modified scheme using cental difference as much as possible does really improve the rate of convergence from two to four. The accuracy of the scheme is also improved from 10^{-3} to 10^{-4} on the finest girds. Both Table 7.5 and Table 7.6 show that the average number of nonlinear iteration is about two, which indicates that the nonlinear iteration converges rapidly.

Our numerical results can be used to verify a series solution to the HJB equation (7.2.9). Zhu & Ma (2018) applied the Homotopy Analysis Method (HAM) to solve the HJB equation with the mixed power utility function (7.6.14) and obtained an explicit series solution as follows:

$$V(\tau,x)=\sum_{n=0}^{+\infty}\frac{V_n(\tau,x)}{n!}. \tag{7.6.17}$$

However, when the numerical results are required, their infinite series solution has to be truncated as a finite one.

$$S_L=\sum_{n=0}^{L}\frac{V_n(\tau,x)}{n!}. \tag{7.6.18}$$

Here we would like to verify their series solution with our numerical results to confirm how many terms is necessary for the truncated series solution. Their series solution is obtained with symbolic calculation. Here we list first two terms as follows:

$$V_0(\tau,x)=\frac{k}{\gamma_1}(xe^{r\tau})^{\gamma_1}+\frac{1-k}{\gamma_2}(xe^{r\tau})^{\gamma_2}, \tag{7.6.19}$$

$$V_1(\tau,x)=\frac{C\tau[(1-k)(xe^{r\tau})^{\gamma_2}+k(xe^{r\tau})^{\gamma_1}]^2}{k(1-\gamma_1)(xe^{r\tau})^{\gamma_1}+(1-k)(1-\gamma_2)(xe^{r\tau})^{\gamma_2}}. \tag{7.6.20}$$

where $C=\frac{(\mu-r)^2}{2\sigma^2}$. Actually, more terms V_n can be explicitly expressed but it becomes more

and more complicated as n increases. For example, the third term $V_2(\tau, x)$ is too messy to express it in full parameters and we substitute the values of some parameters to simplify it as follows:

$$\begin{aligned} V_2(\tau,x) = -C^2\tau^2(&13722y^{4.45} + 13280y^{3.05} + 18491y^{3.75} - 28.226y^{7.95} + 1579y^{5.85} \\ &+ 5941y^{5.15} + 85.79y^{7.25} + 0.2y^{8.65} + 256.4y^{6.55} + 4112.9y^{2.35} + 199.58y^{1.65} \\ &- 1.448y^{9.35})/(672.5y^{3.55} - 43180.66y^{2.15} - 7620y^{2.85} - 122345y^{1.45} - 138658y^{0.75} \\ &- 23.83y^{4.25})^2|_{y=xe^{r\tau}}. \end{aligned} \tag{7.6.21}$$

In this chapter, we only show the series solution with the first three terms and demonstrate such a few terms has converged. The details how these terms are derived haven been provided by Zhu & Ma (2018).

To verify their series solution, the numerical results calculated on grid $(M, N) = (1600, 5120)$ using central difference as much as possible is considered as the reference solution to report the l_∞ error. The numerical results obtained from truncated series solution with L terms are shown in Table 7.7.

Method	S_L	$x = 2$	$x = 4$	$x = 6$	$x = 8$	l_∞ error
HAM	S_0	4.832737	6.128867	7.208718	8.182219	6.79×10^{-2}
	S_1	4.850979	6.163165	7.259378	8.249383	8.42×10^{-4}
	S_2	4.851178	6.163637	7.260200	8.249924	1.56×10^{-4}
Reference solution		4.851182	6.163685	7.260220	8.250080	

Table 7.7: Comparison of values calculated with different terms.

From Table 7.7, the l_∞ error decreases rapidly from 10^{-2} to 10^{-4} as the number of truncated term, L, increases from 0 to 2, which implies that the series solution with only first three terms is really in good agreement with our numerical results. In other words, the truncated series solution S_2 is indeed a good analytical approximate solution to the HJB equation (7.2.9) with the mixed power utility function. Our numerical solution and the approximate truncated series solution with merely three terms can verify with each other perfectly.

Finally, the optimal investment proportion u^* calculated from the monotone scheme and

the truncated series solution $S_2(T,x)$ are presented in Figure 7.3. These curves obtained by numerical scheme and the truncated series solution $S_2(T,x)$ are almost indistinguishable except that some error appear near $x = 0$, which is introduced by approximate boundary condition. More discussion about the optimal investment proportion u^* would be presented in next section.

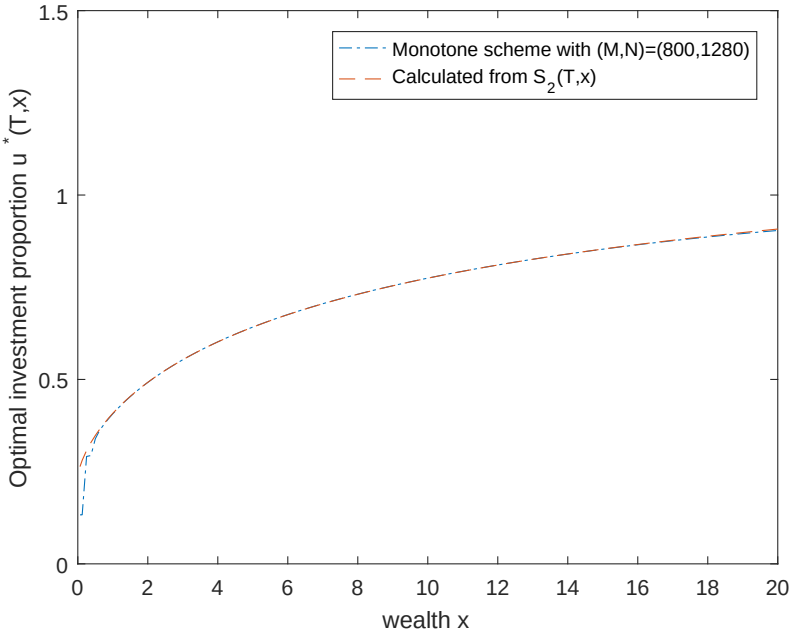


Figure 7.3: Optimal investment proportion u^* for the mixed utility function (7.6.14) with $k = 0.5$, $\gamma_1 = 0.15$, and $\gamma_2 = 0.85$.

7.7 Economic discussions

In last section, three different utility functions have been taken as examples to demonstrate the accuracy and versatility of the monotone numerical scheme. These three utility functions actually represent three different kinds of investors. We provide some economic discussions about the corresponding optimal investment policies in different examples.

The power utility function in Example 1 belongs to the CRRA class, which indicates that the investor's attitude to the risk would never change no matter how much money he has. Due to the constant relative risk aversion, the optimal investment proportion u^* is also constant in

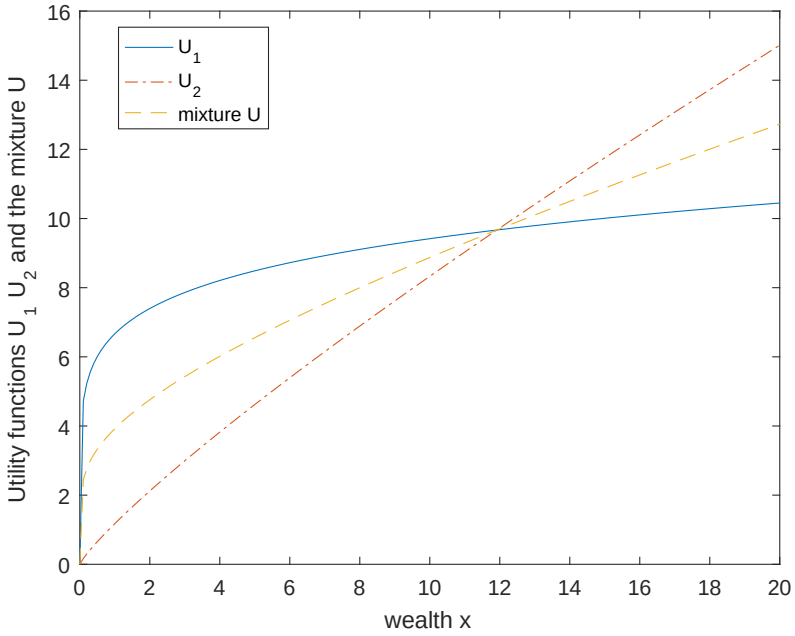
Example 1, which was ever observed by Merton (1969).

The exponential utility function in Example 2 is obviously not a member of the CRRA class, because its Arrow-Pratt measure of relative risk is an increasing function of wealth x as shown in Equation (7.6.9). In other words, the more money the investor has, the more risk aversion he shows. As a result, the corresponding optimal investment proportion u^* is a decreasing function of wealth x as shown in Figure 7.2. In other words, with his wealth rising up, the investor prefers the risk-free asset to the risky asset and consequently, he would allocate more and more money on the risk-free asset.

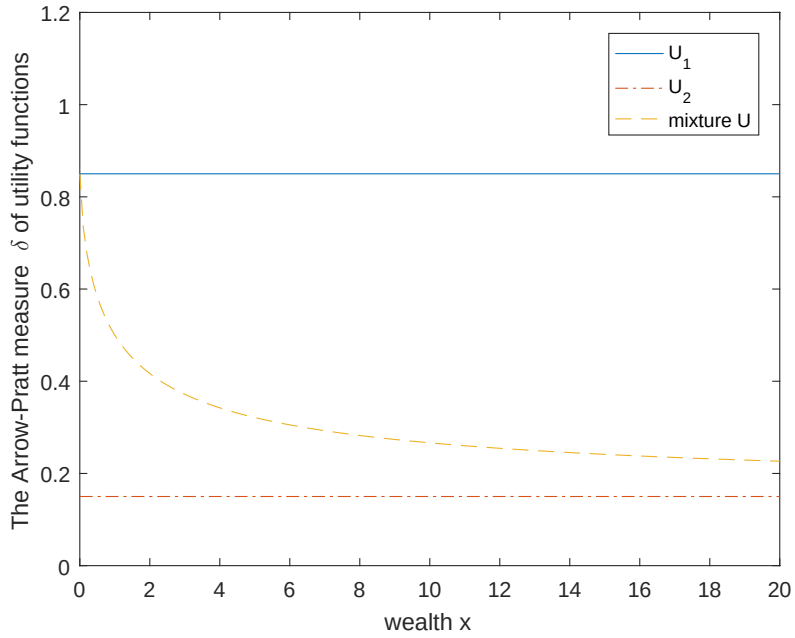
Since the analytical solutions for the first two examples have been obtained in the literature, the corresponding economic discussions have been provided by Merton (1969) and Ma & Zhu. (2017). In this chapter, discussions are mainly focused on the new mixed utility function in Example 3, since no closed-form analytical solution has been obtained so far. From the definition of the mixed power utility function (7.6.14), it is indeed a linear combination of two power utility functions as shown in Example 1. When we set $k = 1$ (or $k = 0$), it degenerates to a single power utility function U_1 (or U_2). According to Equation (7.6.15), the Arrow-Pratt measure becomes a function of wealth, instead of being constant in the case of single power utility function. In other words, it is a non-constant relative risk aversion utility function if it is not degenerate. One can appreciate more of the complexity associated with the mixture of two simple power utilities through their visualization shown in Figure 7.4(a) (a graphic display of U_1, U_2 and the mixture U) and Figure 7.4(b) (the corresponding Arrow-Pratt measure of relative risk aversion) with a set of parameters being

$$k = 0.5, \gamma_1 = 0.15, \gamma_2 = 0.85. \tag{7.7.1}$$

From Figure 7.4(b), The Arrow-Pratt measure of the mixed utility function U varies with wealth x; while that of the single power utility function U_1 or U_2 is a constant. In contract to the exponential utility function in Example 2, the Arrow-Pratt measure in Example 3 is



(a) Utility function U_1, U_2 and the mixture U



(b) Arrow-Pratt risk aversion $\delta = -\frac{U''}{U'}x$

Figure 7.4: Mixed power utility with $k = 0.5, \gamma_1 = 0.15, \gamma_2 = 0.85$.

a decreasing function of wealth x, which implies that the risk aversion would be reduced as his wealth increases. This different behavior of risk aversion results in a significantly different optimal investment proportion u^* as shown in Figure 7.3. As the wealth increases, the optimal investment proportion u^* increases, too. In other words, the more money the investor has, the less risk aversion he shows and consequently, he allocate higher proportion of his wealth on the risky asset.

From Figure 7.4(b), it is noted that the Arrow-Pratt measure of the mixed utility function is bounded by that of U_2 and U_1 from blew and above. It is expected that the optimal investment proportion corresponding to the mixture should also be bounded by the counterpart of U_1 and U_2. The optimal investment proportion u^* corresponding to U_1, U_2 and the mixture U are pictured in Figure 7.5 with a large enough value of $X_{\max}$.

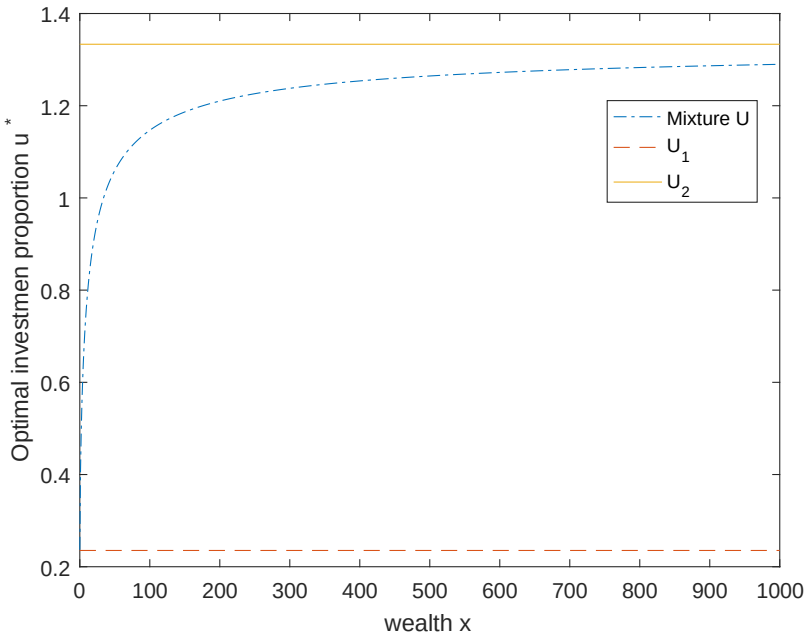


Figure 7.5: Optimal investment proportion u^*.

Obviously, the optimal investment proportion u^* is also bounded by that corresponding to the single utility function U_1 and U_2 as shown in Figure 7.5, which agrees with what we expect.

7.8 Conclusions

In this chapter, we present a monotone scheme for the HJB equation arising from the Merton problem with general utility functions. To ensure the monotonicity of the scheme, the discretization is applied to guarantee that the positive coefficient condition always holds. After demonstrating the l_∞ stability, consistency, and monotonicity, we prove that our numerical scheme really converges to the viscosity solution of the HJB equation.

To show the accuracy and versatility of the monotone scheme, three different utility functions are taken as examples. In the first two examples, the analytical solutions obtained in the literature are adopted to verify the results of the numerical scheme. In the last example, after demonstrating the convergence of our scheme, we also in turn use the numerical results to verify a truncated series solution obtained with homotopy analysis method. Some economic discussions about the optimal investment proportion u^* are provided finally.

Chapter 8

Option pricing with short selling bans being imposed

8.1 Introduction

During the Global Financial Crisis 2007-2009, most regulatory authorities around the world imposed restrictions or bans on short selling to reduce the volatility of financial market and to limit the negative impacts of a downturn market (Beber & Pagano 2013). These interventions were implemented to restore the orderly financial markets and limit drops in stock price. However, these regulations imposed on short selling also result in some new problems, one of which is how to price options or contingent claims. In this chapter, we focus on our interest on the valuation of the contingent claim in a financial market with short selling being banned.

In a complete market, any contingent claim can be replicated perfectly by some self-financing dynamic portfolio strategies and, under a no-arbitrage condition, the price of contingent claim must equal to the cost of constructing such a portfolio. However, imposing short selling ban makes it impossible to hedge a contingent claim perfectly even in the classic Black-Scholes model, which implies that the market becomes incomplete due to the restriction or even ban on short selling. In the literature, how to price contingent claims in an incomplete market has been studied extensively and a large number of approaches and techniques have been provided.

The literature can be grouped into two categories.

Chapters in the first category share a common feature that an equivalent martingale measure is chosen as pricing measure according to some optimal criterion. Follmer & Schweizer (1991) first proposed a criterion to choose the *minimal martingale measure* in order to price option in incomplete market. Then *minimal entropy martingale measure* was proposed by Frittelli (2000) to minimize the entropy difference between the objective probability measure and the risk-neutral measure. Similar concepts, such as the *minimal distance martingale measure* and *minimax measure* were also put forward by Goll & Rüschendorf (2001) and Bellini & Frittelli (2002), respectively. Each measure will lead to a different price, which is "fair" according to the criteria they chose the measure. It is hard to justify which choice of these equivalent martingale measures is "correct".

Chapters in the second category include Karatzas & Kou (1996), Davis (1997), Rouge & El Karoui (2000), Musiela & Zariphopoulou (2004) and Hugonnier et al. (2005) . The key idea of these chapters is utility indifference pricing. An investor chooses a utility function first according to his risk preference. The utility indifference buying price p^b is the price at which the utility of the investor is indifferent between (1) paying nothing and not having the claim and (2) paying p^b now to receive the contingent claim at expire time (Henderson and Hobson, 2004). The utility indifference selling price is defined similarly. In finance literature, utility indifference price is also referred to as "private valuation", which emphasizes the proposed price is for an individual with particular risk preference and not a transactional price (Detemple & Sundaresan 1999, Tepla 2000). In contrast to Black-Scholes price, utility indifference price is nonlinear due to the concavity of the utility function. In addition, it degenerates to the unique fair price when the market is complete.

For any contingent claim in an incomplete market, El Karoui & Quenez (1995) demonstrate that there will be a price interval within which the price must lead to arbitrage opportunities. The maximum price of this interval, called *selling price*, is the lowest price that allows the seller to hedge completely with an optimal hedging strategy. Similarly, the minimum price of this interval, called *buying price*, is the highest price that the buyer is willing to pay for a

contingent claim. Both of these two concepts have been addressed in the literature on contingent claims hedging and pricing under transactions cost (Hodges 1989, Davis 1997, Constantinides & Zariphopoulou 1999, Munk 1999). Obviously, either selling price or buying price is a private price for the seller or buyer because they just consider to minimize unilateral risk. The buyer and seller have to negotiate and compromise with each other in order to reach an agreement on the transactional price.

Recently, Guo & Zhu (2017) proposed a completely new approach, referred to as the *equal-risk pricing approach*, which determines the derivative price by simultaneously analyzing the risk exposure of both parties involved in the contract. It appears to be, but not the same as, the existing utility indifference price method as pointed out in Remark 3.2 in their chapter. They aimed to find out an equal-risk price which distributes expected loss evenly between the two involved parties. Such an *equal-risk price* is interpreted as a fair price that both parties are happy to accept during the negotiation if they intend to enter into a derivative contract. Equal-risk price is a transactional price and it must lie in the price interval consisting of selling price and buying price. Both the seller and buyer would face the same amount of risk when they accept such a price. They also established the existence and uniqueness of equal-risk price for arbitrage European and American options and demonstrated that their model is consistent with standard arbitrage-free pricing model to cover the degenerated case when the market is complete. Although they have derived an analytical pricing formula for European call and put options, it is still hard to extend their analysis method for a general contingent claim, which has limited the application of equal-risk pricing approach.

The main contribution of this chapter is that we have established a PDE framework for the *equal-risk pricing approach* in order to expand the range of its application. Under our PDE framework, we first recover the analytical pricing formula for European call and put options which demonstrates that our PDE approach is consistent with the previous work of Guo & Zhu (2017). Furthermore, we provide an efficient numerical scheme to solve the PDE system when the payoff function is non-monotonic. Taking a butterfly spread option as an example, we compare its equal-risk price from this new pricing approach with Black-Scholes price to show

how short selling bans affect the valuation of contingent claims.

The chapter is organized as follows. In Section 8.2, a financial market with short selling ban is introduced first. The PDE framework is established to derive equal-risk price of general contingent claims. In Section 8.3, analytical pricing formulae are derived when claims are European call and put options. In Section 8.4, a finite difference numerical scheme is introduced to solve the PDE system and two numerical experiments are conducted accordingly. Conclusions are provided in the last section.

8.2 A framework of equal-risk pricing approach

8.2.1 The financial market model

Consider a financial model on a complete probability space $(\Omega, \mathcal{F}, \mathbb{Q})$. Let $\mathbb{F} = \{\mathcal{F}_t : t > 0\}$ be the filtration that represents the information flow available to market participants. For simplicity, we assume there are only two assets traded continuously in this market. One is a risk-free asset, the price of which satisfies the ordinary differential equation

$$dP_t = rP_t dt, \tag{8.2.1}$$

where r is the risk-free interest rate. The other one is a risky asset with its price following the Black-Scholes model

$$dS_t = rS_t dt + \sigma S_t dW_t, \tag{8.2.2}$$

where σ is the volatility of the underlying and W_t is a standard Brownian motion. Since this is the first chapter to set up a PDE framework for equal-risk pricing approach, we choose the simple Black-Scholes model to illustrate how the short selling bans affect the derivative price. Of course, some complicated stock models, such as stochastic volatility and interest rate models, can also be adopted under our PDE framework in the future.

Without short selling bans, market is complete and there exists a unique equivalent martingale measure $\mathbb{Q}$. The price of European contingent claims that expires at time T with a payoff

function $Z(S_T)$ can be easily calculated as $v = \mathbf{E}_{\mathbb{Q}}[e^{-rT}Z(S_T)]$. Such a price is accepted by both the seller and buyer of the claim since they are able to perfectly replicate the claim by corresponding self-financing trading strategies.

When short selling is banned, market becomes incomplete for perfect replication is not valid. In this case, an admissible self-financing trading strategy is a progressively measurable non-negative process ϕ_t, which represent the shares of stock at time t. Given an initial wealth v, an investor, who adopt the trading strategy ϕ_t, would hold ϕ_t shares of stock at time t and leaves all the remaining on the risk-free bond account. Then the wealth process of such a portfolio, denoted as v_t, follows

$$dv_t = \underbrace{d(\phi_t S_t)}_{\text{stock account}} + \underbrace{d(v_t - \phi_t S_t)}_{\text{free-risk account}} = \phi_t dS_t + r(v_t - \phi_t S_t)dt = rv_t dt + \phi_t \sigma S_t dW_t, \tag{8.2.3}$$

where ϕ_t comes from the set of all self-financing, progressively measurable, non-negative and square integrable trading strategy

$$\mathbf{\Phi} := \{\phi(t,\omega) : [0,T] \times \Omega \to R^+| \quad \mathbf{E}\int_0^T \phi^2(t,\omega)dt < \infty, \quad \phi \geq 0\}, \tag{8.2.4}$$

where the positivity condition on ϕ represents the short-selling ban in the market.

8.2.2 Equal-risk price for general contingent claims

As mentioned above, short selling bans have made perfect hedging strategy impossible even the stock price follows the classic Black-Scholes model. Applying any unilateral utility-based arguments would lead to a price interval consisting of buying price as the lower bound and selling price as the upper bound. Any price that lies in this interval would make both the buyer and seller face risk whatever they do to hedge. Intuitively, a higher price of the contingent claim increases the risk exposure of the buyer and decreases that of the seller; while a lower price has the opposite effect. Guo & Zhu (2017) proposed an idea to look for a price that lies in this interval and distributed the risk between the buyer and seller equally. To measure the

risk exposure, they introduced the risk function as follows.

Definition 8.2.1. A function $R : \mathbb{R} \to \mathbb{R}$ is called a *risk function* if it satisfies the following conditions:

1. $R(x)$ is non-decreasing convex and has a finite lower bound LB.

2. $R(0) = 0$ and $R(x) > 0$ for all $x > 0$.

Remark 8.2.1. It is easy to check that both $R_1(x) = x^+$ and $R_2(x) = e^x - 1$ are risk functions we defined. The former is adopted by Guo & Zhu (2017), while the latter is the one we choose in this chapter. Here we provide two reasons for our choice. From the view of mathematics, $R_2(x)$ is a smooth function, while $R_1(x)$ is not. In addition, $R_1(x)$ maps all the negative x to zero. However, from the view of finance, a company that owns one million dollars should be more riskless than a company that has only one dollar. $R_1(x)$ cannot tell the difference between these two companies, while $R_2(x)$ can do it.

Suppose an investor has the opportunity to sell one unit of European contingent claim $Z(S_T)$ at a transaction price v. After receiving the payment, he would establish an hedging account with his initial wealth v. $Z(S_T)$ is a future liability for the seller, which represents the possible risk. The terminal wealth process of the hedging account v_T is an income that reduces the risk he takes at expire date T. As a result, the risk exposure of the seller who sells a European contingent claim $Z(S_T)$ at a price v with the current stock price S is defined

$$\rho^s(S, v; Z) = \inf_{\phi(\cdot)\in\mathbf{\Phi}} \mathbf{E}_{\mathbb{Q}}^{S,v} R(Z(S_T) - v_T^{v,\phi(\cdot)}), \tag{8.2.5}$$

where $\mathbb{E}_{\mathbb{Q}}^{S,v}$ denotes the conditional expectation under the measure $\mathbb{Q}$ with $S_0 = S, v_0 = v$ and $v_t^{v,\phi(\cdot)}$ is the solution of Equation (8.2.3) given trading strategy $\phi(\cdot)$ and initial wealth v.

How to calculate the minimum risk exposure for the seller has becomes an optimal stochastic control problem with objective function (8.2.5) and dynamics S_t and v_t governed by Equations (8.2.2) and (8.2.3). According to the dynamic programming method (Yong & Zhou 1999), the

HJB equation governing the value function $F^s(t,S,v)$ is derived as

$$\begin{cases} 0 = \dfrac{\partial F^s}{\partial t} + \inf\limits_{\phi \geq 0} \mathcal{L}_1^{\phi} F^s, \\ F^s(T,S,v) = R(Z(S) - v), \end{cases} \tag{8.2.6}$$

where

$$\mathcal{L}_1^{\phi} F = \frac{1}{2} S^2 \sigma^2 \frac{\partial^2 F}{\partial S^2} + \phi S^2 \sigma^2 \frac{\partial^2 F}{\partial S \partial v} + \frac{1}{2} S^2 \sigma^2 \phi^2 \frac{\partial^2 F}{\partial v^2} + rS \frac{\partial F}{\partial S} + rv \frac{\partial F}{\partial v}. \tag{8.2.7}$$

The value function at $t = 0$ corresponds to the minimum risk exposure of the seller, i.e. $F^s(0,S,v) = \rho^s(S,v;Z)$.

Similarly, we come to the analysis of the buyer's risk exposure. Assume the buyer offers a price v for a European contingent claim $Z(S_T)$ and his offer is accepted by a seller. The buyer has to borrow money v right now to purchase the claim, which corresponds to the deterministic liability ve^{rT} for the buyer at the expire date. Although the initial wealth is zero, the buyer would also establish a hedging account with a hedging strategy $\phi(\cdot)$, which comes from the admissible set Φ defined in (8.2.4). Then the risk exposure of the buyer, who pays v to purchase a European contingent claim $Z(S_T)$ with current stock price S, is defined as

$$\rho^b(S,v;Z) = \inf_{\phi(\cdot) \in \mathbf{\Phi}} \mathbf{E}_{\mathbb{Q}}^{v,S}[R(ve^{rT} - v_T^{0,\phi(\cdot)} - Z(S_T))] = \inf_{\phi(\cdot) \in \mathbf{\Phi}} \mathbf{E}_{\mathbb{Q}}^{v,S}[R(v_T^{v,-\phi(\cdot)} - Z(S_T))]. \tag{8.2.8}$$

To solve this optimal stochastic control problem associated with the buyer, another HJB equation governing the value function $F^b(t,S,v)$ is also established as

$$\begin{cases} 0 = \dfrac{\partial F^b}{\partial t} + \inf\limits_{\phi \geq 0} \mathcal{L}_2^{\phi} F^b, \\ F^b(T,S,v) = R(v - Z(S)), \end{cases} \tag{8.2.9}$$

where

$$\mathcal{L}_2^{\phi} F = \frac{1}{2} S^2 \sigma^2 \frac{\partial^2 F}{\partial S^2} - \phi S^2 \sigma^2 \frac{\partial^2 F}{\partial S \partial v} + \frac{1}{2} S^2 \sigma^2 \phi^2 \frac{\partial^2 F}{\partial v^2} + rS \frac{\partial F}{\partial S} + rv \frac{\partial F}{\partial v}. \tag{8.2.10}$$

Remark 8.2.2. It is pointed out that the differences between these two HJB equations (8.2.6)

and (8.2.9) lie in the sign of the cross-derivative term and the terminal condition.

From the view of finance, a buyer who purchases a European contingent claim $Z(S)$ at a price v is equivalent to a seller who sells a contingent claim $-Z(S)$ at the price of $-v$. Mathematically, it is expressed as

$$\rho^b(S, v; Z) = \rho^s(S, -v; -Z). \tag{8.2.11}$$

Such a relation plays an important role in the rest of this chapter.

Functions $\rho^s(S, v; Z)$ and $\rho^b(S, v; Z)$ represent the minimum risk exposure of the seller and buyer through selecting an optimal hedging strategy when the transactional price of claim is v and the underlying stock price is S. The following lemma describes some properties of these functions.

Lemma 8.2.1. Assume that Z, Z_1, Z_2 are square integrable, $\mathcal{F}_T$-measurable random variables. The monotonicity and limits behavior of both risk functions $\rho^s(S, v; Z)$ and $\rho^b(S, v; Z)$ are described as follows:

1. If $Z_1 \leq Z_2$, then $\rho^s(S, v; Z_1) \leq \rho^s(S, v; Z_2)$ and $\rho^b(S, v; Z_1) \geq \rho^b(S, v; Z_2)$.
   If $v_1 \leq v_2$, then $\rho^s(S, v_1; Z) \geq \rho^s(S, v_2; Z)$ and $\rho^b(S, v_1; Z) \leq \rho^b(S, v_2; Z)$.

2. As v tends toward ∞ or $-\infty$, the asymptotic behavior of them are

$$\lim_{v\to\infty} \rho^s(S, v; Z) = \text{LB}, \lim_{v\to\infty} \rho^b(S, v; Z) = \infty,$$
$$\lim_{v\to-\infty} \rho^s(S, v; Z) = \infty, \lim_{v\to-\infty} \rho^b(S, v; Z) = LB.$$

Proof. We leave the proof of Lemma 8.2.1 in Appendix C.1. □

Now we present the definition of equal-risk price for a European contingent claim based on the risk exposure functions $\rho^s(S, v; Z)$ and $\rho^b(S, v; Z)$.

Definition 8.2.2. Consider a European contingent claim Z. When the underlying stock price is S and the transactional price of this claim is v, the minimum risk for the seller and buyer

are denoted as $\rho^s(S, v; Z)$ and $\rho^b(S, v; Z)$, respectively. Then equal-risk price of this claim with current price S is the price $\bar{v}(S)$ such that the seller and buyer face the same amount of risk, i.e.

$$\rho^s(S, \bar{v}(S); Z) = \rho^b(S, \bar{v}(S); Z). \tag{8.2.12}$$

In order to demonstrate that equal-risk price is well-defined, the following theorem states its existence and uniqueness.

Theorem 8.2.1. Consider a market where the stock follows the Black-Scholes model and short selling is banned. For a European contingent claim $Z(S_T)$, there exists a unique equal-risk price $\bar{v}(S)$ such that it satisfies the following equation,

$$\rho^s(S, \bar{v}(S); Z) = \rho^b(S, \bar{v}(S); Z). \tag{8.2.13}$$

Proof. The proof of this theorem is left in Appendix C.2. □

In a brief summary, the equal-risk pricing approaching consists of two steps. In the first step, we calculate the risk exposure of the seller and buyer respectively through solving two stochastic optimal control problems. In the second step, equal-risk price is implied by Equation (8.2.12). Obviously, the first step is the significantly important and complicated. To solve the stochastic control problems associated with the seller and buyer, we have to deal with two nonlinear PDE systems (8.2.6) and (8.2.9). For some special contingent claims, the corresponding HJB equations can be solved analytically and the pricing formula for equal-risk price can be derived easily. However, for general claims, analytical solution of these HJB equations are unavailable and hence numerical scheme would be an alternative to solve them.

8.3 Equal-risk price of European call and put options

When the contingent claim is a European call option, the PDE systems (8.2.6) and (8.2.9) can be solved analytically and the risk exposure of the seller and buyer is derived in the following propositions.

Proposition 8.3.1. When the contingent claim is a European call option with payoff $Z(S) = (S-K)^+$, the seller's risk exposure is

$$\rho^s(S,v;Z) = R(e^{rT}[C^{BS}(S,K,r,\sigma,T) - v]), \tag{8.3.1}$$

where $C^{BS}(S,K,r,\sigma,T)$ is the classic Black-Scholes formula for a European call option with the underlying price S, strike price K, risk-free interest rate r, volatility σ and time to expiration $T-t$.

Proof. In order to derive the risk exposure of the seller, we would focus on the PDE system (8.2.6) with $Z = (S-K)^+$. Consider a trial solution to the PDE system (8.2.6) as

$$F^s(t,S,v) = R(e^{r(T-t)}[C^{BS}(S,K,r,\sigma,T-t) - v]). \tag{8.3.2}$$

Assuming that $R(x)$ is twice differential, it follows from the chain rule that

$$\begin{aligned}
\tfrac{\partial F^s}{\partial t} &= R'e^{r(T-t)}(\tfrac{\partial C^{BS}}{\partial t} - rC^{BS} + rv), & \tfrac{\partial F^s}{\partial S} &= R'e^{r(T-t)}\tfrac{\partial C^{BS}}{\partial S},\\
\tfrac{\partial F^s}{\partial v} &= -R'e^{r(T-t)}, & \tfrac{\partial^2 F^s}{\partial S\partial v} &= -R''e^{2r(T-t)}\tfrac{\partial C^{BS}}{\partial S},\\
\tfrac{\partial^2 F^s}{\partial S^2} &= R''e^{2r(T-t)}(\tfrac{\partial C^{BS}}{\partial S})^2 + R'e^{r(T-t)}\tfrac{\partial^2 C^{BS}}{\partial S^2}, & \tfrac{\partial^2 F^s}{\partial v^2} &= R''e^{2r(T-t)}.
\end{aligned}$$

Based on the convexity of function $\mathcal{L}_1 F$ with respect to ϕ, the optimal hedging strategy is

$$\phi^* = \max\{-\frac{\partial^2 F^s}{\partial S\partial v}(\frac{\partial^2 F^s}{\partial v^2})^{-1}, 0\} = \max\{\frac{\partial C^{BS}}{\partial S}, 0\} \tag{8.3.3}$$

The Black-Scholes Delta of a European call option $\dfrac{\partial C^{BS}}{\partial S}$ is always non-negative, which leads to $\phi^* = \dfrac{\partial C^{BS}}{\partial S}$. After substituting ϕ^* back into the HJB equation (8.2.6), we have

$$\begin{aligned}
&\frac{\partial F^s}{\partial t} + \inf_{\phi\geq 0}\{\frac{1}{2}\sigma^2 S^2\frac{\partial^2 F^s}{\partial S^2} + \phi S^2\sigma^2\frac{\partial^2 F^s}{\partial S\partial v} + \frac{1}{2}S^2\sigma^2\phi^2\frac{\partial^2 F^s}{\partial v^2} + rS\frac{\partial F^s}{\partial S} + rv\frac{\partial F^s}{\partial v}\}\\
&= R'e^{r(T-t)}[\frac{\partial C^{BS}}{\partial S} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 C^{BS}}{\partial S^2} + rS\frac{\partial C^{BS}}{\partial S} - rC^{BS}],\\
&= 0.
\end{aligned} \tag{8.3.4}$$

The last equation holds just because C^{BS} satisfies the Black-Scholes PDE. Consequently, the trial solution (8.3.2) is exactly the solution to the HJB equation (8.2.6). Therefore, the risk exposure of the seller is expressed as (8.3.1) because $\rho^s(S, v; Z) = F^s(0, S, v)$. □

Remark 8.3.1. The seller of European call options would adopt the same optimal hedging strategy in the classic Black-Scholes model, i.e. $\phi^* = \frac{\partial C^{BS}}{\partial S}$, which means that the ban of short selling does not affect his hedging strategy. That is because such an optimal hedging strategy for European call options is always non-negative when the payoff function is $Z(S_T) = (S_T - K)^+$ which is non-decreasing with respect to S.

Proposition 8.3.2. When the contingent claim is a European call option with payoff $Z(S) = (S - K)^+$, the buyer's risk exposure is

$$\rho^b(S, v; Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} R(ve^{rT} - (Se^{(r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}x} - K)^+)e^{-\frac{x^2}{2}} dx. \tag{8.3.5}$$

Proof. We first claim that the optimal hedging strategy ϕ^* for the buyer should be zero when $Z(S) = (S - K)^+$, i.e.

$$\rho^b(S, v; Z) = \mathbf{E}_{\mathbb{Q}} R(v_T^{v,0} - Z). \tag{8.3.6}$$

It suffices to demonstrate that $\mathbf{E}_{\mathbb{Q}} R(v_T^{v,-\phi(\cdot)} - Z) \geq \mathbf{E}_{\mathbb{Q}} R(ve^{rT} - Z)$ for any $\phi(\cdot) \in \Phi$. According to the dynamics (8.2.3), we have $v_t^{v,-\phi(\cdot)} = ve^{rt} - \sigma \int_0^t e^{r(t-u)} \phi_u S_u dW_u$. Since $R(x)$ is a convex function, we have

$$\mathbf{E}_{\mathbb{Q}}[R(v_T^{v,-\phi(\cdot)} - Z) - R(ve^{rT} - Z)] \geq \mathbf{E}_{\mathbb{Q}}[-R'(ve^{rT} - Z)\sigma \int_0^T e^{r(T-u)} \phi_u S_u dW_u]. \tag{8.3.7}$$

According to the martingale representation theorem, random variable $-R'(ve^{rT} - Z(S_T))$ can be expressed as

$$-R'(ve^{rT} - Z(S_T)) = -\mathbf{E}_{\mathbb{Q}} R'(ve^{rT} - Z(S_T)) + \int_0^T \psi_u \sigma S_u dW_u, \tag{8.3.8}$$

which financially indicates that such a random variable is decomposed into an initial wealth

$-\mathbf{E}_{\mathbb{Q}}R'(ve^{rT}-Z)$ and a trading strategy $\psi(\cdot)$, which is a predictable process. Obviously, such a random variable is non-decreasing of S_T. Following Lemma 3.2 in Guo & Zhu (2017), we come to a conclusion that such a trading strategy $\psi(\cdot)$ is non-negative. Based on stochastic calculation, we have

$$\mathbf{E}_{\mathbb{Q}}[-R'(ve^{rT}-Z)\sigma\int_0^T e^{r(T-u)}\phi_u S_u dW_u] = \mathbf{E}_{\mathbb{Q}}\int_0^T \sigma^2 e^{r(T-u)}\phi_u\psi_u S_u^2 du \geq 0, \tag{8.3.9}$$

which completes the proof of our claim (8.3.6). Since the optimal trading strategy ϕ^* is zero, the HJB equation (8.2.9) becomes

$$\begin{cases} 0 = \dfrac{\partial F^b}{\partial t} + \dfrac{1}{2}\sigma^2 S^2\dfrac{\partial^2 F^b}{\partial S^2} + rS\dfrac{\partial F^b}{\partial S} + rv\dfrac{\partial F^b}{\partial v}. \\ F^b(T,S,v) = R(v-Z(S)). \end{cases} \tag{8.3.10}$$

By introducing time reversal $\tau = T-t$ and function $G(\tau,S,v) = F^b(t,S,v)$, we have

$$\begin{cases} \dfrac{\partial G}{\partial \tau} = \dfrac{1}{2}\sigma^2 S^2\dfrac{\partial^2 G}{\partial S^2} + rS\dfrac{\partial G}{\partial S} + rv\dfrac{\partial G}{\partial v}, \\ G(0,S,v) = R(v-Z(S)). \end{cases} \tag{8.3.11}$$

According to Feynman-Kac formula, the solution the such a linear PDE system can be written as a condition expectation

$$\begin{aligned} G(\tau,S,v) &= \mathbf{E}_{\mathbb{Q}}^{v,S}R(v_\tau^{v,0} - (S_\tau - K)^+) \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} R(ve^{r\tau} - (Se^{(r-\frac{\sigma^2}{2})\tau+\sigma\sqrt{\tau}x} - K)^+)e^{-\frac{x^2}{2}}dx. \end{aligned} \tag{8.3.12}$$

The risk exposure is expressed as (8.3.5) since $\rho^b(S,v;Z) = F^b(0,S,v) = G(T,S,v)$. □

Remark 8.3.2. It is noted that the optimal hedging strategy for the buyer of European call options is doing nothing when short selling is banned, which is totally different from the counterpart in the classic Black-Scholes model. The reason is that the optimal hedging strategy in the classic Black-Scholes model $\phi^* = -\frac{\partial C^{BS}}{\partial S}$ is non-positive, which is infeasible due to the short selling bans.

After deriving the risk exposure of both the seller and buyer, the analytical pricing formula for European call options is provided in the following theorem.

Theorem 8.3.1. When short selling is banned in the Black-Scholes model, equal-risk price of European call options is produced as follows according to different risk functions.

1. When risk function is $R(x) = x^+$, equal-risk price v is implied by

$$v = C^{BS}(S, K, r, \sigma, T) - [P^{BS}(S, K + ve^{rT}, r, \sigma, T) - P^{BS}(S, K, r, \sigma, T)], \tag{8.3.13}$$

   where $P^{BS}(S, K, r, \sigma, T)$ is classic Black-Scholes formula for a European put option with underlying price S, strike price K, risk-free interest rate r, volatility σ and time to expiration T.

2. When risk function is $R(x) = e^x - 1$, equal-risk price v is explicitly expressed as

$$v = \frac{1}{2}\{C^{BS}(S, K, r, \sigma, T) - e^{-rT} \ln [\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(Se^{(r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}x}-K)^+-\frac{x^2}{2}} dx]\}. \tag{8.3.14}$$

Proof. The risk exposure of both the seller and buyer have been derived in Propositions 8.3.1 and 8.3.2. According to Definition 8.2.2, equal-risk price of European call options is the root of Equation

$$\rho^s(S, v; (S - K)^+) = \rho^b(S, v; (S - K)^+). \tag{8.3.15}$$

When risk function is taken to be $R(x) = x^+$ and $R(x) = e^x - 1$, equal-risk price is derived easily as Equations (8.3.13) and (8.3.14) after some simple calculations. □

Remark 8.3.3. It is remarked that the analytical pricing formula is the same as the one provided by Guo & Zhu (2017) when the risk function is assume to $R(x) = x^+$, which demonstrates that our PDE approach is consistent with Guo and Zhu's method. In addition, we produce an explicit and analytical pricing formula as Equation (8.3.14) when a new risk function $R(x) = e^x - 1$ is adopted. The significant difference between these two formula is that

Equation (8.3.13) is not explicit and it has to be solved by root finding algorithm; while Equation (8.3.14) is explicit.

The pricing formula (8.3.13) has been interpreted in terms of the standard Black-Schole prices and an adjustment term in Guo and Zhu's chapter. In this chapter, we mainly focus on the new explicit equal-risk price (8.3.14) when risk function is assumed to be $R(x) = e^x - 1$. To illustrate how the short selling bans affect the European call option price, we figure out the results computed from the equal-risk pricing formula (8.3.14) and those calculated from the classic Black-Scholes formula in Figure 8.1(a) with the parameters being set as

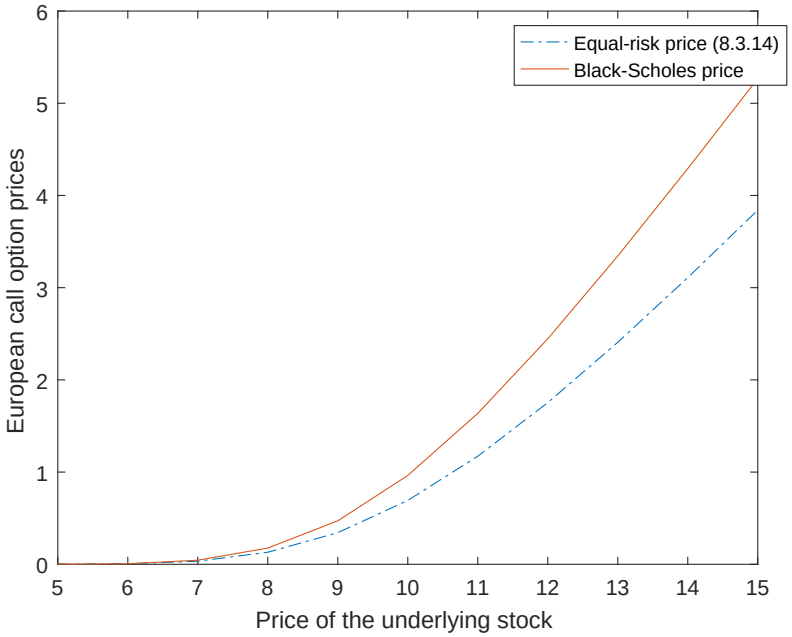
$$K = 10, r = 0.05, T = 0.5, \sigma = 0.3. \tag{8.3.16}$$

As shown in Figure 8.1(a), the absolute difference between equal-risk price and Black-Scholes price is significant for the large underlying price, which indicates that the short selling bans affect the option price substantially. To demonstrate the effect for the small underlying price, the relative difference between equal-risk price and Black-Scholes price is characterized by the percentage distance to Black-Scholes price defined by

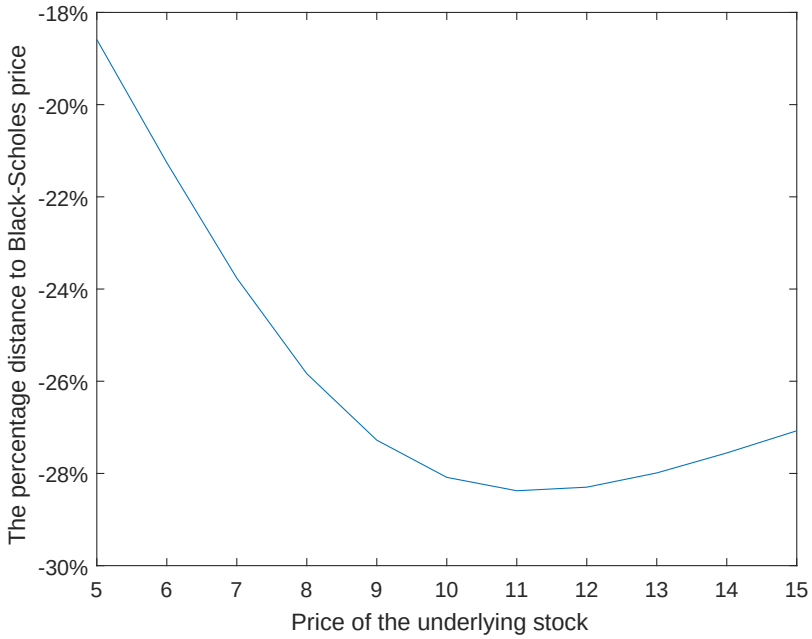
$$\frac{\text{Equal-risk price} - \text{Black-Scholes price}}{\text{Black-Scholes price}} \times 100\%, \tag{8.3.17}$$

which is depicted in Figure 8.1(b). It is also observed that the relative difference is substantial although the absolute difference is not significantly large for small underlying price. From Figures 8.1(a) and 8.1(b), we draw a conclusion that the short selling bans would significantly decrease the European call option price for both small and large underlying prices.

From Propositions 8.3.1 and 8.3.2, the optimal hedging in the classic Black-Scholes model is still available for the seller of European call options; while the counterpart is unavailable for the buyer due the short selling bans. If the transaction price of contingent claim is still set to be Black-Scholes price, the seller would face no risk; while the buyer could not eliminate the risk totally because the optimal hedging strategy is infeasible now. To transfer some risk from buyer to seller so that both of them face the same amount, equal-risk price should be lower



(a) European call option price.



(b) The percentage distance

Figure 8.1: Comparisons between equal-risk price and Black-Scholes price for European call options.

than Black-Scholes price. The premium between these two prices is used to compensate the buyer because he takes too much risk due to the ban of short selling.

According to the relation (8.2.11) between the risk exposure of the buyer and seller, we can derive equal-risk price for European put options as corollaries.

Corollary 8.3.1. When the contingent claim is a European put option with payoff $Z(S) = (K-S)^+$, the buyer's risk exposure is

$$\rho^b(S, v; Z) = R(e^{rT}[v - P^{BS}(S, K, r, \sigma, T)]). \tag{8.3.18}$$

Proof. Consider the seller's risk exposure for a contingent claim $-(K-S)$ first. To calculate $\rho^s(S, v; -(K-S))$, we need to solve the corresponding HJB equation

$$\begin{cases} 0 = \dfrac{\partial F^s}{\partial t} + \inf\limits_{\phi \geq 0} \mathcal{L}_1^{\phi} F^s, \\ F^s(T, S, v) = R(-(K-S)^+ - v). \end{cases} \tag{8.3.19}$$

With the same technique in Proposition 8.3.1, the solution can be produced as

$$F^s(t, S, v) = R(e^{r(T-t)}[-P^{BS}(S, K, r, \sigma, T-t) - v]). \tag{8.3.20}$$

According to the relation (8.2.11), we have

$$\rho^b(S, v; (K-S)^+) = \rho^s(S, -v; -(K-S)^+) = F^s(0, S, -v) = R(e^{rT}[v - P^{BS}(S, K, r, \sigma, T)]).$$

□

Corollary 8.3.2. When the contingent claim is a European put option with payoff $Z(S) = (K-S)^+$, the seller's risk exposure is

$$\rho^s(S, v; Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} R((K - Se^{(r-\frac{\sigma^2}{2})T + \sigma\sqrt{T}x})^+ - ve^{rT}) e^{-\frac{x^2}{2}} dx. \tag{8.3.21}$$

Proof. Consider the buyer's risk exposure for a contingent claim $-(K-S)^+$ first. To compute

$\rho^b(S, v; -(K-S))$, we goes to the HJB equation

$$\begin{cases} 0 = \dfrac{\partial F^b}{\partial t} + \inf\limits_{\phi \geq 0} \mathcal{L}_2^{\phi} F^b, \\ F^b(T, S, v) = R(v + (K-S)^+). \end{cases} \tag{8.3.22}$$

Similar to Proposition 8.3.2, the solution to such a PDE system is

$$F^b(t, S, v) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} R(ve^{r(T-t)} + (K - Se^{(r-\frac{\sigma^2}{2})(T-t)+\sigma\sqrt{T-t}x})^+) e^{-\frac{x^2}{2}} dx. \tag{8.3.23}$$

From relation (8.2.11), the seller's risk exposure of European put options is

$$\begin{aligned} \rho^s(S, v; (K-S)^+) &= \rho^b(S, -v; -(K-S)^+) = F^b(0, S, -v) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} R((K - Se^{(r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}x})^+ - ve^{rT}) e^{-\frac{x^2}{2}} dx. \end{aligned}$$

□

Corollary 8.3.3. When short selling is banned in the Black-Scholes model, equal-risk price of European put options is derived according to different risk functions.

1. When risk function is $R(x) = x^+$, equal-risk price is implied by

$$v = P^{BS}(S, K, r, T, \sigma) + P^{BS}(S, K - ve^{rT}, r, T, \sigma). \tag{8.3.24}$$

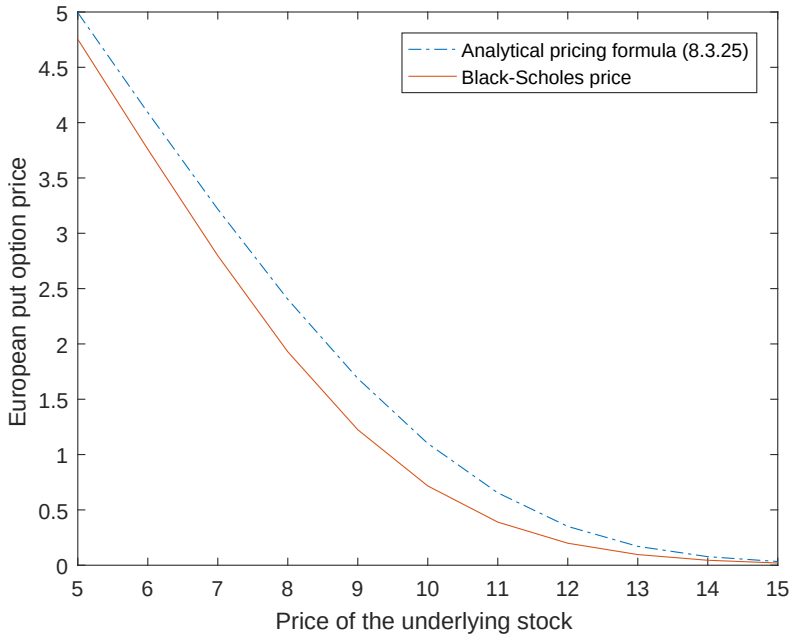
2. When risk function is $R(x) = e^x - 1$, equal-risk price is explicitly expressed as

$$v = \frac{1}{2}\{P^{BS}(S, K, r, T, \sigma) + e^{-rT} \ln \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{(K - Se^{(r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}x})^+ - \frac{x^2}{2}} dx\}. \tag{8.3.25}$$

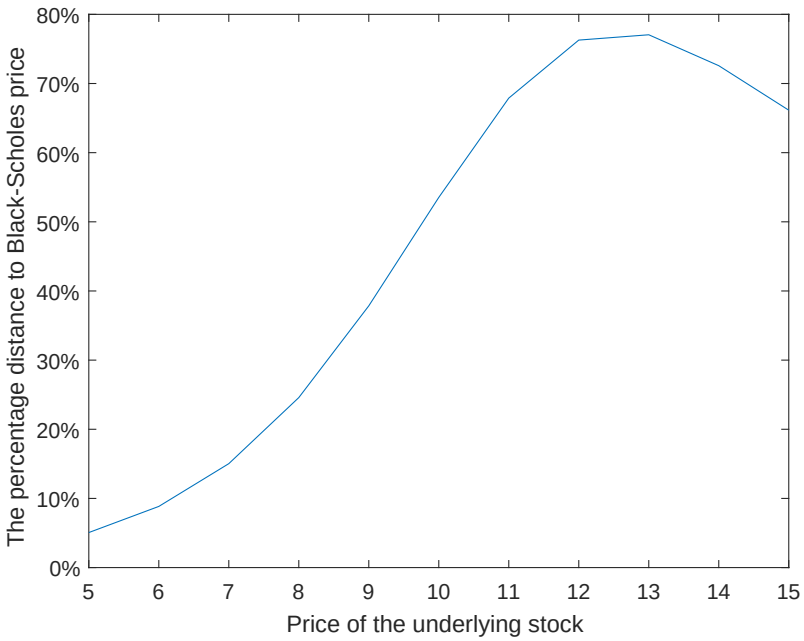
Proof. The proof is similar to Theorem 8.2.1. □

Again, when risk function is taken of $R(x) = x^+$, equal-risk price (8.3.24) for European put options is the same with that produced by Guo & Zhu (2017). With the same parameter as stated in (8.3.16), the comparisons between equal-risk price (8.3.25) and Black-Scholes price

for European put options are plotted in Figure 8.2(a) to demonstrate the effects of short selling bans on the European put option price. From Figure 8.2(a), the absolute difference between



(a) European put option price.



(b) The percentage distance

Figure 8.2: Comparisons between equal-risk price and Black-Scholes price for European put options.

two prices is significant when the underlying price is not very large. The percentage distance

of equal-risk price to Black-Scholes price is depicted in Figure 8.2(b), which indicates that the relative difference is significantly large even though the absolute difference is small for large underlying price. From both Figures 8.2(a) and 8.2(b), we come to a conclusion that equal-risk price of a European put option is higher than Black-Scholes price. In other words, the short selling bans have pumped up the European put option price substantially. Compared with the classic Black-Schole model, the buyer would pay more to purchase a European put option when short selling is banned. Such a premium compensates the seller of the European put option for he cannot short the underlying stock to hedge his risk due to the short selling bans.

As a brief summary of this section, we have analytically produced equal-risk price of European call and put options because the PDE systems (8.2.6) and (8.2.9) can be solved analytically, which is also consistent with the results of Guo & Zhu (2017). However, equal-risk price is still hard to produce analytically when the payoff function of contingent claim is not monotonic, such as the butterfly spread option. In a complete market, a butterfly spread option can be replicated by a linear combination of European call and put options. As a result, its price is actually also a linear combination of the price of the corresponding European call and put options. Guo & Zhu (2017) pointed out that such a replication method does not work any more due to the ban of short selling. To provide equal-risk price of the butterfly spread option, we have to apply a numerical scheme in the next section.

8.4 Numerical scheme for the PDE system

In this section, we provide a numerical scheme to solve the HJB equations (8.2.6) and (8.2.9). Mathematically, they are of the same type and we only take the former as an example when we demonstrate our numerical scheme. The other one can also be numerically solved similarly. In the following, the numerical discretization is implemented first and then two experiments are conducted accordingly.

8.4.1 Discretization

In order to solve the HJB equation (8.2.6) effectively, we introduce time reversal $\tau = T - t$ to change the terminal value problem to be a initial value problem as

$$\begin{cases} F^s_\tau = \inf_{\phi \geq 0} \{ \frac{1}{2} S^2 \sigma^2 F^s_{SS} + \phi S^2 \sigma^2 F^s_{Sv} + \frac{1}{2} S^2 \sigma^2 \phi^2 F^s_{vv} + rSF^s_S + rvF^s_v \}, \\ F^s(0, S, v) = R(Z(S) - v), (\tau, S, v) \in \Omega := [0, T] \times [0, \infty) \times \mathbf{R}. \end{cases} \tag{8.4.1}$$

To implement the numerical scheme, we truncate the unbounded domain into a finite one:

$$\bar{\Omega} = [0, T] \times [0, S_{\max}] \times [-v_{\max}, v_{\max}].$$

It is noted that there is only a terminal condition in the PDE system (8.4.1). In order to establish the properly-closed PDE system, some boundary conditions are needed. In this subsection, we focus on the numerical scheme and assume that some Dirichlet boundary conditions have been imposed properly. The details of how to impose these boundary conditions according to the financial reasoning are left in the next subsection. Of course, such a truncation would introduce some errors. As pointed out by Barles *et al* (1995), we can expect these errors incurred by imposing approximate boundary to be arbitrarily small by extending the computational domain.

The discretization is performed by placing a set of uniformly distributed grids in the computation domain $\bar{\Omega}$ as

$$\begin{aligned} S_i &= (i-1) \cdot \Delta S, i = 1, \cdots, N_1, \\ v_j &= (j-1) \cdot \Delta v, j = 1, \cdots, N_2, \\ \tau_l &= (l-1) \cdot \Delta \tau, l = 1, \cdots, M, \end{aligned}$$

where N_1, N_2 and M are the number of grids in the S,v and τ directions and the step sizes are correspondingly $\Delta S = \frac{S_{\max}}{N_1 - 1}$, $\Delta v = \frac{v_{\max}}{N_2 - 1}$, and $\Delta \tau = \frac{T}{M - 1}$. The value of the unknown function $F^s(\tau, S, v)$ at a grid point thus is thus denoted by $F^n_{i,j} = F^s(\tau_n, S_i, v_j)$.

We first adopt an explicit scheme to approximate the unknown function ϕ as follows:

$$\phi^n_{i,j} := \phi(\tau_n, S_i, v_j) = \max\{-\frac{\Delta v}{4\Delta S}\frac{F^n_{i+1,j+1} + F^n_{i-1,j-1} - F^n_{i+1,j-1} - F^n_{i-1,j+1}}{F^n_{i,j+1} - 2F^n_{i,j} + F^n_{i,j-1}}, 0\}, \tag{8.4.2}$$

and then apply an implicit scheme for the unknown function F

$$\frac{F^{n+1}_{i,j} - F^n_{i,j}}{\Delta\tau} = \mathcal{L}_3(\phi^n_{i,j})F^{n+1}_{i,j}, \tag{8.4.3}$$

where

$$\mathcal{L}_3(\phi)F = aF_{SS} + \rho F_{Sv} + bF_{vv} + cF_S + dF_v, \tag{8.4.4}$$

with $a = \frac{1}{2}\sigma^2S^2, b = \frac{1}{2}\phi^2\sigma^2S^2, \rho = \phi\sigma^2S^2, c = rS, d = rv$.

The alternative direction implicit (ADI) scheme is then applied to discretize the linear operator $\mathcal{L}_3$. In the first step, only the derivatives with respect to S are evaluated in terms of unknown values F^{2n+1}, while the other derivatives are replaced in terms of known values of F^{2n}. The difference equation obtained in the first step is implicit in the S-direction and explicit in v-direction. The procedure is then repeated at next step with the difference equation implicit in the v-direction and explicit in the S-direction. The cross derivative is always treated explicitly. Thus, we have two difference equations:

$$\begin{aligned}
\frac{F^{2n+1}_{i,j} - F^{2n}_{i,j}}{\Delta\tau} &= a_i\frac{F^{2n+1}_{i+1,j} - 2F^{2n+1}_{i,j} + F^{2n+1}_{i-1,j}}{\Delta S^2} + c_i\frac{F^{2n+1}_{i+1,j} - F^{2n+1}_{i-1,j}}{2\Delta S} \\
&\quad + b_{i,j}\frac{F^{2n}_{i,j+1} - 2F^{2n}_{i,j} + F^{2n}_{i,j-1}}{\Delta v^2} + d_j\frac{F^{2n}_{i,j+1} - F^{2n}_{i,j-1}}{2\Delta v} \\
&\quad + \rho_{i,j}\frac{F^{2n}_{i+1,j+1} - F^{2n}_{i-1,j+1} - F^{2n}_{i+1,j-1} + F^{2n}_{i-1,j-1}}{4\Delta S\Delta v},
\end{aligned} \tag{8.4.5}$$

$$\begin{aligned}
\frac{F^{2n+2}_{i,j} - F^{2n+1}_{i,j}}{\Delta\tau} &= b_{i,j}\frac{F^{2n+2}_{i,j+1} - 2F^{2n+2}_{i,j} + F^{2n+2}_{i,j-1}}{\Delta v^2} + d_j\frac{F^{2n+2}_{i,j+1} - F^{2n+2}_{i,j-1}}{2\Delta v} \\
&\quad + a_i\frac{F^{2n+1}_{i+1,j} - 2F^{2n+1}_{i,j} + F^{2n+1}_{i-1,j}}{\Delta S^2} + c_i\frac{F^{2n+1}_{i+1,j} - F^{2n+1}_{i-1,j}}{2\Delta S} \\
&\quad + \rho_{i,j}\frac{F^{2n+1}_{i+1,j+1} - F^{2n+1}_{i-1,j+1} - F^{2n+1}_{i+1,j-1} + F^{2n+1}_{i-1,j-1}}{4\Delta S\Delta v}.
\end{aligned} \tag{8.4.6}$$

The unknown functions $F^n_{i,j}$ and $\phi^n_{i,j}$ are both derived by solving these difference equations.

After solving the PDE system (8.2.6) and (8.2.9), the risk exposure of both the seller and buyer are produced numerically on the grids. To find out equal-risk price for the contingent claims, we have to find the root of equation (8.2.12) numerically, which is similar to determining the optimal exercise from the values of American put option through the free-boundary condition. We demonstrate how to produce equal-risk price numerically in the following.

Given a current underlying stock price S, it is assumed to be located between two grid points S_i and S_{i+1}. When the offer price v is larger than equal-risk price $v(S)$, the seller would take less risk for he gets more compensation, i.e

$$\rho^s(S, v; Z) < \rho^b(S, v; Z), \quad v > v(S). \tag{8.4.7}$$

On the other hand, when the offer price v is smaller than equal-risk price $v(S_i)$, the buyer takes less risk because he pays less, i.e.

$$\rho^s(S_i, v; Z) > \rho^b(S_i, v; Z), \quad v < v(S_i). \tag{8.4.8}$$

Consequently, equal-risk price of the claim Z with current price S_i is produced as

$$v(S_i) = \max_j \{v_j, j = 1, \cdots, N_2 | \quad \rho^s(S_i, v_j; Z) > \rho^b(S_i, v_j; Z)\}. \tag{8.4.9}$$

Similarly, equal-risk price of the claim Z with current price S_{i+1} is obtained as

$$v(S_{i+1}) = \max_j \{v_j, j = 1, \cdots, N_2 | \quad \rho^s(S_{i+1}, v_j; Z) > \rho^b(S_{i+1}, v_j; Z)\}. \tag{8.4.10}$$

As a result, equal-risk price of the contingent claim Z with current price S is

$$v(S) = \frac{v(S_i) + v(S_{i+1})}{2}. \tag{8.4.11}$$

8.4.2 Numerical experiments

In this subsection, two numerical experiments are conducted to illustrate the performance and convergence of our numerical scheme provided above. Both of them experiments were carried out with Matlab 2016a on an Intel(R) Xeon (R) CPU and risk function is assumed to be $R(x) = e^x - 1$.

Experiment 1: European call option

In the first experiment, the contingent claim is European call option, of which the value functions $F^s(t, S, v)$ and $F^b(t, S, v)$ have been obtained analytically according to Propositions 8.3.1 and 8.3.2. The analytical solutions are considered as the benchmark to illustrate the performance of our numerical scheme. Before implementing our numerical scheme, we need provide the proper boundary conditions for the PDE systems (8.2.6) and (8.2.9).

First of all, we consider the boundary condition on $S = 0$. The stock price stays at zero once it hits zero for it follows geometric Brownian motion. As a result, the European call option is worthless at the expire date. The seller of such a claim faces no liability; while the buyer gets nothing. In addition, the hedging strategies for both seller and buyer must be $\phi^* = 0$ because they could not invest on a stock whose price is zero. Therefore, the boundary conditions at $S = 0$ are

$$\begin{cases} F^s(t, 0, v) = R(-ve^{r(T-t)}), \\ F^b(t, 0, v) = R(ve^{r(T-t)}). \end{cases} \tag{8.4.12}$$

On the other hand, $S \to \infty$ implies $S_T \to \infty$, which indicates that the European call option is priceless. The buyer of such a claim would have an infinite income at the expire date. The boundary condition for the buyer at $S \to \infty$ is imposed as

$$\lim_{S\to\infty} F^b(t, S, v) = \lim_{S\to\infty} \inf_{\phi(\cdot)\in\Phi} \mathbf{E}_{\mathbb{Q}}^{S,v}[R(v_T^{v,\phi(\cdot)} - (S_T - K)^+)] = \lim_{S\to\infty} R(-S) = -1. \tag{8.4.13}$$

Such a bounded Dirichlet boundary condition is approximated by

$$F^b(t, S_{\max}, v) = -1. \tag{8.4.14}$$

As $S \to \infty$, we have

$$\lim_{S\to\infty} F^s(t, S, v) = \lim_{S\to\infty} \inf_{\phi(\cdot)\in\Phi} \mathbf{E}_{\mathbb{Q}}^{S,v}[R((S_T - K)^+ - v_T^{v,\phi(\cdot)})] = \infty. \tag{8.4.15}$$

When the value function approaches infinity on the boundary, we have to do growth order analysis so that such a boundary condition can be imposed on the truncated boundary. For any admissible hedging strategy ϕ, by applying the Jensen's inequality to the risk function $R(x)$, we have

$$\begin{aligned}\mathbf{E}_{\mathbb{Q}}^{S,v}[R(Z(S_T) - v_T^{v,\phi(\cdot)})] &\geq R(\mathbf{E}_{\mathbb{Q}}^{S,v}[Z(S_T) - v_T^{v,\phi(\cdot)}]) \\ &= R(e^{r(T-t)}(C^{BS}(S, K, r, \sigma, T-t) - v)).\end{aligned} \tag{8.4.16}$$

Consequently, the asymptotic behavior of the value function $F^s(t, S, v)$ is described as

$$\lim_{S\to\infty} F^s(t, S, v) \geq \lim_{S\to\infty} R(C^{BS}(S, K, r, \sigma, T-t)e^{r(T-t)} - ve^{r(T-t)}) \to \infty \quad \text{for } t \in [0, T], \tag{8.4.17}$$

which means that the growth order of $F^s(t, S, v)$ with respect to S is higher than that of the right hand for any t. On the other hand, at the specific time $t = T$, it follows that

$$\lim_{S\to\infty} F^s(T, S, v) = \lim_{S\to\infty} R((S - K)^+ - v), \tag{8.4.18}$$

which implies that the growth order of $F^(t, S, v)$ is the same as the right hand of the above equation at $t = T$. In order to make sure the boundary condition at $S \to \infty$ is consistent with the terminal condition at the corner point, the boundary condition on $S = S_{\max}$ is

$$F^s(t, S_{\max}, v) = R((S_{\max} - K)^+ - ve^{r(T-t)}). \tag{8.4.19}$$

Remark 8.4.1. When the value function is bounded as Equation (8.4.13), it can be directly imposed on the truncated boundary as Equation (8.4.14). When the value function approaches infinity on the boundary, such as Equation (8.4.17), we should do growth order analysis first and then impose an approximate boundary condition as Equation (8.4.19) to make sure that it is consistent with the terminal condition. In the rest of this chapter, such steps would be repeated. Without demonstrating details again, we would directly provide the truncated boundary conditions.

Following Lemma 8.2.1, the boundary conditions along the v direction are

$$\begin{cases} \lim\limits_{v\to\infty} F^b(t,S,v) = \infty, \\ \lim\limits_{v\to\infty} F^s(t,S,v) = -1, \\ \lim\limits_{v\to-\infty} F^b(t,S,v) = -1, \\ \lim\limits_{v\to-\infty} F^s(t,S,v) = \infty. \end{cases} \tag{8.4.20}$$

which are approximated by

$$\begin{cases} F^b(t,S,v_{\max}) = R(v_{\max}e^{r(T-t)} - (S-K)^+), \\ F^s(t,S,v_{\max}) = -1, \\ F^b(t,S,-v_{\max}) = -1, \\ F^s(t,S,-v_{\max}) = R((S-K)^+ + v_{\max}e^{r(T-t)}). \end{cases} \tag{8.4.21}$$

After providing these proper boundary conditions for the value functions $F^s(t,S,v)$ and $F^b(t,S,v)$, we now implement our numerical scheme. The parameters used in the this experiment are listed in Table 8.1.

Parameters	K	T	r	σ	$S_{\max}$	$v_{\max}$	v_0
values	5	0.5	0.05	0.3	10	5	2

Table 8.1: Parameters.

Given $\tau = T$ and $v = v_0$, the values of $F^s(\tau,S,v)$ and $F^b(t,S,v)$ are computed at different values of S and then listed in Tables 8.2 and 8.3. To determine the numerical rates of con-

vergence, we choose a sequence of meshes by successively halving the mesh parameters. The analytical solutions (8.3.1) and (8.3.5) obtained in Propositions 8.3.1 and 8.3.2 are considered as a benchmark when we report the l_2 error. The *ratio* column of Tables 8.2 and 8.3 is the ratio of successive l_2 error as the grid is refined by a factor of two.

(N_1, N_2, M)	$S = 4$	$S = 4.5$	$S = 5$	$S = 5.5$	$S = 6$	l_2 error	ratio
(21,21,160)	-0.8604	-0.8399	-0.7981	-0.7216	-0.5889	0.0452	
(41,41,320)	-0.8595	-0.8371	-0.7915	-0.7074	-0.5601	0.0123	3.7
(81,81,640)	-0.8593	-0.8364	-0.7898	-0.7039	-0.5528	0.0040	3.1
(161,161,1280)	-0.8591	-0.8361	-0.7891	-0.7026	-0.5503	0.0012	3.4
Benchmark (8.3.1)	-0.8592	-0.8362	-0.7892	-0.7023	-0.5492		

Table 8.2: The values of $F^s(T, S, v_0)$ with different meshes for European call options.

(N_1, N_2, M)	$S = 4$	$S = 4.5$	$S = 5$	$S = 5.5$	$S = 6$	l_2 error	ratio
(21,21,160)	6.3860	5.7689	4.8250	3.7099	2.6162	0.1635	
(41,41,320)	6.3423	5.6985	4.7540	3.6598	2.5889	0.0403	4.1
(81,81,640)	6.3307	5.6812	4.7369	3.6475	2.5819	0.0099	4.1
(161,161,1280)	6.3302	5.6791	4.7348	3.6465	2.5820	0.0071	1.4
Benchmark (8.3.5)	6.3268	5.6755	4.7313	3.6435	2.5800		

Table 8.3: The values of $F^b(T, S, v_0)$ with different meshes for European call options.

From Tables 8.2 and 8.3, it is observed that the successive l_2 error is approaching to zero as the grid spacing is diminished, which show that our numerical results are in good agreement with the benchmark solution. Therefore, we choose the numerical results calculated on the grid $(161, 161, 1280)$ to produce equal-risk price numerically.

In Figure 8.3, we demonstrate how risk exposure functions $F^s(T, S, v)$ and $F^b(T, S, v)$ changes as v varies with $S = 5$. As expected, the seller's risk exposure is increasing; while the buyer's risk exposure is decreasing as v goes toward infinity. Equal-risk price of European call options with current price $S = 5$ corresponds to the offer price v that makes $\rho^s(S, v; Z) = \rho^b(S, v; Z)$, which is numerically solved according to formula (8.4.11).

We repeat the above steps again and again with different values of S. Then how equal-risk price varies with the underlying stock price is plotted in Figure 8.4(a), compared with the results calculated from analytical pricing formula (8.3.14). The absolute error between them is plotted in Figure 8.4(b). From Figures 8.4(a) and 8.4(b), our numerical equal-risk price is in a

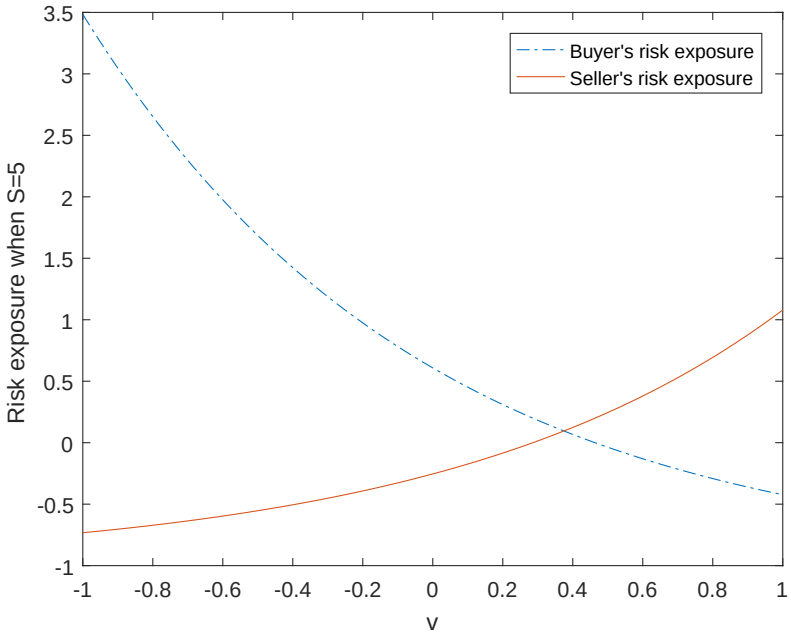
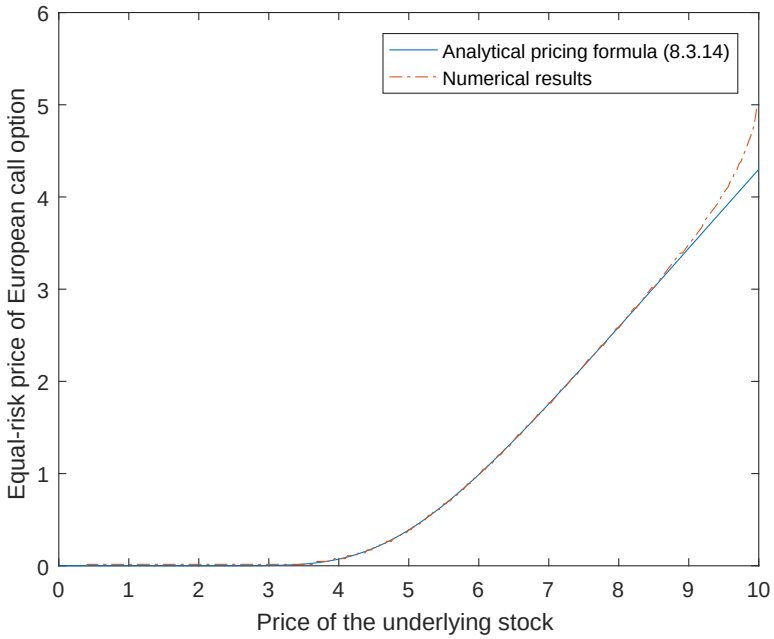


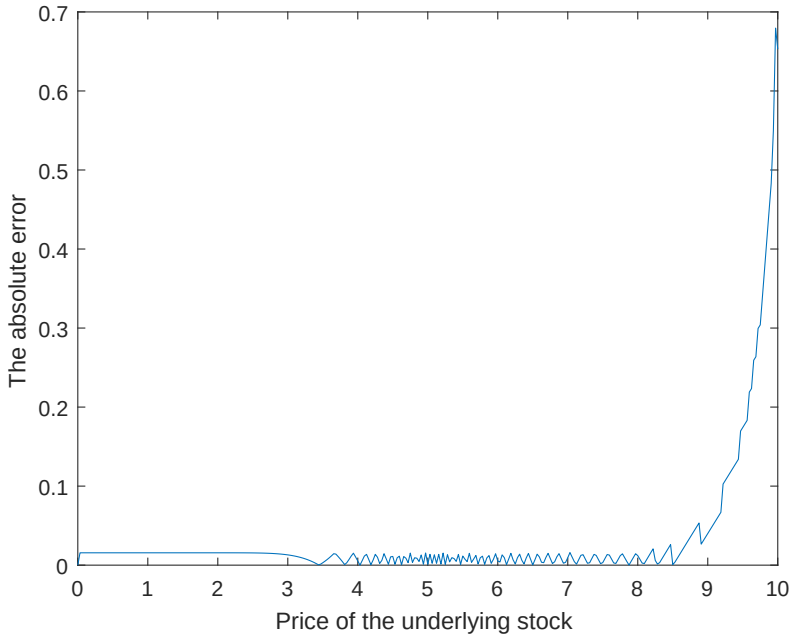
Figure 8.3: Risk exposure for both seller and buyer of European call option with $S = 5$.

good agreement with the analytical pricing formula except near the boundary $S = S_{\max}$. This error is actually incurred by truncating the domain and imposing an approximate boundary condition there. As pointed out by Barles *et al* (1995), by extending the computational domain, it is possible to make the near-field error arbitrarily small. The first experiment demonstrates that our method to produce equal-risk price by solving HJB equation numerically is consistent with analytical pricing formula, which provides us more confidence to apply it to the general contingent claim.

It has to been pointed out that oscillation can be observed in the absolute error in Figure 8.4(b). That's because the numerical scheme we applied in Section 8.4.2 is not the monotone scheme we proposed in Chapter 7. The reason is that it is hard to implement our monotone scheme when the PDE system is two dimensional, especially when the cross derivative term appears.



(a) Equal-risk price of European call options.



(b) The percentage distance

Figure 8.4: Comparisons between analytical pricing formula (8.3.14) and numerical results.

Experiment 2: Butterfly spread option

The second experiment we conduct is to derive equal-risk price for a butterfly spread option, of which the payoff function is defined as

$$Z(S) = (S - K_1)^+ - 2(S - \frac{K_1 + K_2}{2})^+ + (S - K_2)^+. \tag{8.4.22}$$

Figure 8.5 provides a diagram of such a payoff function. Obviously, it is not monotonic and Guo

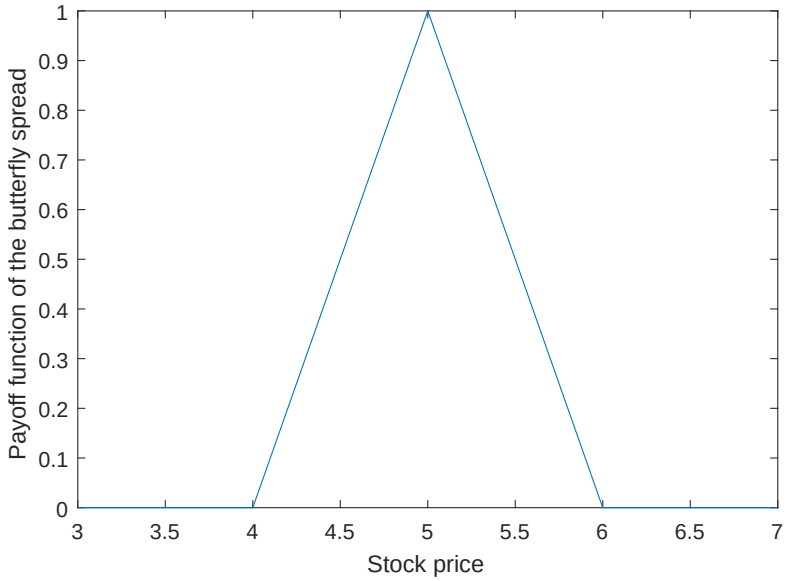


Figure 8.5: Payoff of a butterfly option with $K_1 = 4, K_2 = 6$.

& Zhu (2017) could not provide its corresponding equal-risk price according to their methods. We now apply our numerical scheme to solve the corresponding HJB equations first and then derive its equal-risk price numerically.

Before implementing the scheme, we also need to specify the boundary conditions according to the financial reasoning. Similar to the analysis in the first experiment, the stock price would stay at zero (or infinity) once it hits zeros (or infinity) at any time t because it follows geometric Brownian motion. According to the payoff function of the butterfly spread option, it becomes worthless at both $S = 0$ and $S \to \infty$. The seller of the claim faces no liability and he has no motivation to hedge. Consequently, he would invests his initial wealth on the risk-free account and obtains the profits $ve^{r(T-t)}$ at time T. Consequently, the boundary conditions at $S = 0$

and $S \to \infty$ are imposed as

$$\begin{cases} F^s(t,0,v) = R(-ve^{r(T-t)}), \\ \lim\limits_{S\to\infty} F^s(t,S,v) = R(-ve^{r(T-t)}). \end{cases} \tag{8.4.23}$$

According to the same financial reasoning, on the boundaries $S = 0$ and $S \to \infty$, the buyer pays v at time t for a worthless contingent claim and has no motivation to hedge. At the expire date, the buyer only faces a deterministic liability $ve^{r(T-t)}$ and we impose the boundary conditions as

$$\begin{cases} F^b(t,0,v) = R(ve^{r(T-t)}), \\ \lim\limits_{S\to\infty} F^b(t,S,v) = R(ve^{r(T-t)}). \end{cases} \tag{8.4.24}$$

The boundary condition along the v direction are also implied by Lemma 8.2.1, i.e

$$\begin{cases} \lim\limits_{v\to\infty} F^s(t,S,v) = -1, \\ \lim\limits_{v\to\infty} F^b(t,S,v) = \infty, \\ \lim\limits_{v\to-\infty} F^s(t,S,v) = \infty, \\ \lim\limits_{v\to-\infty} F^b(t,S,v) = -1, \end{cases} \tag{8.4.25}$$

which are approximated by

$$\begin{cases} F^s(t,S,v_{\max}) = -1, \\ F^b(t,S,v_{\max}) = R(v_{\max}e^{r(T-t)} - Z(S)), \\ F^s(t,S,-v_{\max}) = R(Z(S) + v_{\max}e^{r(T-t)}), \\ F^b(t,S,v_{\max}) = -1, \end{cases} \tag{8.4.26}$$

to make sure they are consistent with the terminal condition.

After all the boundary conditions are provided properly, we apply our numerical scheme to numerically solve the PDE system associated with the butterfly spread option. The parameters in the second experiment are listed in Table 8.4

When the contingent claim is a butterfly spread option, we do not have an analytical solution

Parameters	K_1	K_2	T	r	σ	$S_{\max}$	$v_{\max}$	v_0
values	4	6	0.5	0.05	0.3	10	3	1

Table 8.4: Parameters.

in hand and we choose the results computed on the uniform mesh with $321 \times 321 \times 2560$ nodes as a benchmark solution. The numerical results of the value functions $F^s(T, S, v_0)$ and $F^b(T, S, v_0)$ calculated on different meshes are reported in Tables 8.5 and 8.6.

(N_x, N_y, N_T)	$S = 4$	$S = 4.5$	$S = 5$	$S = 5.5$	$S = 6$	l_2 error	ratio
(11,11,40)	-0.5654	-0.4601	-0.3767	-0.4398	-0.4925	0.1183	
(21,21,80)	-0.5429	-0.4816	-0.4548	-0.4715	-0.5106	0.0294	4.0
(41,41,160)	-0.5445	-0.4919	-0.4696	-0.4832	-0.5172	0.0068	4.3
(81,81,320)	-0.5451	-0.4944	-0.4729	-0.4859	-0.5189	0.0015	4.7
(321, 321, 2560)	-0.5453	-0.4951	-0.4739	-0.4867	-0.5194		

Table 8.5: The values of $F^s(T, S, v_0)$ on different meshes for a butterfly spread option.

(N_x, N_y, N_T)	$S = 4$	$S = 4.5$	$S = 5$	$S = 5.5$	$S = 6$	l_2 error	ratio
(11,11,40)	-0.5480	-0.4491	-0.3658	-0.4112	-0.4385	0.1368	
(21,21,80)	-0.5427	-0.4807	-0.4508	-0.4568	-0.4669	0.0324	4.2
(41,41,160)	-0.5444	-0.4914	-0.4665	-0.4704	-0.4763	0.0071	4.5
(81,81,320)	-0.5451	-0.4939	-0.4701	-0.4736	-0.4786	0.0013	5.6
(321, 321, 2560)	-0.5452	-0.4946	-0.4710	-0.4742	-0.4786		

Table 8.6: The values of $F^b(T, S, v_0)$ on different meshes for a butterfly spread option.

The l_2 error reported in Tables 8.5 and 8.6 indicates the numerical results have converged and they can be used to produce equal-risk price for the butterfly spread option by solving Equation (8.2.12). Given $S = 5$, we plot the risk exposure of both the seller and buyer in Figure 8.6. The equal-risk price for the butterfly spread option with current price $S = 5$ should be the offer price that makes the risk exposure of seller and buyer equal, which can be numerically solved by formula (8.4.11).

When short selling is allowed and the market is complete, a butterfly spread option can be replicated by three European call options as shown in Equation (8.4.22). Its price is a linear combination of three call options

$$v = C^{BS}(S, K_1, r, \sigma, T) - 2C^{BS}(S, \frac{K_1 + K_2}{2}, r, \sigma, T) + C^{BS}(S, K_2, r, \sigma, T). \tag{8.4.27}$$

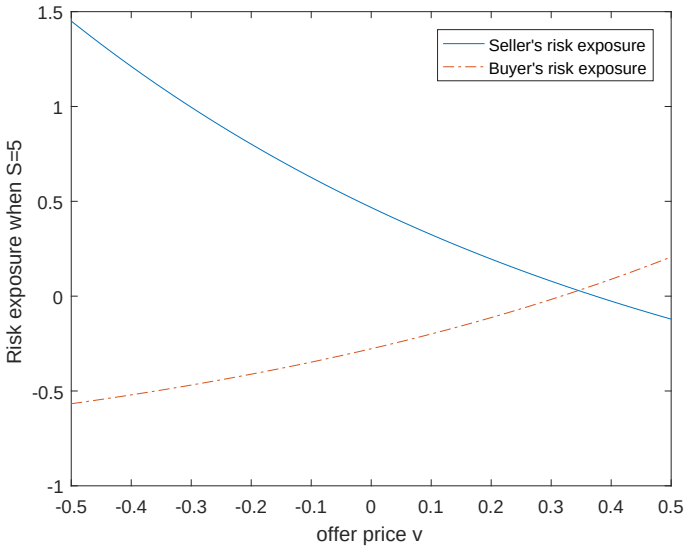
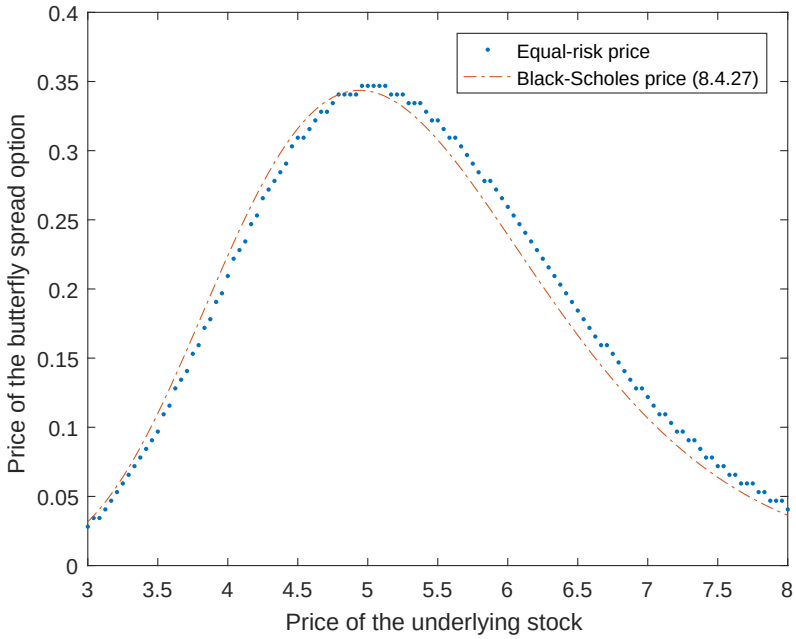


Figure 8.6: Risk exposure for both seller and buyer of a butterfly spread option with $S = 5$.

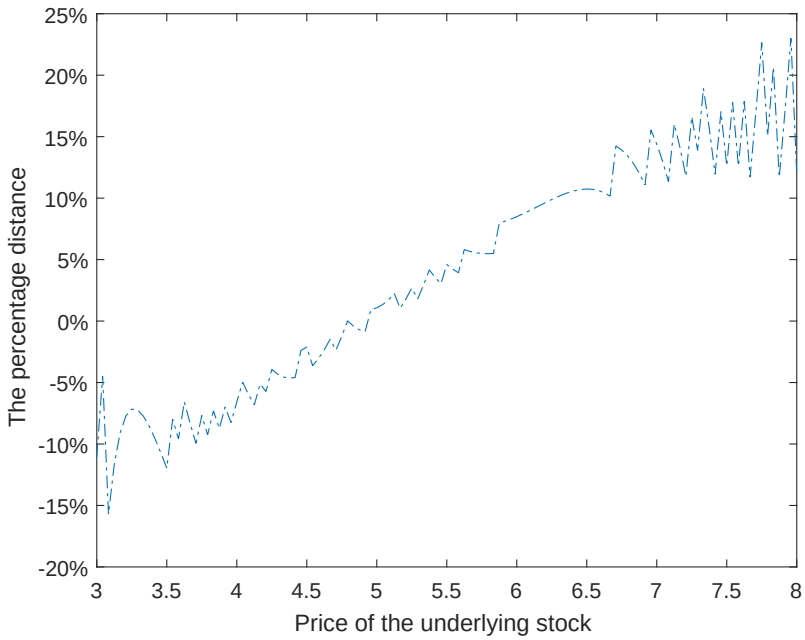
Such a Black-Scholes price is taken as the benchmark solution to illustrate how short selling bans affect the price of the butterfly spread option. Equal-risk price calculated through our PDE method and numerical results calculated from the formula (8.4.27) are plotted in Figure 8.7(a) and the percentage distance to Black-Scholes price is depicted in Figure 8.7(b).

Unlike the cases of European call (put) options where short selling decreases (increases) the option price for all the underlying stock prices, it is observed from Figure 8.7(a) that equal-risk price is higher than Black-Scholes price when $S > 5$; while it is lower than Black-Scholes price on the other side. Figure 8.7(b) shows that the relative difference between equal-risk price and Black-Scholes price is significant even though the absolute difference is tiny, which demonstrates that short selling bans indeed affect the price of the butterfly spread option. The effects of short selling bans depend on the current underlying stock price, which is totally different from the cases we considered before. To explain the effects of short selling bans, we come to its payoff function displayed in Figure 8.5. Locally, the payoff function is monotonically increasing with the underlying stock price when $S < 5$. In this region, the short selling bans push down the option price as it does in the case of European call option. On the other side, the short selling bans have an opposite effects.

Finally, we consider how the hedging strategy is affected by the ban of short selling. Take the



(a) European put option price.



(b) The percentage distance

Figure 8.7: Comparisons between equal-risk price and Black-Scholes price.

seller of this claim as an example. The optimal hedging strategy for the seller is numerically calculated from the PDE system (8.4.1). To make comparisons, the corresponding optimal hedging strategy in the Black-Scholes model without short selling bans is as

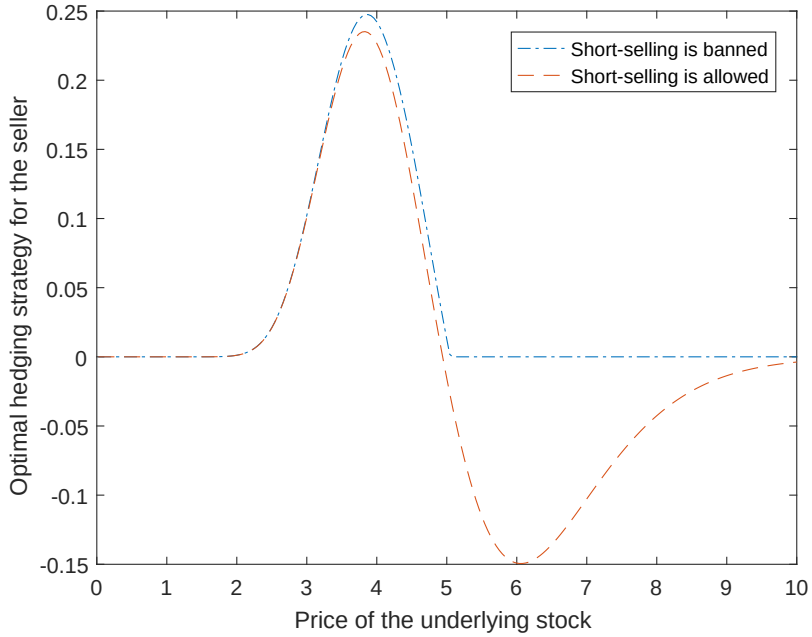
$$\phi^{BS} = \frac{\partial C^{BS}(S, K_1, r, \sigma, T)}{\partial S} - 2\frac{\partial C^{BS}(S, \frac{K_1+K_2}{2}, r, \sigma, T)}{\partial S} + \frac{\partial C^{BS}(S, K_2, r, \sigma, T)}{\partial S}. \tag{8.4.28}$$

The numerical results calculated from the PDE system and the formula (8.4.28) are plotted in Figure 8.8(a) with $v = 0.5$.

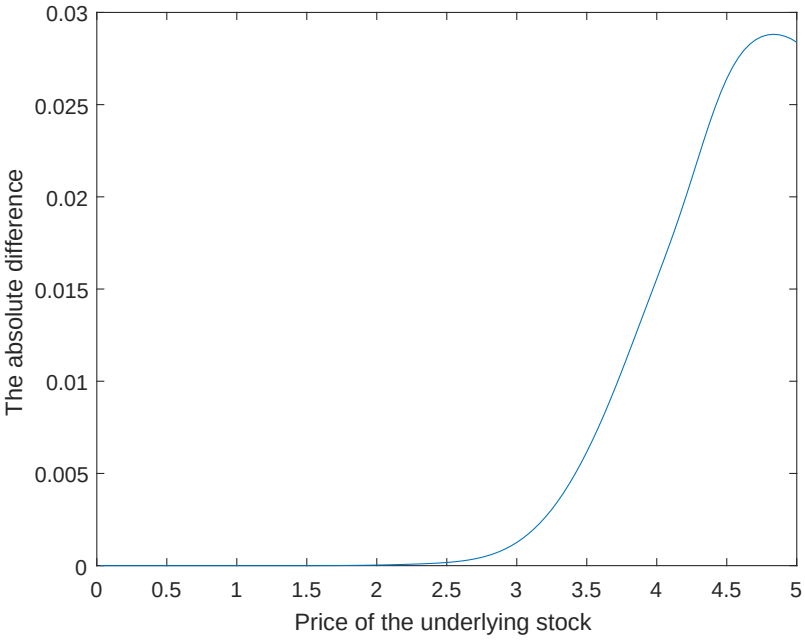
It is observed from Figure 8.8(a) that the optimal hedging strategy takes both positive and negative values as the underlying stock price varies when short selling is allowed. After imposing the short selling bans, the negative part becomes zero and the positive negative part becomes larger as the absolute difference between them is plotted in Figure 8.8(b) when $S < 5$.

8.5 Conclusions

This chapter has discussed how to price the contingent claim when short selling bans are imposed in the Black-Scholes model. We have successfully established and implemented a PDE framework of equal-risk pricing approach proposed by Guo & Zhu (2017). Under our PDE framework, analytical pricing formula has been derived for European call and put options, which indicates that our PDE approach degenerates to the results by Guo & Zhu (2017). In addition, our PDE approach has also been applied to deal with the case where the payoff function is non-monotonic, such as the butterfly spread option. According to the numerical results, the effects of short selling bans are discussed through comparisons between equal-risk price and Black-Scholes price. Generally, short selling bans would decrease the price of European call option; while it has an opposite effect on European put options. As for the butterfly spread option, short selling ban draws down the option price when payoff function is increasing with underlying stock price; while it pushes up the option price when the payoff function is decreasing.



(a) Optimal hedging strategy for the seller .



(b) The absolute difference

Figure 8.8: Comparison between the optimal hedging strategy for the seller of a butterfly spread option.

Chapter 9

Concluding Remarks

In this thesis, we have explored how regulations on short selling affect option pricing. In particular, our study is divided into two parts: the first one focuses on option with short selling restrictions and the second one is about option pricing with short selling bans.

In the case where short selling is restricted, a new dynamic model proposed by Avellaneda & Lipkin (2009) is adopted to describe the underlying stock. We provide a PDE approach to European option pricing problem under this new dynamic model. Then numerical results are obtained with two carefully chosen methods, the difference of which lies in the treatment of the jump term. In addition, a Monte Carlo scheme is also provided to implement the semi-explicit pricing formula so that we can compare the numerical results from our PDE system with those from the semi-explicit pricing formula. Through comparisons, it is verified that semi-explicit pricing formula is a good approximate solution when the independence is reasonable. However, in the event that this is not the case, it is our PDE approach that can solve the option problem correctly instead of the semi-explicit pricing formula. In other words, the derivation of the semi-explicit pricing formula requires an independence assumption, which has limited its application; while our PDE method always works whether the independence assumption is valid or not.

The PDE system established for European option pricing has also laid a solid foundation for our study on American option pricing. The PDE method is easily extended to the American case as a linear complimentary problem (LCP); while it is impossible for such an extension under

the semi-explicit pricing formula. Lagrange multiplier approach is applied to solve the LCP and then numerical results are provided to demonstrate how option prices and optimal exercise price are affected by the value of parameters in the hard-to-borrow stock model. A significant conclusion from our study is that American call option may be exercised early although the underlying stock does not pay dividend in the hard-to-borrow stock model. Such a conclusion supports a recent empirical study conducted by Jensen & Pedersen (2016) and overturns the classic result by Merton (1973). Explanations both from mathematical and financial aspects have been presented.

When short selling is completely banned, Guo & Zhu (2017) recently proposed a new and efficient *equal-risk pricing approach*. However, they only obtained analytical pricing formula for European call and put options as the payoff function is monotonic and it is difficulty to apply their analysis method to the case where the payoff function is non-monotonic. In order to expand the range of its application, we have successfully established a PDE framework by solving the maximization problem with the HJB equation. When the payoff function is monotonic, analytical pricing formula has also been derived from our PDE framework, which is consistent with the results from Guo & Zhu (2017). Furthermore, equal-risk price is also produced through solving the PDE system numerically when the payoff function is non-monotonic, such as a butterfly spread option, which is absent in Guo & Zhu (2017). Comparisons between equal-risk price and Black-Scholes price are provided to demonstrate how short selling bans affect the price of different options. In general, short selling bans would decrease the price of European call options; while it has an opposite effect on European put options. As for the butterfly spread option, short selling ban draws down the option price when payoff function is increasing with underlying stock price; while it pushes up the option price when the payoff function is decreasing.

Along with our research on short selling bans, we have also proposed three different solution approaches to the HJB equation, which has many applications even beyond mathematical finance. The homotopy analysis method has been successfully applied to decompose the highly nonlinear HJB equation into a series of linear PDEs that can be solved analytically. Besides, for

the first time, we derive a closed-form analytical solution to the HJB equation from the Merton problem defined on a finite horizon with exponential utility. In the literature where analytical solution is available, they have to make at least one of the following assumptions: (1) the utility function belongs to the constant relative risk aversion (CRRA) class; (2) the utility function is defined over terminal wealth only and consumption is not allowed; (3) the investment horizon is infinite. As one of contributions to the literature, our closed-form analytical is derived without any one of these three assumptions. A monotone numerical scheme is also presented to solve the HJB equation after demonstrating its stability, consistency and monotonicity. All these three approaches, albeit being quite different from each other, have shed some light on the final "exertion" on the very difficult problem of what we initially aimed at, i.e. being able to solve a highly nonlinear HJB equation in order to price options with a non-monotonic payoff function under the PDE framework of equal-risk pricing approach.

Appendix A

Proofs for Chapter 3

A.1 Proof of Proposition 3.3.1

Under the risk-neutral measure, the dynamics are shown in Equation (4.2.5). According to the definition of European call options, we have

$$u(x,S,t) = \mathbb{E}[e^{-r(T-t)}h(S_T)|S_t = S, x_t = x] = \mathbb{E}[e^{-r(T-t)}h(S_T)|\mathcal{G}_t]. \tag{A.1.1}$$

where $\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{H}_t$, $\mathcal{F}_t$ is the filtration generated by the standard Browain motion W_t and Z_t and $\mathcal{H}_t$ is the filtration generated by the Poisson process N_{λ_t}. Supposing that $0 \leq s \leq t \leq T$, using Equation (A.1.1), we obtain

$$\begin{aligned}
\mathbb{E}[e^{-rt}u(x_t,S_t,t)|\mathcal{G}_s] &= \mathbb{E}[\mathbb{E}[e^{-rT}h(S_T)|\mathcal{G}_t]\mathcal{G}_s] \\
&= \mathbb{E}[e^{-rT}h(S_T)|\mathcal{G}_s], \\
&= e^{-rs}\mathbb{E}[e^{-r(T-s)}h(S_T)|\mathcal{G}_s], \\
&= e^{-rs}u(x_s,S_s,s).
\end{aligned}$$

Therefore, $e^{-rt}u(x_t, S_t, t)$ is a martingale with respect to the filtration $\mathcal{G}_t$. By applying Ito's formula to $u(x_t, S_t, t)$, we have

$$
\begin{aligned}
&du(x_t, S_t, t) \\
&= u_t dt + u_S dS_t + u_x dx + \frac{1}{2}u_{SS}(dS_t)^2 + \frac{1}{2}u_{xx}(dx_t)^2 + [u(x - \beta\gamma, S(1-\gamma)) - u]dN_{\lambda_t} \\
&= \{u_t + rS_t u_S + [\alpha(\overline{x} - x_t) + \beta r]u_x + \frac{1}{2}\sigma^2 S_t^2 u_{SS} + \frac{\kappa^2 + \sigma^2\beta^2}{2}u_{xx} + \beta\sigma^2 S_t u_{xS}\}dt \\
&\quad + (u_S\sigma S + u_x\beta\sigma)dW_t + \kappa u_x dZ_t + [u(x - \beta\gamma, S_t(1-\gamma)) - u]dN_{\lambda_t} \\
&= \{rS_t u_S + [\alpha(\overline{x} - x_t) + \beta r]u_x + \frac{1}{2}\sigma^2 S_t^2 u_{SS} + \frac{\kappa^2 + \sigma^2\beta^2}{2}u_{xx} + \lambda_t[u(x - \beta\gamma, S_t(1-\gamma)) - u] \\
&+ \beta\sigma^2 S_t u_{xS} + u_t\}dt + (u_S\sigma S_t + u_x\beta\sigma)dW_t + \kappa u_x dZ_t + [u(x - \beta\gamma, S_t(1-\gamma)) - u](dN_{\lambda_t} - \lambda_t dt)
\end{aligned}
$$

where $u_t = \frac{\partial u(x,S,t)}{\partial t}$, $u_S = \frac{\partial u(x,S,t)}{\partial S}$,$u_x = \frac{\partial u(x,S,t)}{\partial x}$, $u_{xx} = \frac{\partial^2 u(x,S,t)}{\partial x^2}$, $u_{SS} = \frac{\partial^2 u(x,S,t)}{\partial S^2}$, and $u_{xS} = \frac{\partial^2 u(x,S,t)}{\partial x \partial S}$. Consequently, we obtain

$$
\begin{aligned}
&d(e^{-rt}u(x_t, S_t, t)) \\
&= e^{-rt}\{(u_S\sigma S + u_x\beta\sigma)dW_t + \kappa u_x dZ_t + [u(x - \beta\gamma, S(1-\gamma)) - u](dN_{\lambda_t} - \lambda_t dt)\} \\
&\quad + e^{-rt}\{u_t + rSu_S + [\alpha(\overline{x} - x_t) + \beta r]u_x + \frac{1}{2}\sigma^2 S^2 u_{SS} + \frac{\kappa^2 + \beta^2\sigma^2}{2}u_{xx} + \\
&\quad + \beta\sigma^2 S u_{xS} + \lambda_t[u(x - \beta\gamma, S(1-\gamma)) - u)] - ru\}dt.
\end{aligned}
$$

According to the martingale representation theorem, we set the dt term to be zero. Finally, we obtain

$$
\begin{cases}
-\dfrac{\partial u}{\partial t} = (\mathcal{L}_1 + \mathcal{L}_2)u, \\
u(x, S, T) = \max\{S - K, 0\},
\end{cases}
\tag{A.1.2}
$$

where

$$
\begin{cases}
\mathcal{L}_1 u = \dfrac{\kappa^2 + \beta^2\sigma^2}{2}u_{xx} + \dfrac{1}{2}\sigma^2 S^2 u_{SS} + \beta\sigma^2 S u_{xS} + [\alpha(\overline{x} - x) + \beta r]u_x + rSu_S - ru, \\
\mathcal{L}_2 u = e^x[u(x - \beta\gamma, S(1-\gamma), t) - u(x, S, t)].
\end{cases}
\tag{A.1.3}
$$

A.2 The derivation for Equation (3.4.10)

The space discretization is performed first. The governing PDE system (3.4.9) becomes:

$$\frac{\partial u_{i,j}^n}{\partial \tau} = a_i \delta_{xx} u_{i,j}^n + b_i \delta_{yy} u_{i,j}^n + c_i \delta_x u_{i,j}^n + d_i \delta_y u_{i,j}^n + \rho_i \delta_{xy} u_{i,j}^n - r u_{i,j}^n \tag{A.2.1}$$

where

$$\begin{aligned} (\delta_{xx} u)_{i,j} &= \tfrac{u_{i+1,j}-2u_{i,j}+u_{i-1,j}}{\Delta x^2} & (\delta_{yy} u)_{i,j} &= \tfrac{u_{i,j+1}-2u_{i,j}+u_{i,j-1}}{\Delta y^2} \\ (\delta_x u)_{i,j} &= \tfrac{u_{i+1,j}-u_{i-1,j}}{2\Delta x} & (\delta_y u)_{i,j} &= \tfrac{u_{i,j+1}-u_{i,j-1}}{2\Delta x} \end{aligned}$$

$$(\delta_{xy} u)_{i,j} = \frac{u_{i+1,j+1} + u_{i-1,j-1} - u_{i-1,j+1} - u_{i+1,j-1}}{4\Delta x \Delta y}.$$

The mixed derivative, the spatial derivatives in the x direction and the spatial derivatives in the y direction are denoted as linear operators A_0, A_1, A_2 respectively so that we can demonstrate the ADI method more clearly.

$$\begin{aligned} A_0 u_{i,j}^n &= \Delta\tau \rho_i \delta_{xy} u_{i,j}^n \\ A_1 u_{i,j}^n &= \Delta\tau (a_i \delta_{xx} u_{i,j}^n + c_i \delta_x u_{i,j}^n - \frac{r}{2} u_{i,j}^n) \\ A_2 u_{i,j}^n &= \Delta\tau (b_i \delta_{yy} u_{i,j}^n + d_i \delta_y u_{i,j}^n - \frac{r}{2} u_{i,j}^n) \end{aligned}$$

Thus, the weighted average of the fully implicit scheme and explicit scheme can be represented as :

$$[I - \theta(A_0 + A_1 + A_2)] u^{n+1} = [I + (1-\theta)(A_0 + A_1 + A_2)] u^n + \mathcal{O}((\Delta\tau)^3) \tag{A.2.2}$$

Note that when $\theta = 0$ or $\theta = 1$, (A.2.2) becomes fully explicit or fully implicit respectively. When $\theta = 0.5$, it is equivalent to apply Crank-Nicolson scheme to the time derivative $\dfrac{\partial u}{\partial \tau}$.

After some simple algebra, we obtain:

$$\begin{aligned} [I - \theta(A_0 + A_1 + A_2) + \theta^2 A_1 A_2] u^{n+1} &= [I + (1-\theta)A_0 + (1-\theta)A_1 + (1-\theta)A_2) + \theta^2 A_1 A_2] u^n \\ &\quad + \theta^2 A_1 A_2 (u^{n+1} - u^n) + \mathcal{O}((\Delta\tau)^3). \end{aligned} \tag{A.2.3}$$

As the term $\theta^2 A_1 A_2(u^{n+1} - u^n) \sim \mathcal{O}((\Delta\tau)^3)$, it can be taken into the error term.

Thus, (A.2.3) turns to

$$(I-\theta A_1)(I-\theta A_2)u^{n+1}-\theta A_0 u^{n+1} = [I+(1-\theta)A_0+(1-\theta)A_1+(1-\theta)A_2+\theta^2 A_1 A_2]u^n+\mathcal{O}((\Delta\tau)^3). \tag{A.2.4}$$

Moving the term $\theta A_0 u^{n+1}$ to the right side of the (A.2.4), we obtain :

$$\begin{aligned}(I-\theta A_1)(I-\theta A_2)u^{n+1} &= [I+A_0+(1-\theta)A_1+(1-\theta)A_2+\theta^2 A_1 A_2]u^n \\ &\quad +\theta A_0(u^{n+1}-u^n)+\mathcal{O}((\Delta\tau)^3) \\ &= [I+A_0+(1-\theta)A_1+(1-\theta)A_2+\theta^2 A_1 A_2]u^n+\mathcal{O}((\Delta\tau)^2).\end{aligned} \tag{A.2.5}$$

The finite difference equation for PDE system (A.2.1) is of the form :

$$(I-\theta A_1)(I-\theta A_2)u^{n+1} = [I+A_0+(1-\theta)A_1+(1-\theta)A_2+\theta^2 A_1 A_2]u^n. \tag{A.2.6}$$

A.3 The matrix forms in Methods 1 and 2

Both in Method 1 and Method 2, the boundary conditions are the same:

$$\begin{aligned}\mathbf{bnd1} &= u(x_1, y, \tau_n) &&= C^{BS}(e^y, K, r, n\Delta\tau, \sigma); \\ \mathbf{bnd2} &= u(x_{N_x}, y, \tau_n) &&= \max(e^y - K, 0); \\ \mathbf{bnd3} &= u(x, y_1, \tau_n) &&= 0; \\ \mathbf{bnd4} &= u(x, y_{N_y}, \tau_n) &&= \max(e^{y_N} - K, 0);\end{aligned}$$

It is pointed out that **bnd2** is an approximate Dirichlet boundary condition at $x \to \infty$. For the convenience of our implementation, we adopt such a condition at the beginning of each time step and then apply the Neumann boundary condition $\lim_{x\to\infty} \frac{\partial u}{\partial x}(x, y, \tau) = 0$ as $u(x_{N_x}, y, \tau) = u(x_{N_x-1}, y, \tau)$ to correct it at the end of each time step.

For Method 1, the matrix form are

$$\mathbf{H_1} u_j^{2n+1} = P_j^{2n} + \mathbf{xBnd}_j^1 \tag{A.3.1}$$

$$\mathbf{H_2} u_i^{2n+2} = Q_i^{2n+1} + \mathbf{yBnd}_i^1 \tag{A.3.2}$$

with

$$\mathbf{xBnd}_j^1 = \begin{pmatrix} \Delta\tau(\frac{a}{\Delta x^2} - \frac{c_2}{2\Delta x})\mathbf{bnd1}_j \\ 0 \\ \Delta\tau(\frac{a}{\Delta x^2} + \frac{c_{N_x-1}}{2\Delta x})\mathbf{bnd2}_j \end{pmatrix},$$

$$\mathbf{yBnd}_i^1 = \begin{pmatrix} \Delta\tau(\frac{b}{(\Delta y)^2} - \frac{d}{2\Delta y})\mathbf{bnd3}_i \\ 0 \\ \Delta\tau(\frac{b}{(\Delta x)^2} + \frac{d}{2\Delta y})\mathbf{bnd4}_i \end{pmatrix},$$

$$P_j^{2n} = (p_{1,j}^{2n}, p_{1,j}^{2n}, \cdots, p_{N_x,j}^{2n})^T,$$

$$Q_j^{2n+1} = (q_{i,1}^{2n+1}, q_{i,2}^{2n+1}, \cdots, q_{i,N_y}^{2n+1})^T,$$

$$p_{i,j} = u_{i,j} + \Delta\tau[\frac{b}{\Delta y^2}(u_{i,j+1} - 2u_{i,j} + u_{i,j-1}) + \frac{d}{2\Delta y}(u_{i,j+1} - u_{i,j-1}) - \frac{r}{2}u_{i,j}] + \Delta\tau[e^{x_i}(u_{\text{int}} - u_{i,j})],$$

$$q_{i,j} = u_{i,j} + \Delta\tau[\frac{a}{\Delta x^2}(u_{i+1,j} - 2u_{i,j} + u_{i-1,j}) + \frac{c_i}{2\Delta x}(u_{i+1,j} - u_{i-1,j}) - \frac{r}{2}u_{i,j}]\Delta\tau[e^{x_i}(u_{\text{int}}^n - u_{i,j})],$$

$$\mathbf{H_1} = \begin{pmatrix} 1 + \Delta\tau\left(\frac{2a}{\Delta x^2} + \frac{r}{2}\right) & -\Delta\tau\left(\frac{a}{\Delta x^2} + \frac{c_2}{2\Delta x}\right) & & & \\ -\Delta\tau\left(\frac{a}{\Delta x^2} - \frac{c_3}{2\Delta x}\right) & 1 + \Delta\tau\left(\frac{2a}{\Delta x^2} + \frac{r}{2}\right) & -\Delta\tau\left(\frac{a}{\Delta x^2} + \frac{c_3}{2\Delta x}\right) & & \\ & \ddots & \ddots & & \\ & -\Delta\tau\left(\frac{a}{\Delta x^2} - \frac{c_{N_x-2}}{2\Delta x}\right) & 1 + \Delta\tau\left(\frac{2a}{\Delta x^2} + \frac{r}{2}\right) & -\Delta\tau\left(\frac{a}{\Delta x^2} + \frac{c_{N_x-2}}{2\Delta x}\right) & \\ & & -\Delta\tau\left(\frac{a}{\Delta x^2} - \frac{c_{N_x-1}}{2\Delta x}\right) & 1 + \Delta\tau\left(\frac{2a}{\Delta x^2} + \frac{r}{2}\right) \end{pmatrix},$$

$$\mathbf{H_2} = \begin{pmatrix} 1+\Delta\tau\left(\frac{2b}{\Delta y^2}+\frac{r}{2}\right) & -\Delta\tau\left(\frac{b}{\Delta y^2}+\frac{d}{2\Delta y}\right) & & \\ -\Delta\tau\left(\frac{b}{\Delta y^2}-\frac{d}{2\Delta y}\right) & 1+\Delta\tau\left(\frac{2b}{\Delta y^2}+\frac{r}{2}\right) & -\Delta\tau\left(\frac{b}{\Delta y^2}+\frac{d}{2\Delta y}\right) & \\ & \ddots & \ddots & \\ & -\Delta\tau\left(\frac{b}{\Delta y^2}-\frac{d}{2\Delta y}\right) & 1+\Delta\tau\left(\frac{2b}{\Delta y^2}+\frac{r}{2}\right) & -\Delta\tau\left(\frac{b}{\Delta y^2}+\frac{d}{2\Delta y}\right) \\ & & -\Delta\tau\left(\frac{b}{\Delta y^2}-\frac{d}{2\Delta y}\right) & 1+\Delta\tau\left(\frac{2b}{\Delta y^2}+\frac{r}{2}\right) \end{pmatrix},$$

As for Method 2, we have the matrix form as :

$$\mathbf{A}Z_j = P_j + \mathbf{xBnd}_j^2 \tag{A.3.3}$$

$$\mathbf{B}u_i = Q_i + \mathbf{yBnd}_i^2 \tag{A.3.4}$$

with

$$\mathbf{xBnd}_j^2 = \begin{pmatrix} \theta\Delta\tau(\frac{a_2}{\Delta x^2}-\frac{c_2}{2\Delta x})\mathbf{bnd1}_j \\ 0 \\ \theta\Delta\tau(\frac{a_{N_x-1}}{\Delta x^2}+\frac{c_{N_x-1}}{2\Delta x})\mathbf{bnd2}_j \end{pmatrix},$$

$$\mathbf{xBnd}_i^2 = \begin{pmatrix} \Delta\tau(\frac{b_i}{(\Delta y)^2}-\frac{d_i}{2\Delta y})\mathbf{bnd3}_i \\ 0 \\ \Delta\tau(\frac{b_i}{(\Delta x)^2}+\frac{d_i}{2\Delta y})\mathbf{bnd4}_i \end{pmatrix},$$

$$P_j = (p_{1,j}, p_{1,j}, \cdots, p_{N_x,j})^T,$$

$$Q_j = (q_{i,1}, q_{i,2}, \cdots, q_{i,N_y})^T,$$

$$\begin{aligned} p_{i,j} &= u_{i,j} + \Delta\tau\rho_i\frac{u_{i+1,j+1}+u_{i-1,j-1}-u_{i+1,j-1}-u_{i-1,j+1}}{4\Delta x\Delta y} + \Delta\tau[\frac{b_i}{\Delta y^2}(u_{i,j+1}-2u_{i,j}+u_{i,j-1}) \\ &\quad +\frac{d_i}{2\Delta y}(u_{i,j+1}-u_{i,j-1}) - \frac{r}{2}u_{i,j}] + (1-\theta)\Delta\tau[\frac{a_i}{\Delta x^2}(u_{i+1,j}-2u_{i,j}+u_{i-1,j}) \\ &\quad +\frac{c_i}{2\Delta x}(u_{i+1,j}-u_{i-1,j}) - \frac{r}{2}u_{i,j}], \\ q_{i,j} &= z_{i,j} - \theta\Delta\tau[\frac{b_i}{\Delta x^2}(u_{i,j+1}-2u_{i,j}+u_{i,j-1}) + \frac{c_i}{2\Delta x}(u_{i+1,j}-u_{i-1,j}) - \frac{r}{2}u_{i,j}], \end{aligned}$$

$$\mathbf{A} = \begin{pmatrix} 1+\theta\Delta\tau(\frac{2a_2}{\Delta x^2}+\frac{r}{2}) & -\theta\Delta\tau(\frac{a_2}{\Delta x^2}+\frac{c_2}{2\Delta x}) & & & \\ -\theta\Delta\tau(\frac{a_3}{\Delta x^2}-\frac{c_3}{2\Delta x}) & 1+\theta\Delta\tau(\frac{2a_3}{\Delta x^2}+\frac{r}{2}) & -\theta\Delta\tau\Big(\frac{a_3}{\Delta x^2}+\frac{c_3}{2\Delta x}\Big) & & \\ & \ddots & \ddots & \ddots & \\ & -\theta\Delta\tau(\frac{a_{M-2}}{\Delta x^2}-\frac{c_{M-2}}{2\Delta x}) & 1+\theta\Delta\tau(\frac{2a_{M-2}}{\Delta x^2}+\frac{r}{2}) & -\theta\Delta\tau\Big(\frac{a_{M-2}}{\Delta x^2}+\frac{c_{M-2}}{2\Delta x}\Big) & \\ & & -\theta\Delta\tau(\frac{a_{M-1}}{\Delta x^2}-\frac{c_{M-1}}{2\Delta x}) & 1+\theta\Delta\tau(\frac{2a_{M-1}}{\Delta x^2}+\frac{r}{2}) & \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} 1+\theta\Delta\tau\Big(\frac{2b_i}{\Delta y^2}+\frac{r}{2}\Big) & -\theta\Delta\tau\Big(\frac{b_i}{\Delta y^2}+\frac{d_i}{2\Delta y}\Big) & & \\ -\theta\Delta\tau\Big(\frac{b_i}{\Delta y^2}-\frac{d_i}{2\Delta y}\Big) & 1+\theta\Delta\tau\Big(\frac{2b_i}{\Delta y^2}+\frac{r}{2}\Big) & -\theta\Delta\tau\Big(\frac{b_i}{\Delta y^2}+\frac{d_i}{2\Delta y}\Big) & \\ & \ddots & \ddots & \\ & -\theta\Delta\tau\Big(\frac{b_i}{\Delta y^2}-\frac{d_i}{2\Delta y}\Big) & 1+\theta\Delta\tau\Big(\frac{2b_i}{\Delta y^2}+\frac{r}{2}\Big) & -\theta\Delta\tau\Big(\frac{b_i}{\Delta y^2}+\frac{d_i}{2\Delta y}\Big) \\ & & -\theta\Big(\frac{a_j\Delta\tau}{\Delta x^2}-\frac{d_j'\Delta\tau}{2\Delta x}\Big) & 1+\theta\Delta\tau\Big(\frac{2b_i}{\Delta y^2}+\frac{r}{2}\Big) \end{pmatrix}.$$

A.4 Proof of Proposition 3.4.1

Following the standard procedure of von Neumann stability analysis (Strikwerda, 2004), $u^n_{k,m}$ in (3.4.11) and (3.4.12) is expressed by $g_1^n e^{ik\phi}e^{im\psi}$ and $Z_{k,m}$ by $g_1^n g_2 e^{ik\phi}e^{im\psi}$, where g_1 is the amplification factor of (3.4.11) and g_2 is the amplification factor of (3.4.12), with $\phi, \psi \in [-\pi, \pi]$. Therefore, Equations (3.4.11) and (3.4.12) are transformed to :

$$\begin{aligned} g_2(1-\theta z_1) &= 1 + z_0 + (1-\theta)z_1 + z_2 \\ g_1(1-\theta z_2) &= g_2 - \theta z_2 \end{aligned}$$

After simple calculations, we obtain :

$$g_1 = 1 + \frac{z_0 + z_1 + z_2}{(1-\theta z_2)(1-\theta z_2)},$$

where

$$\begin{aligned} z_0 &= -\frac{\rho\Delta\tau}{\Delta x\Delta y}\sin\phi\sin\psi \\ z_1 &= -\frac{4a\Delta\tau}{\Delta x^2}\sin^2\frac{\phi}{2} - \frac{r\Delta\tau}{2} + i\frac{c\Delta\tau}{\Delta x}\sin\phi \\ z_2 &= -\frac{4b\Delta\tau}{\Delta y^2}\sin^2\frac{\psi}{2} - \frac{r\Delta\tau}{2} + i\frac{d\Delta\tau}{\Delta y}\sin\psi \end{aligned}$$

It is easy to check that the coefficients satisfy

$$\begin{aligned} 4ab-\rho^2 &= [\kappa^2+\beta^2\sigma^2+\beta^2\gamma^2e^x][\sigma^2+\ln^2(1-\gamma)e^x] - \gamma^2\beta^2\ln^2(1-\gamma)e^{2x} \\ &\geq \gamma^2\beta^2\ln^2(1-\gamma)e^{2x} - \gamma^2\beta^2\ln^2(1-\gamma)e^{2x} \\ &= 0 \end{aligned}$$

$$\begin{aligned} 4\mathcal{R}(z_1)\mathcal{R}(z_2) - |z_0|^2 &\geq \tfrac{16ab\Delta\tau^2}{\Delta x^2\Delta y^2}\sin^2\tfrac{\phi}{2}\sin^2\tfrac{\psi}{2} - \tfrac{\Delta\tau^2\rho^2}{\Delta x^2\Delta y^2}\sin\phi\sin\psi \\ &= \tfrac{4\Delta\tau^2}{\Delta x^2\Delta y^2}\sin^2\tfrac{\phi}{2}\sin^2\tfrac{\psi}{2}(4ab-\rho^2\cos^2\tfrac{\phi}{2}\cos^2\tfrac{\psi}{2}) \\ &\geq \tfrac{4\Delta\tau^2}{\Delta x^2\Delta y^2}\sin^2\tfrac{\phi}{2}\sin^2\tfrac{\psi}{2}(4ab-\rho^2) \\ &\geq 0 \end{aligned}$$

Define $v_i = (\ \sqrt{-2\mathcal{R}(z_i)},\ \ \frac{|1+\theta z_i|}{\sqrt{2\theta}}\ )'$, where $i = 1, 2$. By Cauchy-Schwarz inequality, we have

$$\begin{aligned} \frac{|1-\theta z_1||1-\theta z_2|}{2\theta} &= \|v_1\|\|v_2\| \\ &\geq v_1\cdot v_2 \\ &= 2\sqrt{\mathcal{R}(z_1)\mathcal{R}(z_2)} + \tfrac{(1+\theta z_1)(1+\theta z_2)}{2\theta} \\ &\geq |z_0| + |\tfrac{(1-\theta z_1)(1-\theta z_2)}{2\theta} + z_1 + z_2| \end{aligned}$$

Dividing both sides with $|1-\theta z_1||1-\theta z_2|$, we have

$$\frac{1}{2\theta} \geq \frac{|z_0|}{|1-\theta z_1||1-\theta z_2|} + |\frac{1}{2\theta} + \frac{z_1+z_2}{|1-\theta z_1||1-\theta z_2|}| \geq |\frac{1}{2\theta} + \frac{z_1+z_2}{|1-\theta z_1||1-\theta z_2|}| \qquad \text{(A.4.1)}$$

Finally, we have such an estimate for the amplification factor g_1

$$\begin{aligned} |g_1| &= |1 + \tfrac{z_0+z_1+z_2}{(1-\theta z_1)(1-\theta z_2)}| \\ &= |1 - \tfrac{1}{2\theta} + \tfrac{1}{2\theta} + \tfrac{z_0+z_1+z_2}{(1-\theta z_1)(1-\theta z_2)}| \qquad (\theta \geq \tfrac{1}{2}) \\ &\leq 1 - \tfrac{1}{2\theta} + |\tfrac{1}{2\theta} + \tfrac{z_0+z_1+z_2}{(1-\theta z_1)(1-\theta z_2)}| \\ &\leq 1. \end{aligned}$$

Therefore, the scheme of DR method (3.4.11) and (3.4.12) is unconditionally stable.

Appendix B

Proofs for Chapter 6

B.1 Proof of Theorem 6.3.2

By introducing a transformation $d = \frac{1}{\gamma}$, we rewrite the functions $a_1(t)$ and $b_1(t)$ as

$$\begin{cases} a_1(t;d) = e^{-A_1(T-t)(d-1)}[\frac{1}{\eta}(1-d)(-d)^{\frac{1}{d-1}} + \frac{B_1}{A_1}(1-e^{A_1(T-t)})]^{d-1}, \\ b_1(t;d) = e^{-\int_t^T A_2(s)ds}[\alpha(\frac{1-d}{\eta})^d - \int_t^T B_2(s)e^{\int_s^T A_2(u)du}ds], \end{cases} \tag{B.1.1}$$

where

$$\begin{cases} A_1(t;d) = \frac{d\rho}{d-1} - \frac{r}{d-1} - \frac{dC}{(d-1)^2}, \\ B_1(t;d) = \frac{d-1}{\eta}(-d)^{\frac{1}{d-1}}, \\ A_2(t;d) = d\rho + \frac{dC}{1-d} + \frac{(d-1)^2}{\eta}(\frac{-a_1}{d})^{\frac{1}{1-d}}, \\ B_2(t;d) = \frac{\alpha}{\eta}\frac{(1-d)}{d}a_1. \end{cases} \tag{B.1.2}$$

According to L'Hospital rules, we obtain

$$\lim_{\gamma\to\infty} V(t,x;\gamma) = \lim_{d\to 0^-} V(t,x;d) = \lim_{d\to 0^-} -e^{-\rho t}e^{\frac{\ln[a_1(t;d)x+b_1(t;d)]}{d}} = -e^{-\rho t}e^{\lim\limits_{d\to 0^-}\frac{\dot{a}_1 x+\dot{b}_1}{a_1 x+b_1}}, \tag{B.1.3}$$

where $\dot{a}_1 = \frac{\partial a_1}{\partial d}$ and $\dot{b}_1 = \frac{\partial b_1}{\partial d}$.

After rearranging the terms, we have

$$a_1 = f_1 f_2 = f_1(f_3+f_4)^{d-1}, \tag{B.1.4}$$

with

$$\begin{cases} f_1 = e^{-A_1(T-t)(d-1)}, \\ f_2 = e^{(d-1)\ln(f_3+f_4)}, \\ f_3 = \frac{1}{\eta}(1-d)(-d)^{\frac{1}{d-1}}, \\ f_4 = \frac{1-e^{A_1(T-t)}}{A_1}\frac{d-1}{\eta}(-d)^{\frac{1}{d-1}}. \end{cases} \tag{B.1.5}$$

It is easy to check that $\lim_{d\to 0^-} f_1 = e^{r(T-t)}$ and

$$\lim_{d\to 0^-} f_2 = \lim_{d\to 0^-} e^{(d-1)\ln(f_3+f_4)} = 0.$$

Combining f_1 and f_2 together, we obtain

$$\lim_{d\to 0^-} a_1 = 0. \tag{B.1.6}$$

To calculate function $\dot{a}_1$, we first obtain

$$\lim_{d\to 0^-} \dot{f}_2 = \lim_{d\to 0^-} \frac{\ln(f_3+f_4)}{(f_3+f_4)^{1-d}} + (d-1)\frac{\dot{f}_3+\dot{f}_4}{(f_3+f_4)^{2-d}} = \lim_{d\to 0^-}(d-1)\frac{\dot{f}_3+\dot{f}_4}{(f_3+f_4)^{2-d}} = -\eta g(t)e^{-r(T-t)},$$

where $g(t) = \dfrac{r}{1+(r-1)e^{-r(T-t)}}$.

According to the chain rule, we come to

$$\bar{a}_1 := \lim_{d\to 0^-} \dot{a}_1 = \lim_{d\to 0^-} \dot{f}_1 f_2 + \dot{f}_2 f_1 = -\eta g(t). \tag{B.1.7}$$

Again, according to L'Hospital rules, we have

$$\lim_{d\to 0^-} A_2 = \lim_{d\to 0^-} \frac{(d-1)^2}{\eta} e^{\frac{1}{1-d}\ln(\frac{-a_1}{d})} = \lim_{d\to 0^-} \frac{1}{\eta} e^{\ln(-\dot{a}_1)} = g, \tag{B.1.8}$$

$$\lim_{d\to 0^-} B_2 = \lim_{d\to 0^-} \frac{\alpha}{\eta}\frac{(1-d)}{d} a_1 = \lim_{d\to 0^-} \frac{\alpha}{\eta}\dot{a}_1 = -\alpha g. \tag{B.1.9}$$

and

$$\lim_{d\to 0^-} (\frac{1-d}{\eta})^d = \lim_{d\to 0^-} e^{d\ln(\frac{1-d}{\eta})} = 1. \tag{B.1.10}$$

Using Equations (B.1.8),(B.1.9) and (B.1.10), we obtain

$$\begin{aligned}\lim_{d\to 0^-} b_1 &= \lim_{d\to 0^-} e^{-\int_t^T A_2 ds}(\alpha(\frac{1-d}{\eta})^d - \int_t^T B_2 e^{\int_s^T A_2 du} ds)\\ &= e^{-\int_t^T \lim_{d\to 0^-} A_2 ds}(\lim_{d\to 0^-} \alpha(\frac{1-d}{\eta})^d - \int_t^T \lim_{d\to 0^-} B_2 e^{\int_s^T \lim_{d\to 0^-} A_2 du} ds)\\ &= e^{-\int_t^T g ds}\alpha(1+\int_t^T g e^{\int_s^T g du} ds).\end{aligned}$$

The term $\lim_{d\to 0^-} \dot{b}_1$ is the last one we need to obtain now.

$$\begin{aligned}\bar{b}_1 := \lim_{d\to 0^-} \dot{b}_1 &= \lim_{d\to 0^-} e^{-\int_t^T A_2 ds}(-\int_t^T \dot{A}_2 ds)[\alpha(1-d)^d - \int_t^T B_2 e^{\int_s^T A_2 du} ds]\\ &\quad + e^{-\int_t^T A_2 ds}[\alpha(\frac{1-d}{\eta})^d(\ln\frac{1-d}{\eta} - \frac{d}{1-d})]\\ &\quad - e^{-\int_t^T A_2 ds}\int_t^T (\dot{B}_2 e^{\int_s^T A_2 du} + B_2 e^{\int_s^T A_2 du}\int_s^T \dot{A}_2 du) ds\\ &= -e^{-\int_t^T g ds}[\alpha\int_t^T \lim_{d\to 0^-} \dot{A}_2 ds(1+\int_t^T g e^{\int_s^T g du} ds) + \alpha\ln\eta\\ &\quad + \int_t^T e^{\int_s^T g du}(\lim_{d\to 0^-} \dot{B}_2 - \alpha g\int_s^T \lim_{d\to 0^-} \dot{A}_2 du) ds].\end{aligned} \tag{B.1.11}$$

It is pointed out that $\lim_{d\to 0^-} \dot{A}_2$ and $\lim_{d\to 0^-} \dot{B}_2$ are calculated as follows:

$$\lim_{d\to 0^-} \dot{B}_2 = \lim_{d\to 0^-} \frac{\partial}{\partial d}(\frac{\alpha}{\eta}\frac{1-d}{d}a_1) = \lim_{d\to 0^-} \frac{\alpha}{\eta}(\frac{\dot{a}_1 d - a_1}{d^2} - \dot{a}_1) = \lim_{d\to 0^-} \frac{\alpha}{\eta}(\frac{\ddot{a}_1}{2}) + \alpha g,$$

and

$$\begin{aligned}\lim_{d\to 0^-} \dot{A}_2 &= \lim_{d\to 0^-} \rho + \frac{C}{(1-d)^2} + 2\frac{(d-1)}{\eta}(\frac{-a_1}{d})^{\frac{1}{1-d}} + \frac{(d-1)^2}{\eta} I\\ &= \rho + C - \frac{2}{\eta}\lim_{d\to 0^-} e^{\frac{1}{1-d}\ln(\frac{-a_1}{d})} + \frac{1}{\eta}\lim_{d\to 0^-} I\\ &= \rho + C - 2g + \frac{1}{\eta}\lim_{d\to 0^-} I,\end{aligned}$$

where

$$\begin{aligned}\lim_{d\to 0^-} I &= \lim_{d\to 0^-} e^{\frac{1}{1-d}\ln(\frac{-a_1}{d})}[\frac{\dot{a}_1 d - a_1}{a_1 d(1-d)} + \frac{1}{(1-d)^2}\ln(\frac{-a_1}{d})] \\ &= \eta g \ln \eta g + \eta g \lim_{d\to 0^-} \frac{\ddot{a}_1}{\dot{a}_1(1-d) + \frac{a_1}{d}(1-2d)} \\ &= \eta g \ln \eta g - \frac{1}{2}\lim_{d\to 0^-} \ddot{a}_1.\end{aligned}$$

The term $\lim_{d\to 0^-} \ddot{a}_1$ is involved in both terms $\lim_{d\to 0^-} I$ and $\lim_{d\to 0^-} \dot{B}_2$. We can also obtain

$$\lim_{d\to 0^-} \dot{f}_1 = \lim_{d\to 0^-} e^{(1-d)A_1}[-A_1 + (1-d)\dot{A}_1](T-t) = e^{r(T-t)}(t-T)(\rho + C). \tag{B.1.12}$$

According to the chain rules, we have

$$\lim_{d\to 0^-} \ddot{a}_1 = \ddot{f}_1 f_2 + 2\dot{f}_1\dot{f}_2 + \ddot{f}_2 f_1 = 2\eta g(T-t)(\rho + C) + e^{r(T-t)}\ddot{f}_2. \tag{B.1.13}$$

Now we turn to the term $\lim_{d\to 0^-} \ddot{f}_2$.

$$\begin{aligned}\lim_{d\to 0^-} \ddot{f}_2 &= \lim_{d\to 0^-} \frac{\partial}{\partial d}\frac{(f_3+f_4)\ln(f_3+f_4) + (d-1)(\dot{f}_3+\dot{f}_4)}{(f_3+f_4)^{2-d}} \\ &= \lim_{d\to 0^-} \frac{(\dot{f}_3+\dot{f}_4)[2+\ln(f_3+f_4)] + (d-1)(\ddot{f}_3+\ddot{f}_4)}{(f_3+f_4)^{2-d}} \\ &\quad - \frac{[(f_3+f_4)\ln(f_3+f_4) + (d-1)(\dot{f}_3+\dot{f}_4)][-\ln(f_3+f_4) + (2-d)\frac{\dot{f}_3+\dot{f}_4}{f_3+f_4}]}{(f_3+f_4)^{2-d}} \\ &= \lim_{d\to 0^-} \frac{(d-1)(\dot{f}_3+\dot{f}_4)}{(f_3+f_4)^{2-d}}[2\ln(f_3+f_4) - 2\frac{\dot{f}_3+\dot{f}_4}{f_3+f_4} + \frac{\ddot{f}_3+\ddot{f}_4}{\dot{f}_3+\dot{f}_4} \\ &\quad + \frac{2}{d-1} + d\frac{\dot{f}_3+\dot{f}_4}{f_3+f_4} + \frac{(f_3+f_4)\ln^2(f_3+f_4)}{(d-1)(\dot{f}_3+\dot{f}_4)}].\end{aligned} \tag{B.1.14}$$

After some complicated calculations, we have

$$\begin{cases} \lim_{d\to 0^-} \frac{(d-1)(\dot{f}_3+\dot{f}_4)}{(f_3+f_4)^{2-d}} = -\eta g e^{-r(T-t)}, \\ \lim_{d\to 0^-} d\frac{\dot{f}_3+\dot{f}_4}{f_3+f_4} = -1, \\ \lim_{d\to 0^-} \frac{(f_3+f_4)\ln^2(f_3+f_4)}{(d-1)(\dot{f}_3+\dot{f}_4)} = 0. \end{cases} \tag{B.1.15}$$

The expression (B.1.14) can be simplified as

$$\lim_{d\to 0^-} \ddot{f}_2 = -\eta g e^{-r(T-t)} \lim_{d\to 0^-} [2\ln(f_3+f_4) - 2\frac{\dot{f}_3+\dot{f}_4}{f_3+f_4} + \frac{\ddot{f}_3+\ddot{f}_4}{\dot{f}_3+\dot{f}_4} - 3]. \tag{B.1.16}$$

Each term of Equation (B.1.16) is explored separately as

$$\begin{aligned}
&\lim_{d\to 0^-} \ln(f_3+f_4) \\
&= \lim_{d\to 0^-} \ln[(-d)^{\frac{1}{d-1}} \frac{1-d}{\eta}(1+\frac{e^{A_1}(T-t)-1}{A_1})] \\
&= r(T-t) - \ln \eta g + \lim_{d\to 0^-} \frac{\ln(-d)}{d-1}.
\end{aligned}$$

and

$$\lim_{d\to 0^-} \frac{\dot{f}_3+\dot{f}_4}{f_3+f_4} = \lim_{d\to 0^-} \frac{\partial}{\partial d} \ln(f_3+f_4) = \lim_{d\to 0^-} \frac{1}{d(d-1)} - \frac{\ln(-d)}{(d-1)^2} + f(t),$$

with $f(t) = g(t)\frac{r-\rho-C}{r}[T-t-\frac{1-e^{-r(T-t)}}{r}] - 1.$

With the help of such limits

$$\begin{cases} \lim_{d\to 0^-} \frac{\ddot{f}_3+\ddot{f}_4}{\dot{f}_3+\dot{f}_4} + \frac{2}{d(1-d)} = 3, \\ \lim_{d\to 0^-} [\frac{\ln(-d)}{d-1} + \frac{\ln(-d)}{(d-1)^2}] = 0, \end{cases} \tag{B.1.17}$$

we have

$$\lim_{d\to 0^-} 2\ln(f_3+f_4) - 2\frac{\dot{f}_3+\dot{f}_4}{f_3+f_4} + \frac{\ddot{f}_3+\ddot{f}_4}{\dot{f}_3+\dot{f}_4} = 2r(T-t) - 2\ln \eta g - 2f(t) + 3. \tag{B.1.18}$$

Finally, we obtain

$$\lim_{d\to 0^-} \ddot{f}_2 = 2\eta g(t) e^{-r(T-t)}[\ln \eta g(t) + f(t) - r(T-t)], \tag{B.1.19}$$

and from Equations (B.1.13) and (B.1.16), we have

$$\lim_{d\to 0^-} \ddot{a}_1 = 2\eta g(t) F(t), \tag{B.1.20}$$

with $F(t) = (T-t)(\rho + C - r) + \ln \eta g(t) + f(t)$.

As a result, expression (B.1.11) can be obtained explicitly as

$$\bar{b}_1(t) = -e^{-h_1(t)}\{\alpha[h_2(t)H(t) + \ln \eta] + \int_t^T e^{h_1(s)}[G(s) - \alpha g(s)H(s)]ds\}, \tag{B.1.21}$$

where $h_1(t), h_2(t), H(t)$ and $G(t)$ are defined in Equation (6.3.19).The unknown terms in (B.1.3) all have been obtained. The limiting form of the solution (6.3.17) is explicitly expressed as

$$\lim_{d \to 0^-} V(t, x; d) = -e^{-\rho t} e^{\frac{\bar{a}_1 x + \bar{b}_1}{\alpha}}. \tag{B.1.22}$$

B.2 Proof of Theorem 6.3.3.

To prove the solutions obtained with different methods are equivalent given $\alpha = 1$, we just need to check that

$$\bar{b}_1 = \ln b_2 - \ln \eta. \tag{B.2.1}$$

Although $\bar{b}_1$ and b_2 have been expressed explicitly in Theorems 6.3.2 and 6.3.3, they need to be simplified further to show the equivalence between solutions (6.3.17) and (6.3.20). We have

$$\begin{aligned} D(t) &= \rho + C - 2g + g(t)[\ln \eta g(t) - F] \\ &= \rho + C - g(t)[\frac{(C+\rho-r)(T-t)(r-1)e^{-r(T-t)}}{1 + (r-1)e^{-r(T-t)}} + \frac{C+\rho-r}{r}\frac{e^{r(T-t)} - 1}{e^{r(T-t)} + r - 1} + 1], \end{aligned}$$

and

$$\begin{aligned} H(t) &= \int_t^T D(s)ds = (\rho + C)(T-t) + \int_t^T g(s)ds - (C+\rho-r)(r-1)\Pi_1 - \frac{C+\rho-r}{r}\Pi_2 \\ &= (\rho + C - r)(T-t) + \ln g - (C+\rho-r)(r-1)\Pi_1 - \frac{C+\rho-r}{r}\Pi_2, \end{aligned} \tag{B.2.2}$$

with $\Pi_1 = \int_t^T \frac{(T-s)g(s)}{e^{r(T-s)}+r-1}ds$ and $\Pi_2 = \int_t^T \frac{(e^{r(T-s)}-1)g(s)}{e^{r(T-s)}+r-1}ds$. Furthermore, we obtain

$$\Pi_1 = \int_t^T \frac{(T-s)g(s)}{e^{r(T-s)}+r-1}ds = \frac{T}{1-r} + \frac{\ln g(t)}{r(r-1)} + \frac{g(t)}{r}[\frac{T}{r-1} + te^{-r(T-t)}], \tag{B.2.3}$$

and

$$\Pi_2 = \int_t^T \frac{(e^{r(T-s)}-1)g(s)}{e^{r(T-s)}+r-1}ds = \frac{1-g(t)}{r} + r(T-t) - \ln g(t) - \frac{g(t)-1}{r(r-1)}. \tag{B.2.4}$$

From Equations (B.2.2), (B.2.3) and (B.2.4), we have

$$H(t) = \int_t^T D(s)ds = \ln g(t) - \frac{C+\rho-r}{r}\frac{e^{r(T-t)-1}}{r-1+e^{r(T-t)}} + \frac{(T-t)(C+\rho-r)(r-1)}{r-1+e^{r(T-t)}}. \tag{B.2.5}$$

Furthermore, we obtain

$$G(t) = g(t)(F+1) = g(t)\ln \eta g(t) + \frac{(C+\rho-r)[(T-t)(r-1)r + e^{r(T-t)} - 1]}{e^{r(T-t)} + 2(r-1) + (r-1)^2 e^{-r(T-t)}}. \tag{B.2.6}$$

In addition, we explore

$$\begin{aligned}\int_t^T e^{r(T-s)}\frac{G(s)}{g(s)}ds &= \int_t^T e^{r(T-s)}\ln \eta g(s) + \frac{(C+\rho-r)[(T-s)(r-1)r + e^{r(T-s)} - 1]}{r[1+(r-1)e^{-r(T-s)}]}ds \\ &= (C+\rho-r)(r-1)\Pi_4 + \frac{C+\rho-r}{r}\Pi_5 + \int_t^T e^{r(T-s)}\ln g(s)\eta ds,\end{aligned} \tag{B.2.7}$$

with

$$\begin{cases} \Pi_4 = \int_t^T \frac{T-s}{1+(r-1)e^{-r(T-s)}}ds, \\ \Pi_5 = \int_t^T \frac{e^{r(T-s)}-1}{1+(r-1)e^{-r(T-s)}}ds = \frac{e^{r(T-t)}-1}{r} - r(T-t) + \ln g(t). \end{cases} \tag{B.2.8}$$

Now we have

$$\begin{aligned}
&\int_t^T e^{r(T-s)}H(s)ds \\
&= \int_t^T e^{r(T-s)}[(\rho+C-r)(T-s)+\ln g-(C+\rho-r)(r-1)\Pi_1-\frac{C+\rho-r}{r}\Pi_2]ds \\
&= \frac{C+\rho-r}{r}[\frac{e^{r(T-t)}-1}{r}-r(T-t)+\ln g]+\frac{(C+\rho-r)T(e^{r(T-t)}-1)}{r}+\int_t^T e^{r(T-s)}\ln g ds \\
&\quad -\frac{C+\rho-r}{r}[\int_t^T \frac{Tre^{r(T-s)}}{1+(r-1)e^{-r(T-s)}}ds+\int_t^T \frac{r(r-1)s}{1+(r-1)e^{-r(T-s)}}ds]. \qquad \text{(B.2.9)}
\end{aligned}$$

According to Equations (B.2.7) and (B.2.9), we obtain

$$\int_t^T e^{r(T-s)}[\frac{G(s)}{g(s)}-H(s)]ds=\frac{e^{r(T-t)}-1}{r}\ln\eta. \qquad \text{(B.2.10)}$$

The term $\bar{b}_1$ has been simplified as follows

$$\begin{aligned}
\bar{b}_1 &= -e^{-h_1(t)}\{(h_2(t)H(t)+\ln\eta)+\int_t^T e^{h_1(s)}[G(s)-g(s)H(s)]ds\} \\
&= -H(t)-g(t)e^{-r(T-t)}[\ln\eta+\int_t^T e^{r(T-s)}(\frac{G}{g}-H)ds] \\
&= -H(t)-g(t)\ln\eta\frac{1+(r-1)e^{-r(T-t)}}{r} \\
&= -H(t)-\ln\eta. \qquad \text{(B.2.11)}
\end{aligned}$$

On the other hand, we obtain

$$\begin{aligned}
&\ln b_2 \\
&= -g(t)e^{-r(T-t)}[(C+\rho)\int_t^T \frac{r-1+e^{r(T-s)}}{r}ds-\int_t^T e^{r(T-s)}ds+\int_t^T \ln g(s)e^{r(T-s)}ds] \\
&= -g(t)e^{-r(T-t)}[\frac{(r-1)(T-t)(\rho+C)}{r}+\frac{C+\rho-r}{r}\frac{e^{r(T-t)}-1}{r}+\int_t^T \ln g(s)e^{r(T-s)}ds] \\
&= -g(t)e^{-r(T-t)}[\frac{(r-1)(T-t)(\rho+C-r)}{r}+\frac{C+\rho-r}{r}\frac{e^{r(T-t)}-1}{r}]-\ln g(t). \qquad \text{(B.2.12)}
\end{aligned}$$

According to the Equations and (B.2.11) and (B.2.12), we finally have

$$\begin{aligned}
&\bar{b}_1 + \ln\eta - \ln b_2 \\
&= -\Big\{ \ln g(t) - \frac{C+\rho-r}{r}\frac{e^{r(T-t)-1}}{r-1+e^{r(T-t)}} + \frac{(T-t)(C+\rho-r)(r-1)}{r-1+e^{r(T-t)}} \Big\} \\
&\quad + g(t)e^{-r(T-t)}[\frac{(r-1)(T-t)(\rho+C-r)}{r} + \frac{C+\rho-r}{r}\frac{e^{r(T-t)}-1}{r}] + \ln g(t) \\
&= 0. \qquad \text{(B.2.13)}
\end{aligned}$$

This completes the proof.

Appendix C

Proofs for Chapter 8

C.1 The proof of Lemma 7.4.1

1. If $Z_1 \leq Z_2$, the following inequality always holds because of the monotonicity of $R(x)$ for any admissible hedging strategy $\phi(\cdot)$,

$$R(Z_1(S_T) - v_T^{v,\phi(\cdot)}) \leq R(Z_2(S_T) - v_T^{v,\phi(\cdot)}). \tag{C.1.1}$$

Taking expectation and infimum on both sides leads to

$$\rho^s(S, v; Z_1) \leq \rho^s(S, v; Z_2). \tag{C.1.2}$$

When $v_1 \leq v_2$, for any admissible hedging strategy $\phi(\cdot)$ the inequality becomes

$$R(Z(S_T) - v_T^{v_1,\phi(\cdot)}) \geq R(Z(S_T) - v_T^{v_2,\phi(\cdot)}), \tag{C.1.3}$$

which results in

$$\rho^s(S, v_1; Z) \geq \rho^s(S, v_2; Z). \tag{C.1.4}$$

By relation (8.2.11), the monotonicity of $\rho^b(S, v; Z)$ is characterized as

$$\rho^b(S, v; Z_1) = \rho^s(S, -v; -Z_1) \geq \rho^s(S, -v; -Z_2) = \rho^b(S, v; Z_2)$$
$$\rho^b(S, v_1; Z) = \rho^s(S, -v_1; -Z) \leq \rho^s(S, -v_2; -Z) = \rho^b(S, v_2; Z).$$

2. If the seller sets his hedging strategy to be zero, we obtain

$$\lim_{v\to\infty} \rho^s(S, v; Z) \leq \lim_{v\to\infty} \mathbf{E}_{\mathbb{Q}} R(Z(S_T) - v_T^{v,0}) = \lim_{v\to\infty} \mathbf{E}_{\mathbb{Q}} R(Z(S_T) - ve^{rT}) = LB. \tag{C.1.5}$$

Due to the fact that $R(Z(S_T) - v_T^{v,\phi(\cdot)}) \geq LB$ always holds for any $\phi(\cdot) \in \Phi$, we have

$$\lim_{v\to\infty} \rho^s(S, v; Z) \geq LB. \tag{C.1.6}$$

Combing Equations (C.1.5) and (C.1.6) together, we have $\lim_{v\to\infty} \rho^s(S, v; Z) = LB$.

For any $\phi(\cdot) \in \Phi$, we apply Jensen's inequality to risk function $R(x)$ and obtain

$$\mathbf{E}_{\mathbb{Q}}^{v,S}[R(v_T^{v,-\phi(\cdot)} - Z(S_T))] \geq R(ve^{rT} - \mathbf{E}_{\mathbb{Q}} Z(S_T)). \tag{C.1.7}$$

Taking infimum and limits on both sides results in

$$\lim_{v\to\infty} \rho^b(S, v; Z) = \lim_{v\to\infty} \inf_{\phi(\cdot)\in\mathbf{\Phi}} \mathbf{E}_{\mathbb{Q}}^{v,S}[R(v_T^{v,-\phi(\cdot)} - Z(S_T))] \geq \lim_{v\to\infty} R(ve^{rT} - \mathbf{E}_{\mathbb{Q}} Z(S_T)) = \infty.$$

Following the relation (8.2.11), it is easy to derive that

$$\lim_{v\to-\infty} \rho^s(S, v; Z) = \lim_{v\to-\infty} \rho^b(S, -v; -Z) = \lim_{v\to\infty} \rho^b(S, v; -Z) = \infty$$
$$\lim_{v\to-\infty} \rho^b(S, v; Z) = \lim_{v\to-\infty} \rho^s(S, -v; -Z) = \lim_{v\to\infty} \rho^s(S, v; -Z) = LB,$$

which completes the proof.

C.2 The proof of Theorem 8.2.1

Given the current underlying price S and the European contingent claim Z, we construct a map:

$$H(v) := \rho^b(S, v; Z) - \rho^s(S, v; Z). \tag{C.2.1}$$

According to Lemma 7.4.1, such a map $H(v)$ is continuous and non-decreasing. On one hand, we have

$$\lim_{v\to-\infty} H(v) = \lim_{v\to-\infty} [\rho^b(S, v; Z) - \rho^s(S, v; Z)] = -\infty. \tag{C.2.2}$$

On the other hand, as v tends toward infinity, we obtain

$$\lim_{v\to\infty} H(v) = \lim_{v\to\infty} [\rho^b(S, v; Z) - \rho^s(S, v; Z)] = \infty. \tag{C.2.3}$$

Hence, we conclude that there exists at least one solution to $H(v) = 0$ on $(-\infty, \infty)$.

To demonstrate the uniqueness of the solution, we first assume that the equation $H(v) = 0$ has two different solutions $v_1 > v_2$. According to the monotonicity described in Lemma 7.4.1, we have

$$\rho^b(S, v_1; Z) \geq \rho^b(S, v_2; Z) = \rho^s(S, v_2; Z) \geq \rho^s(S, v_1; Z) = \rho^b(S, v_1; Z), \tag{C.2.4}$$

which implies that $\rho^b(S, v_1; Z) = \rho^b(S, v_2; Z)$. Again, according to the monotonicity and convexity of $\rho^b(S, v; Z)$ with respect to v, we come to a conclusion that $\rho^b(S, v; Z)$ is constant for $v \leq v_1$. It follows that

$$\rho^s(S, v_1; Z) = \rho^b(S, v_2; Z) = \lim_{v\to-\infty} \rho^b(S, v; Z) = LB \tag{C.2.5}$$

By Jensen's inequality, we have

$$\begin{cases} R(v_1 e^{rT} - \mathbf{E}_{\mathbb{Q}} Z(S_T)) \leq \rho^s(S, v_1; Z) = LB \leq 0, \\ R(\mathbf{E}_{\mathbb{Q}} Z(S_T) - v_2 e^{rT}) \leq \rho^b(S, v_2; Z) = LB \leq 0. \end{cases} \tag{C.2.6}$$

The above equations implies that both $v_1 e^{rT} - \mathbf{E}_{\mathbb{Q}} Z(S_T)$ and $\mathbf{E}_{\mathbb{Q}} Z(S_T) - v_2 e^{rT}$ are non-positive because that $R(x) \geq 0$ for any $x \geq 0$. However, this conclusion contradicts the fact that

$$v_1 e^{rT} - \mathbf{E}_{\mathbb{Q}} Z(S_T) + \mathbf{E}_{\mathbb{Q}} Z(S_T) - v_2 e^{rT} = (v_1 - v_2) e^{rT} > 0. \tag{C.2.7}$$

Therefore, the solution must be unique.

Bibliography

Abbasbandy, S. & Zakaria, F. S. (2008), 'Soliton solutions for the fifth-order kdv equation with the homotopy analysis method', *Nonlinear Dynamics* **51**(1-2), 83–87.

Allegretto, W., Lin, Y. & Yang, H. (2001), 'A fast and highly accurate numerical method for the evaluation of american options', *Dynamics of Continuous Discrete and Impulsive Systems Series B* **8**, 127–138.

Amin, K. I. & Jarrow, R. A. (1992), 'Pricing options on risky assets in a stochastic interest rate economy', *Mathematical Finance* **2**(4), 217–237.

Asquith, P., Pathak, P. A. & Ritter, J. R. (2005), 'Short interest, institutional ownership, and stock returns', *Journal of Financial Economics* **78**(2), 243–276.

Avellaneda, M. & Lipkin, M. (2009), 'A dynamic model for hard-to-borrow stocks', *Risk* **22**(6), 92–97.

Avellaneda, M. & Zhang, S. (2010), 'Path-dependence of leveraged etf returns', *SIAM Journal on Financial Mathematics* **1**(1), 586–603.

Ayub, M., Rasheed, A. & Hayat, T. (2003), 'Exact flow of a third grade fluid past a porous plate using homotopy analysis method', *International Journal of Engineering Science* **41**(18), 2091–2103.

Bachelier, L. (1900), *Théorie de la spéculation*, Gauthier-Villars.

Barberis, N. (2000), 'Investing for the long run when returns are predictable', *The Journal of Finance* **55**(1), 225–264.

Barles, G. (1997), 'Convergence of numerical schemes for degenerate parabolic equations arising in finance theory', *Numerical methods in finance* **13**, 1.

Barles, G., Daher, C. & Romano, M. (1995), 'Convergence of numerical schemes for parabolic equations arising in finance theory', *Mathematical models and methods in applied Sciences* **5**(01), 125–143.

Barles, G. & Jakobsen, E. (2007), 'Error bounds for monotone approximation schemes for parabolic hamilton-jacobi-bellman equations', *Mathematics of Computation* **76**(260), 1861–1893.

Barles, G. & Jakobsen, E. R. (2002), 'On the convergence rate of approximation schemes for hamilton-jacobi-bellman equations', *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique* **36**(1), 33–54.

Barles, G. & Rouy, E. (1998), 'A strong comparison result for the bellman equation arising in stochastic exit time control problems and its applications', *Communications in Partial Differential Equations* **23**(11-12), 552–562.

Barles, G. & Souganidis, P. E. (1991), 'Convergence of approximation schemes for fully nonlinear second order equations', *Asymptotic analysis* **4**(3), 271–283.

Bates, D. S. (1996), 'Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options', *Review of financial studies* **9**(1), 69–107.

Beber, A. & Pagano, M. (2013), 'Short-selling bans around the world: Evidence from the 2007–09 crisis', *The Journal of Finance* **68**(1), 343–381.

Bellini, F. & Frittelli, M. (2002), 'On the existence of minimax martingale measures', *Mathematical Finance* **12**(1), 1–21.

Black, F. & Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of political economy* **81**(3), 637–654.

Bollerslev, T., Gibson, M. & Zhou, H. (2011), 'Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities', *Journal of econometrics* **160**(1), 235–245.

Brennan, M. J., Schwartz, E. S. & Lagnado, R. (1997), 'Strategic asset allocation', *Journal of Economic Dynamics and Control* **21**(8), 1377–1403.

Brennan, M. J. & Xia, Y. (2002), 'Dynamic asset allocation under inflation', *The journal of finance* **57**(3), 1201–1238.

Bris, A., Goetzmann, W. N. & Zhu, N. (2007), 'Efficiency and the bear: Short sales and markets around the world', *The Journal of Finance* **62**(3), 1029–1079.

Broadie, M. & Detemple, J. (1996), 'American option valuation: new bounds, approximations, and a comparison of existing methods', *The Review of Financial Studies* **9**(4), 1211–1250.

Brunnermeier, M. K. & Nagel, S. (2008), 'Do wealth fluctuations generate time-varying risk aversion? micro-evidence on individuals asset allocation (digest summary)', *American Economic Review* **98**(3), 713–736.

Carr, P. (1998), 'Randomization and the american put', *The Review of Financial Studies* **11**(3), 597–626.

Carr, P. & Faguet, D. (1996), Fast accurate valuation of american options., *in* 'JOURNAL OF FINANCE', Vol. 51, AMER FINANCE ASSN 44 WEST FOURTH ST, STE 9-190, NEW YORK, NY 10012, pp. 1030–1031.

Chacko, G. & Viceira, L. M. (2005), 'Dynamic consumption and portfolio choice with stochastic volatility in incomplete markets', *Review of Financial Studies* **18**(4), 1369–1402.

Cheridito, P., Soner, H. M., Touzi, N. & Victoir, N. (2007), 'Second-order backward stochastic differential equations and fully nonlinear parabolic pdes', *Communications on Pure and Applied Mathematics* **60**(7), 1081–1110.

Clarke, N. & Parrott, K. (1999), 'Multigrid for american option pricing with stochastic volatility', *Applied Mathematical Finance* **6**(3), 177–195.

Constantinides, G. M. & Zariphopoulou, T. (1999), 'Bounds on prices of contingent claims in an intertemporal economy with proportional transaction costs and general preferences', *Finance and Stochastics* **3**(3), 345–369.

Cox, J. (1975), 'Notes on option pricing i: Constant elasticity of variance diffusions', *Unpublished note, Stanford University, Graduate School of Business* .

Cox, J. C. & Huang, C.-f. (1989), 'Optimal consumption and portfolio policies when asset prices follow a diffusion process', *Journal of economic theory* **49**(1), 33–83.

Cox, J. C., Ingersoll Jr, J. E. & Ross, S. A. (1985), 'A theory of the term structure of interest rates', *Econometrica: Journal of the Econometric Society* pp. 385–407.

Cox, J. C., Ross, S. A. & Rubinstein, M. (1979), 'Option pricing: A simplified approach', *Journal of financial Economics* **7**(3), 229–263.

Crandall, M. G., Ishii, H. & Lions, P.-L. (1992), 'Users guide to viscosity solutions of second order partial differential equations', *Bulletin of the American Mathematical Society* **27**(1), 1–67.

Davis, M. H. (1997), 'Option pricing in incomplete markets', *Mathematics of derivative securities* **15**, 216–226.

Delbaen, F. & Schachermayer, W. (1994), 'A general version of the fundamental theorem of asset pricing', *Mathematische annalen* **300**(1), 463–520.

Detemple, J., Rindisbacher, M. et al. (2005), 'Closed-form solutions for optimal portfolio selection with stochastic interest rate and investment constraints', *Mathematical Finance* **15**(4), 539–568.

Detemple, J. & Sundaresan, S. (1999), 'Nontraded asset valuation with portfolio constraints: a binomial approach', *The review of financial studies* **12**(4), 835–872.

d'Halluin, Y., Forsyth, P. A. & Vetzal, K. R. (2005), 'Robust numerical methods for contingent claims under jump diffusion processes', *IMA Journal of Numerical Analysis* **25**(1), 87–112.

Diamond, D. W. & Verrecchia, R. E. (1987), 'Constraints on short-selling and asset price adjustment to private information', *Journal of Financial Economics* **18**(2), 277–311.

Douglas, J. & Rachford, H. H. (1956), 'On the numerical solution of heat conduction problems in two and three space variables', *Transactions of the American mathematical Society* pp. 421–439.

Dragulescu, A. A. & Yakovenko, V. M. (2002), 'Probability distribution of returns in the heston model with stochastic volatility*', *Quantitative finance* **2**(6), 443–453.

Duffie, D., Garleanu, N. & Pedersen, L. H. (2002), 'Securities lending, shorting, and pricing', *Journal of Financial Economics* **66**(2), 307–339.

Dupire, B. (1997), 'Pricing and hedging with smiles', *Mathematics of derivative securities* **1**(1), 103–111.

El Karoui, N. & Quenez, M.-C. (1995), 'Dynamic programming and pricing of contingent claims in an incomplete market', *SIAM journal on Control and Optimization* **33**(1), 29–66.

Evans, R. B., Geczy, C. C., Musto, D. K. & Reed, A. V. (2009), 'Failure is an option: Impediments to short selling and options prices', *Review of Financial Studies* **22**(5), 1955–1980.

Fahim, A., Touzi, N. & Warin, X. (2011), 'A probabilistic numerical method for fully nonlinear parabolic pdes', *The Annals of Applied Probability* pp. 1322–1364.

Figlewski, S. (1981), 'The informational effects of restrictions on short sales: Some empirical evidence', *Journal of Financial and Quantitative Analysis* **16**(4), 463–476.

Fleming, W. H. & Soner, H. M. (2006), *Controlled Markov processes and viscosity solutions*, Vol. 25, Springer Science & Business Media.

Follmer, H. & Schweizer, M. (1991), 'Hedging of contingent claims', *Applied stochastic analysis* **5**, 389.

Forsyth, P. A. & Labahn, G. (2007), 'Numerical methods for controlled hamilton-jacobi-bellman pdes in finance', *Journal of Computational Finance* **11**(2), 1.

Fouque, J.-P., Papanicolaou, G. & Sircar, K. R. (2000), *Derivatives in financial markets with stochastic volatility*, Cambridge University Press.

Fouque, J.-P., Sircar, R. & Zariphopoulou, T. (2015), 'Portfolio optimization and stochastic volatility asymptotics', *Mathematical Finance* .

Frittelli, M. (2000), 'The minimal entropy martingale measure and the valuation problem in incomplete markets', *Mathematical finance* **10**(1), 39–52.

Goll, T. & Rüschendorf, L. (2001), 'Minimax and minimal distance martingale measures and their relationship to portfolio optimization', *Finance and Stochastics* **5**(4), 557–581.

Gukhal, C. R. (2001), 'Analytical valuation of american options on jump-diffusion processes', *Mathematical Finance* **11**(1), 97–115.

Guo, I. & Zhu, S.-P. (2017), 'Equal risk pricing under convex trading constraints', *Journal of Economic Dynamics and Control* **76**, 136–151.

Harrison, J. M. & Kreps, D. M. (1978), 'Speculative investor behavior in a stock market with heterogeneous expectations', *The Quarterly Journal of Economics* **92**(2), 323–336.

Harrison, J. M. & Kreps, D. M. (1979), 'Martingales and arbitrage in multiperiod securities markets', *Journal of Economic theory* **20**(3), 381–408.

Harrison, J. M. & Pliska, S. R. (1981), 'Martingales and stochastic integrals in the theory of continuous trading', *Stochastic processes and their applications* **11**(3), 215–260.

Heath, D., Jarrow, R. & Morton, A. (1992), 'Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation', *Econometrica: Journal of the Econometric Society* pp. 77–105.

Henderson, V. (2005), 'Explicit solutions to an optimal portfolio choice problem with stochastic income', *Journal of Economic Dynamics and Control* **29**(7), 1237–1266.

Henderson, V. & Hobson, D. (2004), 'Utility indifference pricing-an overview', *Volume on Indifference Pricing* .

Heston, S. L. (1993), 'A closed-form solution for options with stochastic volatility with applications to bond and currency options', *Review of financial studies* **6**(2), 327–343.

Hintermüller, M., Ito, K. & Kunisch, K. (2002), 'The primal-dual active set strategy as a semismooth newton method', *SIAM Journal on Optimization* **13**(3), 865–888.

Hodges, S. D. (1989), 'Optimal replication of contingent claims under transaction costs', *Review of futures markets* **8**, 223–238.

Hon, Y.-C. & Mao, X.-Z. (1999), 'A radial basis function method for solving options pricing models', *Journal of Financial Engineering* **8**, 31–50.

Huang, J.-z., Subrahmanyam, M. G. & Yu, G. G. (1996), 'Pricing and hedging american options: a recursive integration method', *The Review of Financial Studies* **9**(1), 277–300.

Hugonnier, J., Kramkov, D. & Schachermayer, W. (2005), 'On utility-based pricing of contingent claims in incomplete markets', *Mathematical Finance* **15**(2), 203–212.

Hull, J. & White, A. (1990), 'Pricing interest-rate-derivative securities', *The Review of Financial Studies* **3**(4), 573–592.

Ikonen, S. & Toivanen, J. (2007), 'Componentwise splitting methods for pricing american options under stochastic volatility', *International Journal of Theoretical and Applied Finance* **10**(02), 331–361.

Ikonen, S. & Toivanen, J. (2008), 'Efficient numerical methods for pricing american options under stochastic volatility', *Numerical Methods for Partial Differential Equations* **24**(1), 104–126.

Ingersoll, J. E. (1987), *Theory of financial decision making*, Vol. 3, Rowman & Littlefield.

Ito, K. & Kunisch, K. (2006), 'Parabolic variational inequalities: The lagrange multiplier approach', *Journal de mathématiques pures et appliquées* **85**(3), 415–449.

Ito, K. & Toivanen, J. (2009), 'Lagrange multiplier approach with optimized finite difference stencils for pricing american options under stochastic volatility', *SIAM Journal on Scientific Computing* **31**(4), 2646–2664.

Jensen, M. V. & Pedersen, L. H. (2016), 'Early option exercise: Never say never', *Journal of Financial Economics* **121**(2), 278–299.

Jones, C. M. & Lamont, O. A. (2002), 'Short-sale constraints and stock returns', *Journal of Financial Economics* **66**(2), 207–239.

Karatzas, I. & Kou, S. G. (1996), 'On the pricing of contingent claims under constraints', *The annals of applied probability* pp. 321–369.

Karatzas, I., Lehoczky, J. P. & Shreve, S. E. (1987), 'Optimal portfolio and consumption decisions for a small investor on a finite horizon', *SIAM journal on control and optimization* **25**(6), 1557–1586.

Karatzas, I. & Shreve, S. E. (1998), *Methods of mathematical finance*, Vol. 39, Springer Science & Business Media.

Kim, T. S. & Omberg, E. (1996), 'Dynamic nonmyopic portfolio behavior', *Review of financial studies* **9**(1), 141–161.

Klemkosky, R. C. & Resnick, B. G. (1979), 'Put-call parity and market efficiency', *The Journal of Finance* **34**(5), 1141–1155.

Krylov, N. (2000), 'On the rate of convergence of finite-difference approximations for bellmans equations with variable coefficients', *Probability theory and related fields* **117**(1), 1–16.

Krylov, N. V. (2005), 'The rate of convergence of finite-difference approximations for bellman equations with lipschitz coefficients', *Applied Mathematics & Optimization* **52**(3), 365–399.

Kushner, H. & Dupuis, P. G. (2013), *Numerical methods for stochastic control problems in continuous time*, Vol. 24, Springer Science & Business Media.

Kushner, H. J. (1990), 'Numerical methods for stochastic control problems in continuous time', *SIAM Journal on Control and Optimization* **28**(5), 999–1048.

Landau, H. G. (1950), 'Heat conduction in a melting solid', *Quarterly of Applied Mathematics* **8**(1), 81–94.

Li, X., Lipkin, M. D. & Sowers, R. B. (2014), 'Dynamics of bankrupt stocks', *SIAM Journal on Financial Mathematics* **5**(1), 232–257.

Liao, S. (2003*a*), *Beyond perturbation: introduction to the homotopy analysis method*, CRC press.

Liao, S.-J. (2003*b*), 'On the analytic solution of magnetohydrodynamic flows of non-newtonian fluids over a stretching sheet', *Journal of Fluid Mechanics* **488**, 189–212.

Liu, H. (2004), 'Optimal consumption and investment with transaction costs and multiple risky assets', *The Journal of Finance* **59**(1), 289–338.

Liu, J. (2007), 'Portfolio selection in stochastic environments', *Review of Financial Studies* **20**(1), 1–39.

Liu, X., Yang, F. & Cai, Z. (2014), 'Does relative risk aversion vary with wealth? evidence from households' portfolio choice data', *Evidence from Households' Portfolio Choice Data (July 30, 2014)* .

Longstaff, F. A. & Schwartz, E. S. (2001), 'Valuing american options by simulation: a simple least-squares approach', *The review of financial studies* **14**(1), 113–147.

Ma, G. & Zhu, S.-P. (2017), A closed-form analytical soluiton to the *HJB* equaiton for the merton problem defined on a finite horizion with exponential utility function. Submitted for publication.

Ma, G., Zhu, S.-P. & Chen, W.-T. (2016), Pricing european call options under a hard-to-borrow stock model. Submitted for publication. Available at SSRN: https://ssrn.com/abstract=2975887.

Merton, R. C. (1969), 'Lifetime portfolio selection under uncertainty: The continuous-time case', *The review of Economics and Statistics* pp. 247–257.

Merton, R. C. (1971), 'Optimum consumption and portfolio rules in a continuous-time model', *Journal of economic theory* **3**(4), 373–413.

Merton, R. C. (1973), 'Theory of rational option pricing', *The Bell Journal of economics and management science* pp. 141–183.

Merton, R. C. (1976), 'Option pricing when underlying stock returns are discontinuous', *Journal of financial economics* **3**(1), 125–144.

Merton, R. C., Brennan, M. J. & Schwartz, E. S. (1977), 'The valuation of american put options', *The Journal of Finance* **32**(2), 449–462.

Miller, E. M. (1977), 'Risk, uncertainty, and divergence of opinion', *The Journal of Finance* **32**(4), 1151–1168.

Munk, C. (1999), 'The valuation of contingent claims under portfolio constraints: reservation buying and selling prices', *Review of Finance* **3**(3), 347–388.

Musiela, M. & Zariphopoulou, T. (2004), 'An example of indifference prices under exponential preferences', *Finance and Stochastics* **8**(2), 229–239.

Nielsen, L. T. (1989), 'Asset market equilibrium with short-selling', *The Review of Economic Studies* **56**(3), 467–473.

Nualart, E. (2009), 'Lectures on malliavin calculus and its applications to finance', *Lecture notes* .

Øksendal, B. (2003), *Stochastic differential equations*, Springer.

Ortega, J. M. & Rheinboldt, W. C. (1970), *Iterative solution of nonlinear equations in several variables*, Vol. 30, Siam.

Pooley, D. M., Forsyth, P. A. & Vetzal, K. R. (2003), 'Numerical convergence properties of option pricing pdes with uncertain volatility', *IMA Journal of Numerical Analysis* **23**(2), 241–267.

Pratt, J. W. (1964), 'Risk aversion in the small and in the large', *Econometrica: Journal of the Econometric Society* pp. 122–136.

Rouah, F. D. (2013), *The Heston Model and Its Extensions in Matlab and C*, John Wiley & Sons.

Rouge, R. & El Karoui, N. (2000), 'Pricing via utility maximization and entropy', *Mathematical Finance* **10**(2), 259–276.

Rubinstein, M. (1985), 'Nonparametric tests 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', *The Journal of Finance* **40**(2), 455–480.

Samuelson, P. A. (1969), 'Lifetime portfolio selection by dynamic stochastic programming', *The review of economics and statistics* pp. 239–246.

Schwartz, E. S. (1977), 'The valuation of warrants: Implementing a new approach', *Journal of Financial Economics* **4**(1), 79–93.

Sepp, A. (2003), 'Pricing european-style options under jump diffusion processes with stochastic volatility: Applications of fourier transform', *Acta et Commentationes Universitatis Tartuensis de Mathematica* **8**, 123–133.

Shreve, S. E. (2004), *Stochastic calculus for finance II: Continuous-time models*, Vol. 11, Springer Science & Business Media.

Stein, E. M. & Stein, J. C. (1991), 'Stock price distributions with stochastic volatility: an analytic approach', *The review of financial studies* **4**(4), 727–752.

Strikwerda, J. C. (2004), *Finite difference schemes and partial differential equations*, Siam.

Tankov, P. (2003), *Financial modelling with jump processes*, Vol. 2, CRC press.

Tehranchi, M. (2004), 'Explicit solutions of some utility maximization problems in incomplete markets', *Stochastic Processes and their Applications* **114**(1), 109–125.

Tepla, L. (2000), 'Optimal hedging and valuation of nontraded assets', *Review of Finance* **4**(3), 231–251.

Vasicek, O. (1977), 'An equilibrium characterization of the term structure', *Journal of financial economics* **5**(2), 177–188.

Wachter, J. A. (2002), 'Portfolio and consumption decisions under mean-reverting returns: An exact solution for complete markets', *Journal of financial and quantitative analysis* **37**(01), 63–91.

Wang, J. & Forsyth, P. A. (2008), 'Maximal use of central differencing for hamilton-jacobi-bellman pdes in finance', *SIAM Journal on Numerical Analysis* **46**(3), 1580–1601.

Wilmott, P. (2013), *Paul Wilmott on quantitative finance*, John Wiley & Sons.

Wilmott, P., Howison, S. & Dewynne, J. (1995), *The mathematics of financial derivatives: a student introduction*, Cambridge University Press.

Wu, L. & Kwok, Y.-K. (1997), 'A front-fixing finite difference method for the valuation of american options', *Journal of Financial Engineering* **6**(4), 83–97.

Xing, H. (2017), 'Stability of the exponential utility maximization problem with respect to preferences', *Mathematical Finance* **27**(1), 38–67.

Yong, J. & Zhou, X. Y. (1999), *Stochastic controls: Hamiltonian systems and HJB equations*, Vol. 43, Springer Science & Business Media.

Zariphopoulou, T. (1999), 'Optimal investment and consumption models with non-linear stock dynamics', *Mathematical Methods of Operations Research* **50**(2), 271–296.

Zeng, X. & Taksar, M. (2013), 'A stochastic volatility model and optimal portfolio selection', *Quantitative Finance* **13**(10), 1547–1558.

Zhao, J. & Wong, H. Y. (2012), 'A closed-form solution to american options under general diffusion processes', *Quantitative Finance* **12**(5), 725–737.

Zhu, S.-P. (2006), 'An exact and explicit solution for the valuation of american put options', *Quantitative Finance* **6**(3), 229–242.

Zhu, S.-P. & Chen, W.-T. (2011), 'A predictor–corrector scheme based on the adi method for pricing american puts with stochastic volatility', *Computers and Mathematics with Applications* **62**(1), 1–26.

Zhu, S.-P. & Ma, G. (2018), 'An analytical soluiton for the *HJB* equaiton arising from the merton problem', *International Journal of Financial Engineering* **5**(01), 1850008.

Zvan, R., Forsyth, P. & Vetzal, K. (1998), 'Penalty methods for american options with stochastic volatility', *Journal of Computational and Applied Mathematics* **91**(2), 199–218.

Publication list of the author

Ma, G., Zhu, S.-P. & Chen, W.-T. (2016), 'Pricing European call options under a hard-to-borrow stock model', *Applied Mathematics and Computation*, (Submitted).

Ma, G. & Zhu, S.-P. (2018), 'Pricing American call options under a hard-to-borrow model', *European Journal of Applied Mathematics* **29**(03), 494–514.

Zhu, S.-P. & Ma, G. (2018), 'An analytical soluiton for the *HJB* equaiton arising from the merton problem', *International Journal of Financial Engineering* **5**(01), 1850008.

Ma, G. & Zhu, S.-P. (2017), 'A closed-form analytical solution to the Merton problem defined on a finite horizon with exponential utility', *Applied Mathematical Finance*, (Submitted).

Ma, G., Kang, B.& Zhu, S.-P. (2017), 'A monotone numerical scheme for the HJB equation arising from the Merton problem', *Computers and Mathematics with Applications*, (Submitted).

Ma, G., Zhu, S.-P.& Guo, I. (2017), 'Option pricing with short selling bans being imposed', *SIAM Journal on Financial Mathematics*, (Submitted).