Quantitative Strategies for Derivatives Trading

©Dennis Yang
Independent Consultant
dennis.yang@atmif.com
July 7, 2006

This document is available at www.atmif.com/qsdt.

Contents

Preface

This is a book on option pricing and trading. There are currently several dozen books on the subject of options theories and their applications available on the market. Some are elementary; some are abstract; some are for day traders; some are for academics. Where does this book belong in the spectrum?

This is not an introductory book on options, nor is it a get-rich-quick options technical analysis book for day traders. The closest words I can think of to describe this book are that it is a research monograph on options theories, which contains the results of my research for the past few years. The target audiences for this book are quantitatively oriented options traders, derivative quants, as well as academic researchers.

Almost all option books currently on the market deal with only the pricing problem; they are silent on the issue of what to do when the model and market prices do not agree. Here is what sets this book apart: It offers a *systematic* way of determining the optimal trading size under a given option market price. Despite the importance of the subject to practitioners, I believe that this is the *first* book that offers a solid foundation to address the question of how to trade derivatives.

Let me now describe the mathematical background needed for this book. I assume that you know the basics of the classical Black Scholes option theory, *e.g.*, concepts such as implied volatility, delta, gamma; and also that you understand the classical derivation of the Black Scholes equation using the delta hedging argument, which implies that you know Ito's lemma of stochastic calculus and the concept of partial differential equations (PDE). This book is written entirely from the PDE point of view, so no knowledge of probability measure theory or martingale theory is needed. On the other hand, it will be helpful, but not required, if you know some basics of the stochastic control theory. Nevertheless, even if you are totally new to this subject, you can understand and enjoy this book.

I try to avoid the expected scientific dullness associated with most research monographs by using a more personal and intuitive narrative style of writing. In particular, I have avoided writing from the perspective of a third party "representative agent"; instead I have used the "you" viewpoint throughout this book. I hope by explicitly stressing *you*, the reader, I can get my messages across more effectively.

Because I am not a mathematician, this book contains no formal mathematical propositions, theorems or lemmas, let alone their proofs. However, that does not mean that this book contains no new ideas; on the contrary, unlike most option books, this book is mostly made of new results. In fact the basic idea is so simple and elegant, you might, after finishing this book, say to yourself: "Aha! Why didn't I think of this before?"

I hope you will enjoy reading this book.

Acknowledgment

I am grateful for the following people, who have read the manuscript, provided helpful comments and pointed out typos. They are: Peter Leopold, Larry Lu, Frank Wang, Minjie Yu, Guihua Zhang and Qiang Zhang. I also thank Peter Carr for providing me with some references.

About the Author

Dennis Yang has a Ph.D. in physics. He was an equity option market maker on the American Stock Exchange for a number of years. Currently he is an independent consultant, who can be reached at *dennis.yang@atmif.com*.

Part I

BACKGROUND

Chapter 1

Introduction

1.1 Trading Strategy

This book is written from the perspective of an options[1] trader. While reading this book, it would be helpful for you to pretend to be a trader even if you are not one in real life, because the trader mentality will keep you focused on the bottom line, and avoid the unnecessary distractions of academic jargon.

The most important tool a trader uses to make money is undoubtedly his trading strategies. A trading strategy, may mean different things to different people. In my mind, a systematic trading strategy should consist of two parts, namely the direction and the quantity. At any given point in time, for any given security, a trading strategy should first tell a trader whether to buy, sell or hold; it then should provide the optimal trading size under the given market price.

The first part of the decision relies on finding a fair value, *i.e.*, derivative pricing, a subject with thousands of papers written on it since the seminal work of Black and Scholes (BS). In other words, the direction of a trade is based on the following rule: if the market price of an option is lower/higher than your model's fair value, or model price, then you will buy/sell the option. If you do not agree with this rule, then you have to wonder why you need a model at all. When the market price is precisely the model's fair value, you hold your position, which means you are in equilibrium with the market. For this reason, fair values are also called equilibrium prices; the two terms are used interchangeably in this book.

Contrary to the option pricing problem, there is scant work on how to make the second part of a trader's decision, *i.e.*, a quantitative theory of determining the optimal size to trade when the market price *disagrees* with the model's fair value. According to the arbitrage argument of the BS option theory, the trading size should be infinite! But for all real-life traders, such an answer is obviously wrong. The optimal trading size problem, for which the BS option theory is unable to provide a reasonable answer, is precisely what this book addresses. This fills an important void in the vast literature on derivatives.

The ultimate goal is to build a model-based automatic derivatives trading system. For any given market price, the system should output just one number—the optimal trading size. A positive number means buy; a negative number means sell; zero means hold. Simply put, the trading strategy boils down to the size problem.

[1] Options mean financial derivatives or contingent claims in this book.

1.2 Position-dependent Valuation

In order to obtain concrete quantitative results, you need to use a model. Before setting up a model, I would like to see what kind of qualitative issues that the model needs to address. For that, I need to find out behaviors of real-life options traders. I mostly focus on equity options in this book, since this is the area where I am most familiar. The applicability of the theory to other types of options will be self-evident later in this book.

When trading stocks, you are mostly trading directly with other investors. Such is not the case for options. You will almost surely be trading with option market makers, or dealers, who set the current market price for an option. Therefore familiarity with market makers' actions will be helpful in understanding the movement of option market prices.

Let me make a quick digression to answer two obvious questions before proceeding. First, who are option market makers? The answer is that they are options speculators who happen to lease or own seats on an options exchange. You, the reader of this book, are definitely qualified (if not over-qualified) to be an option market maker. Second, what exactly do option market makers do? The answer is that they make markets, which means that at any given time, for any given option, a market maker for that option posts four numbers—the bid-ask prices and the bid-ask sizes. The bid price and the bid size indicate that he is willing to buy at most the bid size at the bid price; similarly the ask price and the ask size reflect the maximum size he is willing to sell at the ask price.

To understand how actions by market makers change market prices, or equivalently implied volatilities, I need to first examine the tools used by market makers. For over a quarter century, the basic tool was the BS formula (binomial trees for American puts). I think it is still mostly true even to this day. However, the formula is not used with just one volatility input parameter, as it was intended by the BS model, but with a matrix of implied volatilities, one for each strike and maturity.[2] Using the implied volatility matrix, a market maker can generate a fair value for each option, based on which he creates bid and ask prices. The bid-ask sizes are generally determined based on a trader's gut-feeling, or some ad hoc methods.

For market makers, the implied volatility matrix is adjusted after each trade. For example, let us assume that a sizable trade occurred at the bid price, which means that the market makers bought these options. After the trade, the implied volatilities will be lowered somewhat across the board, particularly on the one that has just been traded. The new fair value of the traded option will probably be the old bid price, and of course, the new bid and ask prices will be adjusted accordingly. Obviously the opposite happens when options are traded at the ask prices.

Tradings not only make implied volatilities move up and down, but also cause skews and smiles to change through this adjustment process. For example, if market makers sold the out-of-the-money puts they would raise the implied volatilities on the lower strikes (relative to the current stock price); if they also bought the out-of-the-money calls, they would then lower the implied volatilities on the upper strikes; hence the adjustment creates a negative implied volatility skew. In short, trading causes the implied volatilities to fluctuate in various ways.

---

[2] This is the so-called "sticky strike" case, which means that the implied volatility on an option contract is fixed as the stock moves.

The natural question for now is why market makers adjust the implied volatility downward after buying. The academic answer to this question is that the market's future expected volatility is lower. This statement has certain truth to it, but I do not think it is the main reason. If all market makers think that the future expected volatility would be the bid implied volatility when options are traded at the bid price, then they would not have bid at that implied volatility to begin with.

The main reason for lowering the implied volatility is that it is an ad hoc way of doing inventory control. Here is why: A market maker buys an option because its market price is lower than his model's fair value; if he does not lower the option's fair value after buying, then he would keep on buying at the same price (assuming the stock price has not changed), as long as it is offered, hence potentially end up with a huge position, and the risks associated with it. This is clearly unacceptable. Lowering the implied volatility after buying an option lowers the model's fair value, hence he will stop buying the same option unless its market price decreases. The inventory control reason also explains other implied volatility adjustments after various options trades. There is no doubt that from a market maker's point of view, an option's fair value depends on his current position.

Since this is an extremely important point, it is worth repeating: option market makers use the BS formula in a creative way (inconsistent with the underlying theory) to achieve inventory control, *i.e.*, position-dependent option valuation.

Chances are that you are not an option market maker, and probably use fancier models than the simple BS formula; should you lower an option's fair value after buying it? The answer is that you should, for exactly the same reason. In fact the inventory control argument applies to any risk-averse trader. Therefore, you too should make an option's fair value depend on your current position.

To summarize, in real life, an option's fair value depends on a trader's position. Currently there is no systematic position-dependent option pricing theory in the literature. As a result the position-dependent valuation is achieved through ad hoc manipulations of the model parameters, *e.g.*, adjustments of volatilities.

I will offer a new way to solve to the position-dependent option valuation problem later in this book, one that is systematic and does not rely on ad hoc adjustments of model parameters. In fact, the optimal trading size problem and the position-dependent option valuation problem are closely related. You cannot determine what size to trade if options valuations are position independent.

1.3 The Simulated Options Game

Reality is so complicated, no one can claim to have a perfect strategy for trading options in real life. The best one could do is to set up a model, then look for the optimal strategy under the model market. I now lay out my model market assumptions. Most of the work in this book is based on the model of the simulated options game (SOG); the equity version is described here, the fixed-income related SOG will be discussed in the last part of this book.

Let us assume you are playing a game in front of a computer screen with a bunch of numbers on it, based on which you have to make trading decisions. The numbers on the computer screen represent the movements of a stock, which are generated from the stochastic differential equation

$$ds = \nu s\, dt + v^{\frac{1}{2}} s\, dB^s \tag{1.1}$$

where s is the stock price, and that the stochastic variance v (volatility squared) satisfies a mean-reverting process

$$dv = b\, dt + a\, dB^v \tag{1.2}$$

The two standard Brownian motions dB^s and dB^v are correlated with the correlation function being ρ. Furthermore, all coefficients $\nu(*)$, $a(*)$, $b(*)$ and $\rho(*)$ are assumed to be known functions of t, s and v. In other words, the program used to simulate the stock movement is given to you, thus there is no excuse to adjust the model parameters.

The stock price simulation program does not depend on whether you trade the stock or not (no feedback effects). The game allows you to buy or sell infinite amount of stock at any given price. Unfortunately, every time you trade the stock, you do have to pay a small proportional transaction cost. In addition, you are assumed to be very credit worthy so that you can borrow and lend an infinite amount at the same interest rate in your trading margin account.

Besides the underlying stock, the computer allows you to trade options based on the stock, all of which are assumed to be of European style with known payoff functions. From time to time, the computer shows you a randomly selected option at a given "market" price; it then asks you for your trading size for that option. The game allows an unlimited[3] option trading size at any given "market" price. Note that an implicit assumption made in this book is that you could trade fractional contracts, which means that the number of shares of the stock and the number of options in your portfolio are real numbers, instead of integers.

This is in essence what the simulated options game is. Now I can state unambiguously what this book is about—to offer a rational systematic trading strategy for playing the SOG for a *risk-averse* trader.

I want to point out two subtle but important features of the SOG. The first is that the game does not tell you how the option "market" price is generated. This affects the information set available to you to make your trading decisions. The point is best illustrated using an example: If the long-term stock volatility is 20% in the model, and a long maturity European call option is traded at the implied volatility of 18% (deduced from the "market" price), then it is intuitively clear that you should buy the call option if you have no other options in the portfolio. However, if in addition you are given the information that there is a big seller, and the implied volatility is likely to go down to 16% in the near future, then obviously a better strategy, after receive the new information, is not to buy at 18%, but to sell there and buy later at 16%. Notice that in reality, this type of order flow information is very hard to obtain in the age of anonymous electronic tradings. Also note that if you possess information on near-term movement of the option implied volatility, then you do not need any stock or option models to make your trading decisions. Since in the SOG you are *not* given the information on the option "market" price movement, you must make your trading decisions based on the relative relation between the option "market" price and the model's fair value. This feature of the SOG is designed to mimic the situation in

[3]The unlimited trading size for both the stock and the options should not be taken literally, otherwise the game would allow the well-known doubling-down type of arbitrage strategy in continuous-time.

reality where market prices of options are determined by their short-term supply and demand that are difficult to forecast.

The second feature of the SOG is that there is no guarantee that the game will let you trade the same option again, which means if you put an option into your portfolio, you *may* have to hold it until its maturity. Thus in the SOG, options' "market" prices are not moving continuously, which implies that you trade options only when opportunities arise. This reflects accurately the reality, not only for over-the-counter options, but also for liquid exchange traded options. Because the bid-ask spreads on options are in general an order of magnitude larger than the bid-ask spread of the underlying stock, thus trading options continuously is out of the question for real-life options traders.

Does the SOG resemble options trading in reality? I certainly think so. But ultimately it is up to you to determine whether or not a particular model can help you making real-life trading decisions.

At this point, you may not let me end this section without registering your protest, that the SOG is ill defined, in the sense that the game has not provided you with enough information to make trading decisions. Here is the gist of your argument: In order for you to make trading decisions, you must be able to compute an option's fair value based on the model; to do so in a stochastic volatility environment like the one of the SOG requires knowing a quantity called the market-price-of-risk, which the game has yet to provide.

I now address your concern from two different angles. The first is to offer you a glimpse of what is to come later in this book. Unlike the literature, I emphasize the *personal* nature of options valuations, which is built into the design of the SOG; there is no options market to speak of, just you versus the computer. Instead of the *market*-price-of-risk, which has no meaning in the SOG, your *personal*-price-of-risk can be explicitly derived, which depends on your current position and your risk preference. Hence the game has provided you with enough information to uniquely determine an option's fair value, which is position dependent. The position dependency of an option's valuation induces a natural solution to the optimal trading size problem.

Of course, since you have not finished reading this book yet, this answer is too vague for you right now, so I address your concern from a different angle—through a thought experiment. Assuming the option trading decision has already been made for you, but you still have to make stock trading decisions to hedge your position. Now imagine cloning yourself a billion times, each executes a slightly different hedging strategy. The resulting probability density function of your final wealth in general depends on the initial option trading size. Now you must assign a score to each of the probability density distributions, which induces a way of ranking different hedging strategies. Notice that ranking trading strategies is your problem, not the problem of the SOG's designer. The ranking method I use in this book is that of the expected utility theory, which will be discussed in detail in the next chapter. Once a ranking method is picked, you choose the hedging strategy that has the best score between you and your clones. So far everything is conditioned on the given initial option trading size; now plot the best hedging score as a function of the initial option trading size; if the curve has an interior maximum point, then the x-coordinate of the maximum point is your optimal option trading size. I will argue in the next section that for a risk-averse trader, the maximum point is not at infinity.

1.4 Complete and Incomplete Markets

I provide in this section a qualitative argument for why the optimal trading size problem in the SOG is well defined for a risk-averse trader. The argument is based on the notion of complete and incomplete markets, which will recur later in this book. Here are my heuristic definitions of these concepts: If you can find a non-trivial[4] trading strategy such that the final wealth distribution is a delta function, *i.e.*, no uncertainties (risks), then the model market is a complete market; if no such trading strategy exists, then the model market is an incomplete market.

The BS model is a complete market model, because the combination of an option together with the right stock trading strategy (delta hedging) can make your final wealth distribution a delta function, which is also known as perfect replication. Options in complete markets are riskless. The size you want to trade is infinite if an option is not trading at its BS value. Therefore BS-like complete market models do not provide a sensible solution to the optimal trading size problem that all real-life traders face.

Comparing the assumptions of the BS option theory[5] with those of the SOG, you see that the SOG generalizes BS option theory assumptions in two important areas, namely the change to stochastic volatility and the inclusion of transaction costs. Each of these two modifications breaks the completeness of the BS model. Thus you cannot completely eliminate the risk associated with an option position while playing the SOG. Obviously risk-averse traders will not take on an infinite size position (implying infinite risks) in an incomplete market. Thus being risk averse means that the optimal option trading size in the SOG is not infinite. The bottom line is that the optimal trading size problem occurs naturally when you combine risk aversion and market incompleteness.

Notice that the real-world market is always incomplete, since there are always extra risk factors that are omitted by any given model market. Therefore it is better to model reality with an incomplete market model to begin with. Because the optimal trading size problem is well defined and can be solved naturally for a risk-averse trader in an incomplete market, the essential message of this book can be summarized in the following schematic equation: *Model* + *Risk Aversion* = *How to Trade.*

1.5 What is New?

As mentioned in the previous section that I use the expected utility framework as a way to rank trading strategies in this book, so my solution to the optimal trading size problem must be an application of the utility-based option pricing theory in an incomplete market. In a sense this is the correct answer. If you happen to be familiar with the mathematical finance literature, then you immediately recall that a lot of work has been done on the subject of option pricing in incomplete markets using the expected utility theory.[6] A natural question at this point is: What is new in this book? The answer is plenty, as I now highlight a few in this section.

The first point is intuitiveness. The mathematical finance approach to the utility-based pricing problem so far is of abstract nature, it relies on things such as martingale theory, duality theory. Many derivative

[4] Doing nothing; *i.e.*, keeping all your money in a bank is a trivial trading strategy.
[5] The assumptions used for deriving the famous BS option pricing formula is clearly laid out in the original BS paper [14].
[6] See the latest review article by Henderson and Hobson [42] and the references therein.

quants and traders come from a science and engineering background, and are unfamiliar with these abstract mathematical concepts. On the other hand, I hope the new partial differential equation (PDE) approach presented in this book is more intuitive, as it only requires the understanding of the traditional BS equation derivation and the basics of portfolio optimization theory in continuous-time. I emphasize that the new approach to option pricing using the expected utility theory is novel, it introduces many new concepts for option pricing in incomplete markets. Although many "proofs" in this book are based on intuitive arguments, it is reassuring that some key results can also be obtained from the more abstract mathematical approaches.[7]

The second point is on computation. After more than a decade of research in utility-based option pricing theory in incomplete markets, the whole field so far has had little impact on practitioners. This is partially because some key concepts are missing in the literature, as well as because researchers on this subject so far have been heavy on mathematical concepts, but light on computational aspects. In other words, mathematicians tend to emphasize existence and uniqueness, but real-life traders have to make trading decisions based on specific numbers, which this book will provide by numerically solving a pair of PDEs.

The third point is on explanation. The current mainstream option theory does not satisfactorily explain many real-life options trading behaviors. On the contrary, the new approach in this book offers *natural* explanations to most observations in reality, in the sense that the theory predictions and real-life traders' actions agree.

In short, the application of the expected utility theory to option pricing so far has been abstract in the mathematical finance literature. My approach on this subject is both systematic and coherent; it is not only intuitive, but also computable. In addition, I have made an effort to give financial interpretations to most concepts and results inside this book, so that a real-life trader can easily understand their relevancies to the real world, and will not get lost amidst all the equations.

[7] See the papers by Stoikov [83] and by İlhan, Jonsson and Sircar [49, 50]; they also contain references to the relevant mathematical finance literature.

Chapter 2

Utility Functions

In this chapter, the basics of the expected utility theory are reviewed. The well-known example of the classical pure stock investment problem is used to illustrate the procedure of applying the stochastic control theory. Intuitive criteria for how to choose the risk aversion parameter are given.

2.1 Decisions with Uncertainty

This section addresses the question of how to evaluate a trading strategy. The final wealth of a trader, which is his initial wealth plus the profit and loss (P&L) resulting from applying a certain trading strategy, is in general a random variable. Different trading strategies produce different final wealth probability density distributions. If there is a way to map a probability density distribution into a real number, which can be ranked, then different trading strategies can be evaluated accordingly.

One obvious mapping is to use the mean of the final wealth distribution. This is the risk neutral[1] preference case. A risk neutral preference trader will play the game of doubling his wealth with 50.1% chance, and losing it all with 49.9% chance. In fact his bet size is infinite (allowing leverage) as soon as he believes that the odds are in his favor. However, this attitude towards risk is not typical for most people.

The utility function approach is used to model behaviors of risk-averse traders. Each final wealth w is assigned a utility value $U(w)$, which can be thought as a level of happiness associated with that wealth. By taking expectation of the utility function $E[U(w)] := \int U(w)p(w)\,dw$, a probability density distribution on w, $p(w)$, is mapped into a real number. Using the expected utility theory to make decisions under uncertainty is standard in economics and game theory.

Let me briefly review some basic properties of a generic utility function:[2] a utility function $U(w)$ should be a strictly increasing function, because one always prefers the higher outcome; it should also be a strictly concave function to model risk aversion. Notice that concaveness implies that any bet size will be finite, because the utility associated with negative wealth will have a dominant effect as the bet size

[1]The term risk neutral here does not have the same meaning as the one used in the classical BS option theory.

[2]Expected utility function theory can be found in many standard economics textbooks such as Huang and Litzenberger [48].

gets bigger.

Another property of a utility function is that it is not unique, adding any constant or multiplying any positive constant to it (affine transformation[3]) does not affect the relative rankings of different probability density distributions. The following two functions are invariant under the affine transformation: (i) the absolute risk aversion function

$$R_A(w) := -\frac{U''(w)}{U'(w)} \tag{2.1}$$

where the symbol ":=" means definition and prime denotes taking derivative with respect to w, (ii) the relative risk aversion function

$$R_R(w) := -w\frac{U''(w)}{U'(w)} \tag{2.2}$$

Therefore either of these two risk aversion functions can be used to uniquely specify a family of utility functions.

Three Common Choices

There are three families of utility functions that are commonly used in the literature due to their analytical tractability. The first is the quadratic pseudo utility function, the second is the power utility function, and the third is the exponential utility function.

Let me first discuss the quadratic pseudo utility function, which is often associated with the mean-variance analysis. The quadratic pseudo utility function is

$$U(w) = w - \gamma_q w^2 \tag{2.3}$$

where γ_q is a positive parameter, which controls risk aversion. It is called a pseudo utility function because a true utility function should be monotonically increasing. In real life, I would be delighted if I could discover a trading strategy that has upside potential with limited downside drawback; but unfortunately the quadratic pseudo utility function penalizes such a positive skewed distribution just as it penalizes a negative skewed one. In addition, the utility function approaches negative infinity only quadratically when the wealth level approaches negative infinity, which implies that real bad outcomes are not given big enough penalties. For these reasons, I personally will not base my trading decisions on the results of the quadratic pseudo utility function. Therefore I will not pursue further along this line.

The second family is the power utility function

$$U(w) = \frac{1}{\gamma_p} w^{\gamma_p} \tag{2.4}$$

where the dimensionless risk aversion parameter of γ_p should obviously be less than one in order for $U(w)$ to be concave. The limit $\gamma_p = 1$ is the risk neutral preference case, and the limit $\gamma_p \to 0$ corresponds to the log utility function, $U(w) = \log(w)$ (after discarding a constant term). It is easy to see that the

[3]An affine transformation, $U \to \alpha_1 U + \alpha_0$, is a linear transformation characterized by two constants α_0 and α_1 ($\alpha_1 > 0$).

relative risk aversion function $R_R(w)$ (cf. (2.2)) for the power utility function is a constant, thus the power utility function is also called the constant relative risk aversion (CRRA) utility function.

I did try the power utility function at first, but discovered that it had feature that I did not like. Option pricing equations under the power utility function depend not only on the usual state variables like time and the stock price, but also on the total wealth level w, which is a dynamic state variable that fluctuates. Let me provide an extreme example to illustrate why wealth-dependent option pricing is undesirable for traders. Assuming you have two completely unrelated investments, say a stock and a real estate venture, then the result of the real estate venture will affect your stock option values under the power utility function. In a sense, this is the correct behavior, since the P&L on other investments affects your capital base. The power utility function may be useful in applications with asset allocations, but for traders who need to evaluate each trade, I find this global dependency on the wealth level w overbearing.

The third family, which I will use throughout this book, is the exponential utility function

$$U(w) = -\frac{1}{\gamma}\exp(-\gamma w) \tag{2.5}$$

where γ is a positive parameter that controls risk aversion. A larger γ means more risk aversion, whereas the limit $\gamma \to 0$ corresponds to the risk neutral preference case. I will discuss in detail the practical way to choose the risk aversion parameter γ later in Section 2.4. It is easy to see that the absolute risk aversion function $R_A(w)$ (cf. (2.1)) for the exponential utility function is a constant, thus the exponential utility function is also called the constant absolute risk aversion (CARA) utility function.

An important property of the exponential utility function is that changing the current wealth does not affect the subsequent trading decisions, because the change of the current wealth can be factored out as an affine transformation for the exponential utility function. I believe that I should make my trading decisions based only on my current position and the current market information, regardless of the initial cost for the position (which affects the current wealth), since nobody can go back and change the past. The exponential utility function has this nice memoryless feature.

Finally let me comment on other possible decision theories. As I have argued in the previous chapter, the optimal trading size problem arises naturally due to the combination of market incompleteness and risk aversion. Therefore, in principle, any risk-averse decision theory can be used to solve the optimal trading size problem. My personal bottom line on this issue is simple: The problem must be solvable under the chosen risk-averse decision theory. It is an oxymoron to apply a decision theory that does not provide an answer. If I were to apply an advanced decision theory such that the optimal trading size problem is difficult to solve even numerically, then instead of helping, I would have erected an obstacle for myself in making systematic trading decisions. I would be in no better position than where I started, which was to make trading decisions based on my gut-feeling instinct.

Certainty-Equivalent P&L

The concept of the certainty-equivalent wealth w_c is defined as

$$U(w_c) := E[U(w)] \tag{2.6}$$

The financial interpretation of this is that one is indifferent between the choices of playing the game with a random outcome or taking a certain lump sum w_c (but not playing the game). Since U is a concave function, Jensen's inequality leads to $U(w_c) < U(E[w])$, which means $w_c < E[w]$ because U is a monotonically increasing function.

Notice that the concept of the certainty-equivalent wealth is well defined for a generic utility function, but it is especially useful for the exponential utility function, due to its memoryless feature. Memoryless implies that changing the initial wealth by a certain amount changes the certainty-equivalent wealth by the same amount.

Denote the certainty-equivalent wealth before and after doing a particular trade as $w_c^{\mathcal{B}}$ and $w_c^{\mathcal{A}}$, respectively, the certainty-equivalent P&L (CEPL) Υ for doing the trade is defined to be

$$\Upsilon := w_c^{\mathcal{A}} - w_c^{\mathcal{B}} = -\frac{1}{\gamma} \ln \left(\frac{E[\exp(-\gamma w^{\mathcal{A}})]}{E[\exp(-\gamma w^{\mathcal{B}})]} \right) \tag{2.7}$$

where the second step is only valid for the exponential utility function. The CEPL reflects how much expected utility you have gained or lost by doing the trade. Since the goal is to maximize your expected utility, you never want to do a trade with a negative CEPL voluntarily.

An important financial interpretation of the CEPL is that a trader who uses the exponential utility function is indifferent between the choices of doing the trade or taking the lump sum Υ and forfeiting the trade. Another property is that CEPLs for two different trades are additive, which is again due to the memoryless feature of the exponential utility function.

Lastly let me clarify potential confusions between the concepts of the certainty-equivalent P&L and the realized P&L (the change of your final wealth). Your CEPL is not a random number; using several formulas in Section 4.2, it can be computed before you actually do a trade. On the other hand, your realized P&L is a random number; it becomes known for certain only in the end—your investment horizon. Therefore for making trading decisions, the CEPL is a useful ex-ante measure; whereas the realized P&L is a useless ex-post quantity. Since the realized P&L is in general path dependent in incomplete markets, it is possible for a trade with a positive CEPL to actually lose money, and vice versa. However, if you do the trade many times, *e.g.*, Monte Carlo simulations, a positive CEPL trade will most likely result in a positive realized P&L.

The last point can also be understood from the following angle. The concept of the CEPL is closely associated with the expected utility function $E[U(w)]$, whereas the concept of the *average* realized P&L is closely associated with the expected final wealth $E[w]$. These two concepts coincide in the risk neutral preference case, because $U(w) = w$. Otherwise, the two random variables $U(w)$ and w are closely related, since $U(w)$ is a monotonically increasing function of w. In fact, if your initial portfolio is empty, then using Jensen's inequality, it is easy to show that the *average* realized P&L is always greater than the CEPL. Therefore any strategy that has a positive CEPL is also likely to produce a positive realized P&L.

To Hedge or Not to Hedge

I now provide a simple example on how to use the expected utility theory to make decisions. Assuming the model market consists of a nondividend paying stock in a zero interest rate environment, the stock

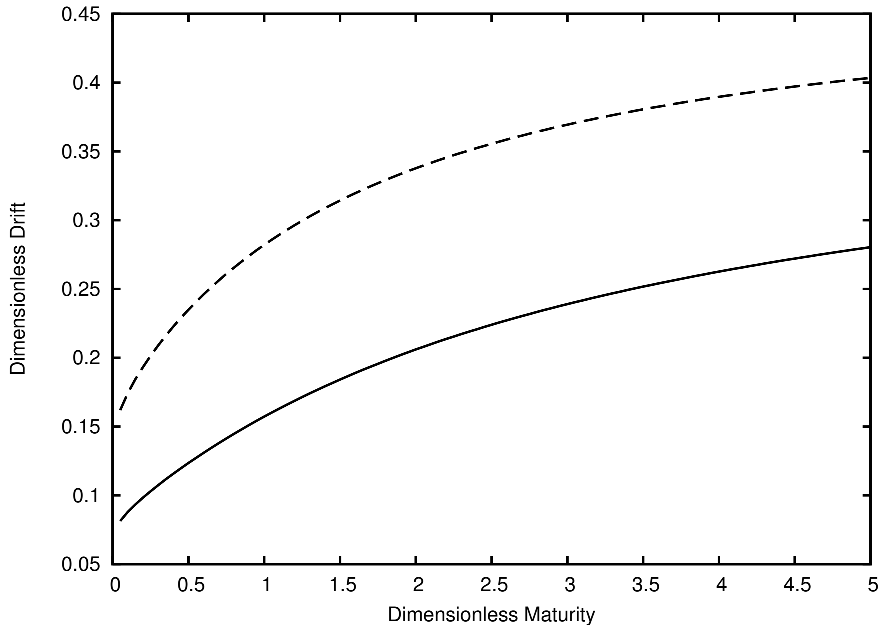

Figure 2.1: The critical dimensionless drift ν/σ^2 is plotted against the dimensionless time-to-maturity $\sigma^2 T$, for two different dimensionless sizes $\gamma n_1 s = 0.2$ (lower curve) and 0.4 (upper curve).

movement is modeled by a geometric Brownian motion with a constant drift ν and a constant volatility σ. Suppose n_1 at-the-money European calls are awarded to you for free (lucky you!), and you are given two choices: either delta hedge them, or do nothing. How do you choose?

The BS delta hedging strategy locks in the values of the call options, which means that your final wealth distribution is a delta function at the value $n_1 C^{BS}$, where C^{BS} is the BS call option value. The expected utility for adopting the hedging strategy is obtained simply by substituting $w = n_1 C^{BS}$ into the utility function definition (2.5).

On the other hand, your final wealth for the do-nothing strategy is

$$w = n_1 \max(0,\, s(T) - s(0))$$

where $s(0)$ is the current stock price, which is also the strike price, and the final stock price $s(T)$ is a lognormal random variable. The calculation of the expected utility value in this case is doing a one-dimensional integral, which does not seem to reduce to an analytical expression; but the integration can easily be done numerically.

Although the BS option value C^{BS} is independent of the drift ν, the final wealth distribution for the do-nothing strategy is highly dependent on ν. Intuitively, it is clearly better to choose the do-nothing strategy if the drift is large; but to choose the hedging strategy if the drift is small. Equating the expected utility values of the two strategies and solving for ν give you the critical value that makes the two strategies indifferent, see Fig. 2.1.

There are two noticeable features in Fig. 2.1, one is obvious, and the other is not. What is obvious is that the curve of higher γ lies above the one of lower γ. This is because a trader with more risk aversion demands a higher drift in order for him to abandon the riskless hedging strategy. What is not so obvious is that the curves are upward sloping. On the one hand, a longer maturity gives more time for the drift to take effect, so it favors the do-nothing strategy; but on the other hand, a longer maturity also makes the BS call more valuable, thus it helps the riskless hedging strategy. In the end the later factor wins out, so a longer maturity requires a higher drift in order for you to abandon the riskless hedge.

2.2 The Present Value Perspective

At the investment time horizon T, the result of applying a certain trading strategy is a probability density distribution of the final wealth $W(T)$, which is then mapped into a number by taking the expectation of a utility function, *i.e.*, $E[U(W(T))]$. Different trading strategies can then be ranked according to their expected utility.

Things become a bit tricky in a nonzero interest rate environment. Let r be the constant interest rate, then the present value of the final wealth is $w(T) := \exp(-rT)W(T)$. Different probability density distributions for $W(T)$ induce corresponding ones for $w(T)$, which can be ranked using the same expected utility method, *i.e.*, $E[U(w(T))]$. The question is whether the rankings based on $W(T)$ is the same as the ones based on $w(T)$.

For users of the power utility function (2.4), the constant discount factor between $W(T)$ and $w(T)$ can be factored out of the utility function as a multiplier. Since an affine transformation of a utility function does not affect the rankings, it does not matter whether one uses $W(T)$ or $w(T)$ to evaluate trading strategies. However, the same argument does not work for the exponential utility function, which implies that the rankings based on $W(T)$ may be different from the ones based on $w(T)$.

Let me provide a concrete example to illustrate the point. Suppose your current wealth is 0.1, and without loss of generality let $\gamma = 1$. You are given the following fair coin flip game: either win 0.12 or lose 0.1, after which you hold your money in a money market account until time T. Will you play this game? A quick computation using (2.5) shows that $\frac{1}{2}[-\exp(0)] + \frac{1}{2}[-\exp(-0.22)] > -\exp(-0.1)$, which means you should play, since the expected utility of playing is larger than the one of not playing. This is the present value $w(T)$ perspective. Now let me show you the $W(T)$ perspective. Assuming that the money has doubled its value between now and T due to interest compounding, the same computation shows that $\frac{1}{2}[-\exp(0)] + \frac{1}{2}[-\exp(-0.44)] < -\exp(-0.2)$, which means that you should decline the game. Therefore you may indeed reach different conclusions depending on whether you use $W(T)$ or $w(T)$.

This immediately raises the question of which perspective to use. I now argue that these two perspectives are equivalent, provided that you let your risk aversion parameter γ be a function of the investment horizon T. Although this seems strange, it is in fact natural. This is because γ is not a dimensionless quantity, it has the dimension of one over the monetary unit. When the interest rate is nonzero, the monetary unit is expanding at the rate of $\exp(rT)$, thus the numerical value of the risk aversion parameter for different investment horizon should be contracting at the same rate, *i.e.*, $\gamma^T = \gamma \exp(-rT)$. Once the adjustment on γ is made, then it does not matter which perspective you choose, as the dimensionless quantity $\gamma^T W(T) = \gamma w(T)$. Since by convention, applying the exponential utility function with the parameter γ implies that the decision is made at the present time, it is natural that you use the present value $w(T)$ perspective. However, if you are adamant about taking the $W(T)$ perspective *without* the adjustment on γ, then bear in mind that the results of the rest of this book may not be applicable to you.

In the present value perspective, every quantity that is measured in the monetary unit is multiplied by the discount factor $\exp(-rt)$. If the price S of a nondividend paying stock is modeled by a geometric Brownian motion

$$dS = \mu S\, dt + \sigma S\, dB^s \tag{2.8}$$

then its present value $s := S\exp(-rt)$ satisfies

$$ds = \nu s\, dt + \sigma s\, dB^s \tag{2.9}$$

where the discounted drift $\nu := \mu - r$. If the current portfolio has n_0 shares of the stock S, and n_1 contracts of an option F in it, then the budget equation (the change of wealth dW during a small time interval dt) is

$$dW = n_0\, dS + n_1\, dF - rn_0 S\, dt - rn_1 F\, dt + rW\, dt \tag{2.10}$$

which becomes the following equation when expressed in the corresponding present value quantities w, s and f

$$dw = n_0\, ds + n_1\, df \tag{2.11}$$

where $f := F\exp(-rt)$. Thus all terms involving the interest rate r drop out of expression (2.11), as if in a zero interest rate environment.

I will use the present value perspective throughout the rest of this book, which is equivalent to presenting all derivations and results in a zero interest rate environment. This treatment simplifies derivations and resulting equations *without* losing any contents. Nonzero interest rate results can be recovered from their corresponding zero interest rate ones through the following two-step procedure: (i) replace the discounted drift ν with $\mu - r$; (ii) replace quantities like the strike price k by its present value $K\exp(-rT)$, where T is the time to maturity. Notice that there are no adjustments for values like s and f, because at the present time $t = 0$, $s = S$, and $f = F$.

2.3 Optimal Strategy for Stock Tradings

Once there is a way to compare different trading strategies, it is natural to ask what the best strategy is. In the expected utility framework, the optimal strategy is defined to be the one that maximizes the final expected utility. This is in essence what the rest of this book is about: finding the optimal strategy for dynamically trading a stock together with its associated options. In this section, I focus on the continuous-time optimal strategy for trading the underlying stock alone. I label this type of portfolio optimization problem as the pure stock investment problem.[4] The stochastic control, or dynamic programming,

[4] This is also known as the Merton investment problem in the literature. Note that the phrase Merton problem is usually associated with problems that involve optimal consumption decisions as well as investment decisions.

method under the power utility function used for the classical Merton's problem[5] can easily be adapted to the current pure stock investment problem under the exponential utility function.

If you know the stochastic control theory (only the basics are needed for this book), then you probably have already seen the material that I am about to present in the next few paragraphs. However, I ask you not to skip this section entirely, because one of the key insights of this book is discussed later in this section.

Despite that the stochastic control theory will be used many times later in this book to determine optimal strategies, it is not the time to quit even if you do not know anything about the subject. You can learn the basics of the stochastic control theory through various sources,[6] or you can simply read on, as I will go through examples to show you the computational procedure involved. The key point is that you can do calculations in the stochastic control theory by following a set of well-defined mathematical steps without having a thorough knowledge of the underlying theory. An analogy to this point in calculus is that you can find a maximum of a function by simply setting its first order derivative to zero, and there is no need to understand the epsilon-delta step in defining the notion of a limit.

I now go through the steps of finding the optimal trading strategy for the continuous-time pure stock investment problem in detail. Assuming the stock pays a constant dividend yield $\tilde{r}$, and its movement is modeled by the standard geometric Brownian motion with a constant drift ν and a constant volatility σ,

$$ds = \nu s\, dt + \sigma s\, dB^s \tag{2.12}$$

Self-financed tradings are done in a margin account with no borrowing or lending restrictions.

Let n_0 be the number of shares of the stock held in the margin account, which is a variable you can control (control variable) through buying or selling. The change of wealth dw, or trading P&L, for your margin account during a small time interval dt is

$$\begin{aligned} dw &= n_0\, ds + \tilde{r} n_0 s\, dt \\ &= n_0(\nu + \tilde{r}) s\, dt + n_0 \sigma s\, dB^s \end{aligned} \tag{2.13}$$

The goal is to choose a trading strategy $n_0(t, w, s)$ to maximize the expected utility at the investment horizon T. The value function J, which is the maximized expected utility conditioned on the current state information, is defined as

$$J(t, w, s) := \sup_{n_0} E[U(w(T))] \tag{2.14}$$

where t is the current time, w is the current wealth, and s is the current stock price. Analogous to the situation in calculus where the necessary condition for finding a maximum point of function is to set its first order derivative to zero, the necessary condition for optimality in the stochastic control theory is the famous Hamilton-Jacobi-Bellman (HJB) equation,

$$\sup_{n_0} \mathcal{L} J = 0 \tag{2.15}$$

[5] See the relevant chapters of Merton [70] and Oksendal [75].

[6] See, for example, the relevant chapters in Oksendal [75]; I also like the elucidating martingale perspective given in the article by Korn [56]. The tutorial by Bressan [17] is also helpful in introducing many related concepts.

with $\mathcal{L}J$ being defined as

$$\begin{aligned}\mathcal{L}J := & J_t + \nu s J_s + n_0(\nu + \tilde{r})sJ_w \\ & + \frac{1}{2}\sigma^2 s^2 J_{ss} + \frac{1}{2}n_0^2\sigma^2 s^2 J_{ww} + n_0\sigma^2 s^2 J_{sw}\end{aligned} \tag{2.16}$$

where the subscripts denote partial derivatives. One way to remember $\mathcal{L}J$ is to recognize that it is the coefficient of the dt term after applying Ito's lemma to $dJ(t, w, s)$.

Equation (2.15) is the result of the Bellman principle of optimality, which states that if the optimal control trajectory between $[t, T]$ is divided into two parts at time t^*, then the sub-trajectory between $[t^*, T]$ is also optimal based on the state information at t^*. Setting t^* to $t + dt$, and using Ito's lemma lead to the HJB equation (2.15).

For the optimal strategy, the expression $\sup_{n_0} \mathcal{L}J$ means that the first order derivative of (2.16) with respect to n_0 should be set to zero,

$$\frac{\partial}{\partial n_0}\mathcal{L}J = 0 \tag{2.17}$$

Solving (2.17) gives the optimal stock holding $\bar{n}_0$,

$$\bar{n}_0(t, w, s) = -\frac{\nu + \tilde{r}}{\sigma^2 s}\frac{J_w}{J_{ww}} - \frac{J_{sw}}{J_{ww}} \tag{2.18}$$

Substituting the optimal strategy $\bar{n}_0$ into the HJB equation (2.15) leads to the following equation after some trivial algebra simplification,

$$\begin{aligned} & J_t + \nu s J_s + \frac{1}{2}\sigma^2 s^2 J_{ss} - \frac{1}{2}\left(\frac{\nu + \tilde{r}}{\sigma}\right)^2 \frac{{J_w}^2}{J_{ww}} \\ & - \frac{1}{2}\sigma^2 s^2 \frac{{J_{sw}}^2}{J_{ww}} - (\nu + \tilde{r})s\frac{J_w J_{sw}}{J_{ww}} = 0 \end{aligned} \tag{2.19}$$

Equation (2.19) is a PDE on $J(t, w, s)$, with the final condition

$$J(T, w, s) = U(w) \tag{2.20}$$

Therefore to obtain the optimal strategy $\bar{n}_0$ using equation (2.18), you need to solve the nonlinear HJB equation first, subject to the final condition (2.20).

For an arbitrary utility function $U(w)$, the nonlinear HJB equation is almost impossible to solve. However, for the exponential utility function (2.5), the variable w can be separated out, which is another main advantage of the exponential utility function. The solution for $J(t, w, s)$ is factored into the following form

$$J(t, w, s) = -\frac{1}{\gamma}\exp(-\gamma w)\exp(-\gamma\bar{\phi}(t, s)) \tag{2.21}$$

Substituting it into the HJB equation (2.19) leads to the following equation for $\bar{\phi}(t, s)$,

$$\bar{\phi}_t - \tilde{r}s\bar{\phi}_s + \frac{1}{2}\sigma^2 s^2\bar{\phi}_{ss} + \frac{1}{2\gamma}\left(\frac{\nu + \tilde{r}}{\sigma}\right)^2 = 0 \tag{2.22}$$

with the final condition

$$\bar{\phi}(T, s) = 0 \tag{2.23}$$

The solution $\bar{\phi}$ does not depend on s, as the final condition is independent of s. It is easy to verify that the solution to equation (2.22) with the final condition (2.23) is

$$\bar{\phi}(t) = \frac{1}{2\gamma}\left(\frac{\nu + \tilde{r}}{\sigma}\right)^2 (T - t) \tag{2.24}$$

The quantity $\bar{\phi}$ has a clear financial interpretation: it is the CEPL for trading the stock. In other words, in order for you to forfeit the right to play the stock trading game, you demand to get paid at least $\bar{\phi}$ amount.

Substituting (2.21) into (2.18), the optimal strategy $\bar{n}_0$ expressed in terms of $\bar{\phi}$ is

$$\bar{n}_0 = \frac{1}{\gamma}\frac{\nu + \tilde{r}}{\sigma^2 s} - \bar{\phi}_s \tag{2.25}$$

Since solution (2.24) means $\bar{\phi}_s = 0$, the optimal stock trading strategy (2.25) says that under the exponential utility function, the amount of money $\bar{\pi} := \bar{n}_0 s$ invested in the stock is a constant that is independent of the investment horizon T. The expression for the optimal investment amount $\bar{\pi}$ is

$$\bar{\pi} = \frac{1}{\gamma}\frac{\nu + \tilde{r}}{\sigma^2} \tag{2.26}$$

Let me summarize the major steps involved in finding an optimal strategy using the stochastic control theory:

- identify the state variables and their dynamical equations (cf. (2.12));
- identify the control variables (*e.g.*, n_0);
- write down the budget equation, *i.e.*, the equation for the change of wealth dw (cf. (2.13));
- apply Ito's lemma to the value function J of the state variables (cf. (2.16));
- set the dt coefficient of dJ to zero (cf. (2.15)), which gives the HJB equation based on the Bellman principle of optimality;
- take first order derivatives of the HJB equation with respect to the control variables to obtain the equations for the optimal control variables (cf. (2.17));
- substitute the optimal control variables into the HJB equation to obtain the nonlinear HJB equation on J (cf. (2.19)), with the value of the utility function as the final condition (cf. (2.20));
- use the special form of the utility function to separate out the wealth variable w (cf. (2.21));
- solve the simplified HJB equation with the appropriate final condition (cf. (2.22) and (2.23));
- obtain the optimal strategy using the solution of the HJB equation (cf. (2.25)).

Since the procedure of the stochastic control theory will be applied many times later in this book to find optimal trading strategies for various problems, it would be helpful to familiar yourself with the list if you are new to the subject.

Notice that HJB equation is only the necessary condition for optimality. Unlike calculus where a simple computation of the second derivative will establish the sufficiency of a local maximum, the sufficiency verification step in the stochastic control theory is more involved. I have left the verification step out of the list, because in many cases it is self-evident whether the resulting strategy corresponds to a maximum or a minimum.

Insight

So far the materials presented in this section are of standard textbook nature, I now offer a new perspective of the optimal trading strategy that seems trivial, but is actually profound.

Recall that the optimal strategy requires me to invest a constant amount of money $\bar{\pi}$ (cf. (2.26)) in the stock; the number of shares, n_0, in my margin account is simply $\bar{\pi}/s$. Rewrite the number of shares–stock price relation into the following form

$$\bar{s} := \frac{\bar{\pi}}{n_0} \tag{2.27}$$

to which I now give a new interpretation. Equation (2.27) defines a fair value, or equilibrium price, $\bar{s}$ of the stock for me based on my current position of n_0 shares. If the market price of the stock is traded below/above $\bar{s}$, then I will buy/sell the stock. Notice that as I am trading, my subjective fair value of the stock $\bar{s}$ changes, since my position n_0 is changing. I will stop trading when my fair value agrees with the market price, *i.e.*, $\bar{s} = s$, at which point I am in an equilibrium with the market. Notice that an equilibrium state implies that the current position is optimal. In other words, as long as my fair value is not the market price, I will keep adjusting my position until my fair value equals the market price, through this mechanism I can determine my optimal trading size.

For a risk neutral ($\gamma = 0$) trader who believes the dividend adjusted drift $\nu + \tilde{r}$ is positive, the stock fair value is infinite, as $\bar{\pi}$ is infinite. Thus he keeps on buying the stock no matter what the market price is. This result is intuitive because a risk neutral trader puts on an infinite size position as soon as he believes that the odds are in his favor.

It is a profound observation that an optimal trading strategy establishes a link between a security's position and its fair value. This is a harbinger of things to come later in this book. Unlike that of a stock, the relationship between an option's position and its fair value is not through an algebraic expression, but instead through a pair of PDEs.

Suboptimal Strategy

The continuous-time optimal stock trading strategy depends on the drift parameter ν. Unfortunately nobody will tell you what the exact value for ν is in real life; besides, ν is impossible to estimate to a high degree of confidence based on historical data. Therefore you can only act on your best guess in real life. The material here attempts to answer the question of what happens if your best estimation $\tilde{\nu}$ is wrong.

Based on your belief, the amount of money you invest in the stock is $n_0 s = \tilde{\pi}$, with $\tilde{\pi}$ being defined as

$$\tilde{\pi} := \frac{1}{\gamma} \frac{\tilde{\nu} + \tilde{r}}{\sigma^2} \tag{2.28}$$

The change of wealth dw, or trading P&L, associated with the suboptimal strategy for your margin account during a small time interval dt is (cf. (2.13))

$$dw = \gamma \sigma^2 \bar{\pi} \tilde{\pi}\, dt + \sigma \tilde{\pi}\, dB^s \tag{2.29}$$

where $\bar{\pi}$ is the true optimal investment amount given by (2.26). Therefore the final wealth $w(T)$ is a Gaussian random variable. Using the elementary identity of the standard Brownian motion $E[\exp(cB(T))] = \exp(c^2 T/2)$, it is easy to compute the expected final utility of the suboptimal strategy, which can be written in the form of $-\exp(-\gamma w)\exp(-\gamma \tilde{\phi})/\gamma$, where w is the current wealth and $\tilde{\phi}$ can be computed to be

$$\tilde{\phi}(t) = \gamma \sigma^2 \left(\bar{\pi} \tilde{\pi} - \frac{1}{2} \tilde{\pi}^2 \right) (T - t) \tag{2.30}$$

Of course, the financial interpretation of $\tilde{\phi}$ is simply the CEPL associated with the suboptimal strategy. If you happen to guess correctly, *i.e.*, $\tilde{\pi} = \bar{\pi}$, then $\tilde{\phi}$ reaches its maximum value $\bar{\phi}$ given by (2.24).

One reason to study suboptimal strategies is to understand when you will get a negative CEPL, after applying a suboptimal strategy. From (2.30), it is clear that $\tilde{\phi}$ is positive only if (i) $\tilde{\pi}$ and $\bar{\pi}$ have the same sign, and (ii) $|\tilde{\pi}| < 2|\bar{\pi}|$, which means that your CEPL is negative if your best guess of the dividend adjusted drift $\nu + \tilde{r}$ has the wrong sign, or if it is more than twice the true value. The first condition is obvious, in which case you will likely to lose money; whereas the second condition is not so obvious, in which case you will probably still make money, but the associated risk is too big.

There is only one strategy, namely the directional neutral strategy, $\tilde{\pi} = 0$, which guarantees that $\tilde{\phi}$ is nonnegative (in fact it is identically zero) no matter what ν is. This seemingly trivial observation will have some implications in options trading to be discussed later in this book.

2.4 Choosing the Risk Aversion Parameter

This section is mostly a digression, you can skip it without affecting the reading of the rest of this book. On the other hand, it is a very important digression, especially for practitioners.

The reason the risk aversion parameter γ is so important is that it *always* appears as a product with the position size in all optimal strategies based on the exponential utility function. Doubling γ cuts the position size in half, assuming everything else being equal. Therefore no other parameters have as much a direct impact on the optimal trading size as γ.

How should you choose what γ to use in practice? One way to find out is to perform the following thought experiment: assuming you are offered to play a game with 50% chance of losing P, and 50% chance of winning $P + 1$, will you play this game? The answer depends on P, if P is small, then you will play, since the game has positive expectation; on the other hand if P is your entire personal wealth, you will probably decline the game, because you do not want to lose it all on a coin flip. Thus there is a value P_0 that makes you indifferent between these two choices; *i.e.*, you play when $P < P_0$, but decline when $P > P_0$. The moment of truth has now arrived, you must decide what your personal P_0 is. Once you have made the decision, then by the definition of indifference, the following equation holds

$$\frac{1}{2}\exp(\gamma P_0) + \frac{1}{2}\exp(-\gamma(P_0 + 1)) = 1 \tag{2.31}$$

where the common factor $-\exp(-\gamma w)/\gamma$ (w is your current wealth) has been removed from both sides of the equation. Assuming $P_0 \gg 1$, then a simple Taylor expansion leads to the solution $\gamma \approx 1/{P_0}^2$.

If you have no difficulty carrying out this thought experiment, and know exactly what your personal P_0 is, then there is little reason to read the rest of this section. However, I find that it is not easy to determine my personal P_0 based on this thought experiment. Hence I will consider the problem from a different angle to shed some light on how to determine γ.

The concept of risk is most naturally associated with the percentage wealth change, rather than the absolute wealth change. For example, a million dollar loss for a billionaire does not have the same impact as that to a mere millionaire. The power utility function (2.4) is often used when dealing with the percentage wealth change. It is natural to seek a link between the exponential and the power utility functions.

Since the value of a utility function is not well defined (subject to an affine transformation), the linkage is based on the absolute risk aversion function (2.1) or the relative risk aversion function (2.2), which is uniquely defined for each family. Equating the corresponding risk aversion functions associated with the two utility functions leads to the relation

$$\gamma = \frac{1 - \gamma_p}{w_0} \tag{2.32}$$

where w_0 is the current wealth level.

Relation (2.32) between γ and w_0 makes intuitive sense: Increasing the wealth level w_0 decreases γ, indicating less risk aversion. The infinite wealth level limit corresponds to the risk neutral preference case. However, there is a problem with relation (2.32), namely the current wealth level w_0 is a dynamic state variable, not a constant. To make γ a constant, as required by the definition of the exponential utility function, w_0 in (2.32) should be replaced by a constant C_r, which is of the same order of magnitude as w_0. The constant C_r can be interpreted as a fixed risk capital for a trader. Therefore the exponential utility function absorbs the total wealth level effect through the risk aversion parameter γ. In practice, the risk capital C_r can be changed manually or made *slowly* adaptive to the wealth level w_0 using a long-run moving average.

The problem of finding what γ to use for the exponential utility function is now transformed into the one of what γ_p to use for the power utility function.

Fractional Kelly Strategy

In order to choose γ_p, I first examine the properties of the optimal strategies associated with the power utility function. The stock movement in the model market is still assumed to be a geometric Brownian motion with drift and volatility being constant (cf. equation (2.12)). The well-known optimal strategy in this case is to invest a constant portion of wealth $\bar{\pi}_p$ in the stock,[7] where $\bar{\pi}_p$ is defined as

$$\bar{\pi}_p := \frac{1}{1-\gamma_p}\frac{\nu+\tilde{r}}{\sigma^2} \tag{2.33}$$

Notice that $\bar{\pi}$ (cf. (2.26)) in the exponential utility case represents the *amount* of wealth, but $\bar{\pi}_p$ (cf. (2.33)) in the power utility case represents the *portion* of wealth. Obviously the risk aversion parameter γ_p controls the size of the bet in the stock (risky asset). The optimal strategy associated with the log utility function ($\gamma_p = 0$) has a special name, it is called the growth optimal strategy, or the Kelly's criterion.[8] The class of constant portion strategies (2.33) for $\gamma_p \leq 0$ is called the fractional Kelly strategies.[9]

I now argue that you should not choose $\gamma_p > 0$ in practice. The budget equation for the fractional Kelly strategy is

$$dw = \frac{\bar{\pi}_p w}{s}\, ds + \bar{\pi}_p w \tilde{r}\, dt \tag{2.34}$$

Substituting equation (2.12) for ds and expression (2.33) for $\bar{\pi}_p$ into (2.34), an application of Ito's lemma leads to

$$\begin{aligned} d\ln w &= \frac{\frac{1}{2}-\gamma_p}{(1-\gamma_p)^2}\left(\frac{\nu+\tilde{r}}{\sigma}\right)^2 dt + \frac{1}{1-\gamma_p}\left(\frac{\nu+\tilde{r}}{\sigma}\right) dB^s \\ &:= \alpha\, dt + \beta\, dB^s \end{aligned} \tag{2.35}$$

where the second equal sign serves as the definitions for the constants α and β. The expression for $d\ln w$ says that $\ln w$, which is the wealth growth rate, is a Brownian motion with a constant drift α and a constant volatility β. It is also immediately clear that $\alpha \leq 0$ if $\gamma_p \geq \frac{1}{2}$, which means if you use the power utility function with $\gamma_p \geq \frac{1}{2}$, then the optimal strategy, which maximizes the expected power utility function, surely leads to ruin ($\ln w \to -\infty$ as $t \to \infty$)! Furthermore, it is easy to see that α is a decreasing function of γ_p when $\gamma_p > 0$, and that the magnitude of the volatility coefficient β is an increasing function of γ_p. Therefore the wealth growth rate of applying the $\gamma_p = 0$ strategy dominates the ones of $\gamma_p > 0$.[10] I am amazed that no textbook mentions this simple fact—no sane person would ever choose a power utility function with $\gamma_p > 0$.

Notice that the stock price in the model market here is a simple geometric Brownian motion, if one were to perturb this simple model market into a more complicated one, then it is likely that the optimal strategies corresponding to the power utility function with $\gamma_p > 0$ will lead to troubles.[11] Therefore I propose a pseudo mathematical conjecture that only those optimal strategies corresponding to the power utility function with $\gamma_p \leq 0$ are "well-behaved" for all "reasonable" model markets. This means that the log utility function should be the riskiest choice in the family of power utility functions in practice.

---

[7] If you are unfamiliar with this result, you can follow the steps of finding an optimal strategy outlined in Section 2.3 to derive it. In addition to the power utility function, the constant portion strategy also maximizes (or minimizes) other objective functions, see the paper by Browne [18].

[8] I find the paper by Thorp [84] to be a good introduction on Kelly's criterion in gambling and investing.

[9] The paper by MacLean and Ziemba [65] discusses some properties of the fractional Kelly strategy and its application in various betting situations.

[10] The dominance argument of the $\gamma_p = 0$ strategy is also made in the paper by MacLean, Ziemba and Li [66] (their proposition 3).

[11] The papers by Kim and Omberg [54], and by Korn and Kraft [57] provide several examples on how things can go wrong for power utility functions with positive γ_p.

Intuitively speaking, a larger bet $\bar{\pi}_p$ on the risky asset leads to larger risks. However, the utility function formulation for the portfolio optimization problem is non-intuitive, as far as risks are concerned. In other words, γ_p is not an intuitive risk measure; just pause for a moment and ask yourself what your personal γ_p is before proceeding. What I will do next is to explicitly define and compute two intuitive risk measures for the fractional Kelly strategy, which will establish a link between γ_p and an intuitive risk level. Thus you can choose γ_p indirectly by selecting an easily understood risk level.

Drawdown Risk Measures

I need to introduce a few more notations before I can define the intuitive drawdown[12] related risk measures. Let $\breve{w}(t) := \sup_{0<s<t} w(s)$ be the maximum level of wealth achieved during the time interval $[0, t]$, since $\ln w$ is a simple Brownian motion with a constant drift, the joint probability density distribution of $\ln w$ and $\ln \breve{w}$ is known to be[13]

$$\rho(t, \ln w, \ln \breve{w}) = \sqrt{\frac{2}{\pi}}(\beta^2 t)^{-\frac{3}{2}}(2\ln \breve{w} - \ln w - \ln w_0) \exp\left(-\frac{\alpha^2 t^2 + (2\ln \breve{w} - \ln w - \ln w_0)^2 - 2\alpha t(\ln w - \ln w_0)}{2\beta^2 t}\right) \tag{2.36}$$

where $w_0 := w(0)$ is the initial value of w, and the formula assumes that the condition $\alpha > 0$ is satisfied. Define the current drawdown to be

$$x(t) := \ln \breve{w}(t) - \ln w(t) \tag{2.37}$$

the probability density distribution for x can be obtained by using (2.37) to substitute for $\ln w(t)$ in (2.36) then integrating over $\breve{w}$ from w_0 to ∞

$$\rho(t, x) = \sqrt{\frac{2}{\pi\beta^2 t}} \exp\left(-\frac{(x + \alpha t)^2}{2\beta^2 t}\right) + \frac{2\alpha}{\beta^2} \exp\left(-\frac{2\alpha}{\beta^2} x\right) N\left(-\frac{x - \alpha t}{\sqrt{\beta^2 t}}\right) \tag{2.38}$$

where $N(x) := \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$ is the cumulative normal distribution function. In the limit $t \to \infty$, $\rho(t, x)$ approaches a stationary distribution

$$\rho(x) = \frac{2\alpha}{\beta^2} \exp\left(-\frac{2\alpha}{\beta^2} x\right) \tag{2.39}$$

The first drawdown related intuitive risk measure is the average-percent drawdown D, which is defined as

$$D := \lim_{T\to\infty} \frac{1}{T} \int_0^T \frac{\breve{w}(t) - w(t)}{\breve{w}(t)}\, dt \tag{2.40}$$

---

[12] Various aspects of drawdown risks in the context of continuous-time portfolio management are discussed in the following papers [19, 26, 39, 68, 76].

[13] See the Handbook of Brownian Motion [16].

The meaning of the average-percent drawdown is easily understood. Using the definition of the current drawdown (2.37), and the fact that long-term time average is the same as the ensemble average due to ergodicity, D can be explicitly calculated to be

$$\begin{aligned} D &= \lim_{T\to\infty} \frac{1}{T}\int_0^T \left[1 - e^{-x(t)}\right]\, dt \\ &= \int_0^\infty (1-e^{-x})\rho(x)\, dx \\ &= \frac{1}{2}\frac{1}{1-\gamma_p} \end{aligned} \tag{2.41}$$

where the stationary distribution (2.39) and the definitions of α and β (cf. (2.35)) were used in the last step. Equation (2.41) establishes a link between the easily understood risk measure D and the non-intuitive risk aversion parameter γ_p. Therefore γ_p can be chosen by first picking an average drawdown risk level D, say 10%, then using (2.41) to find the corresponding γ_p, which is -4 for this example. It is known in the literature that the Kelly criterion ($\gamma_p = 0$) is risky, now you know how risky—the average-percent drawdown for the strategy is 50%!

The second drawdown related intuitive risk measure is the equal-percent drawdown, which is defined as the current drawdown d such that the probability of being down d (percent) is d (percent). The risk measure d is also easily understood, for example, a risk level of 30% means that you can tolerate a 30% chance of being down 30%. Anyone with an equal-percent drawdown risk level of 1% probably puts almost all the wealth into a bank account, on the other hand the one with a risk level of 99% uses almost all the money to buy lottery tickets.

A current drawdown of d (percent) means $(\breve{w} - w)/\breve{w} = d$, the corresponding current drawdown variable x_d is

$$x_d := \ln \breve{w} - \ln w = -\ln(1-d) \tag{2.42}$$

Applying integration by parts on the second term of expression (2.38), the probability of being down d (percent) at time t is

$$\begin{aligned} P_d(t) &:= \int_{x_d}^\infty \rho(t,x)\, dx \\ &= N\left(-\frac{x_d+\alpha t}{\sqrt{\beta^2 t}}\right) + \exp\left(-\frac{2\alpha}{\beta^2}x_d\right) N\left(-\frac{x_d-\alpha t}{\sqrt{\beta^2 t}}\right) \end{aligned} \tag{2.43}$$

By the definition of equal-percent drawdown, d satisfies

$$\max_t P_d(t) = d \tag{2.44}$$

It is not hard to show that $P_d(t)$ is an increasing function of t, as the tail area of the unsteady distribution $\rho(t,x)$ in (2.38) is an increasing function of t. By using the asymptotic value of P_d at infinity,[14] equation (2.44) becomes

$$\exp\left(-\frac{2\alpha}{\beta^2}x_d\right) = d \tag{2.45}$$

[14] This overestimates the risk of a finite investment horizon trader. But overestimation is prudent, it is better safe than sorry.

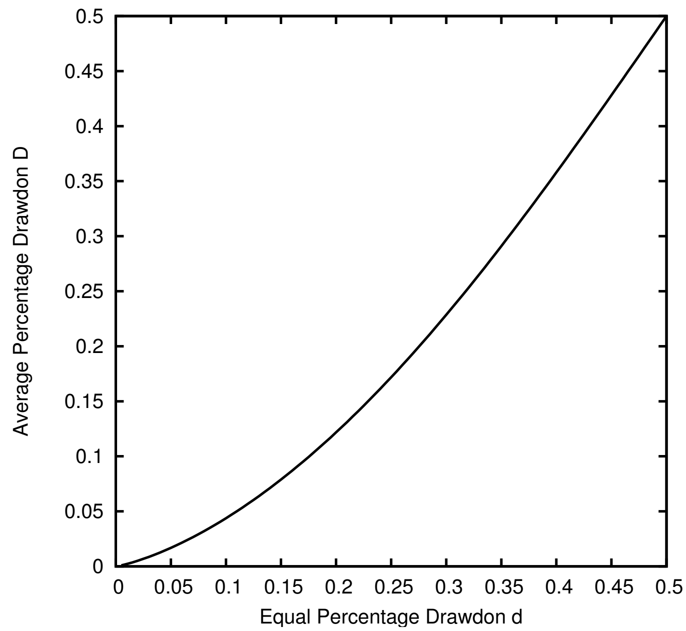

Figure 2.2: Relationship between the two drawdown related risk measures

which reduces to the following relation after substituting x_d (cf. (2.42)), α and β (cf. (2.35)) into (2.45),

$$\gamma_p = \frac{1}{2} - \frac{1}{2}\frac{\ln d}{\ln(1-d)} \tag{2.46}$$

Relation (2.46) links γ_p with an intuitively defined risk measure d. The risk associated with the Kelly criterion measured in the equal-percent drawdown risk is that it has a 50% chance of being down 50%!

The average-percent drawdown D and the equal-percent drawdown d are, of course, related. It is straightforward to show that

$$D = \frac{\ln(1-d)}{\ln d + \ln(1-d)} \tag{2.47}$$

Relationship (2.47) between the two drawdown related risk measures is plotted in Fig. 2.2. Notice that the curve lies below the diagonal line $D = d$, which means for the same numerical value, the equal-percent drawdown is more conservative. For example, a trader with a 10% equal-percent drawdown is more risk averse than a trader with a 10% average-percent drawdown. In fact from (2.47), a 10% equal-percent drawdown corresponds to a 4.4% average-percent drawdown.

The suggested procedure for choosing the risk aversion parameter γ of the exponential utility function is summarized as follows

- use a drawdown risk measure that you can intuitively relate to;
- identify the maximum risk level you are comfortable with;
- find the corresponding γ_p associated with that risk level;

- choose your risk capital, or trading equity, C_r;
- set γ to $(1-\gamma_p)/C_r$.

Be Conservative

Finally, a word of caution, the risk level associated with any risk measure is model specific. In addition, model parameters have so far been implicitly assumed to be known precisely. However, the risk encountered in the real world is always greater than that of the model world, because the model misspecification and the parameter uncertainty risks are unaccounted for in the model world, no matter what models you use. Therefore it pays to be a little more conservative, which means using a γ larger than the one outlined in the aforementioned steps. Because I have tremendous respect for unquantifiable risks, my personal rule of thumb is to use a γ several times as big, which implies reducing the position size by quite a bit.

I now present you with a concrete example illustrating why this cautious attitude is warranted. As discussed earlier in Section 2.3 that the continuous-time trading strategy you use in real life is almost surely suboptimal, since the drift parameter ν is not known precisely. Recall that in order for the suboptimal strategy to have a positive CEPL, you must guess correctly the sign of the dividend adjusted drift $\nu+\tilde{r}$, and that your educated guess cannot be more than twice the true value. An implicit assumption used to reach this conclusion is that you use the same risk aversion parameter γ for the two cases of ν known and ν unknown.

Let us investigate what happens if you decide to use a larger risk aversion parameter $\tilde{\gamma}$ (more risk averse) for the situation of uncertain ν, which is always the case in reality. Expression (2.30) remains valid, which means that the CEPL is negative when the magnitude of your investment amount $|\tilde{\pi}|$ is more than twice the true optimal amount $|\bar{\pi}|$. Substituting expressions for $\tilde{\pi}$ and $\bar{\pi}$ (cf. (2.28) and (2.26)) into this condition leads to the conclusion that the CEPL is positive unless the magnitude of the educated guess of the dividend adjusted drift is larger than a higher threshold, *i.e.*, $|\tilde{\nu}+\tilde{r}| > 2|\nu+\tilde{r}|\tilde{\gamma}/\gamma$. You see that you gain more room for error when you are more risk averse, or you gain robustness by reducing position size.[15] Needless to say that you still need to guess the sign of $\nu+\tilde{r}$ correctly in order for the CEPL to be positive. The downside of being cautious is that your reward is not as big if your educated guess happens to be correct.

2.5 Global or Local Equilibrium

I end this chapter on a philosophical discussion about the concept of equilibrium, which is often associated with the expected utility maximization framework in the literature.

It seems to me that most of work on equilibriums in economics ended up in determining how an asset should be priced when the market is in a global equilibrium, by balancing the supply and demand. I am philosophically wary of such equilibrium theories, as I do not believe the whole market will ever reach a global equilibrium, which is a viewpoint shared by many real-life practitioners. In theory nothing will

[15] Robust portfolio selection problems are discussed in the paper by Maenhout [67], and references therein.

trade in a global equilibrium market. If record stock and option trading volumes cannot convince you that the market is not in a global equilibrium, then nothing will.

On the contrary, the equilibrium notion in this book is used in the local sense only. When I say I am in equilibrium with the market, I mean my subjective fair values of securities agree with their respective market prices, hence I do not trade with the market, at least temporarily. Other traders will not be in local equilibriums with the market at the same time, their tradings with the market will cause market prices to fluctuate. After which I need to trade with the market to re-establish a local equilibrium. In short, I believe the market will never reach a global equilibrium, but that does not preclude me from constantly updating my position to reach a local equilibrium with the market.

Notice that in the economics literature, the term global equilibrium is referred to as general equilibrium, whereas the term local equilibrium is referred to as partial equilibrium. I choose to use the terms global and local partly because they are more descriptive, and partly because I want to make the following analogy to an example in physical science which illustrates my point well.

The laws of thermal dynamics tell us that all things will eventually reach a thermal equilibrium (but it is difficult to determine how long it will take). Based on this, one could ask the question what the temperature would be if the whole earth atmosphere were in a global thermal equilibrium. I think this is a doable physics problem; one could get a rough answer by balancing the energy the earth radiates into the surrounding space, which depends on the assumed uniform atmospheric temperature, against the radiation energy it receives from the sun (maybe include the radioactive decay energy from the earth's interior as well). The answer to this problem might be of interest to planetary scientists, but personally I am not so interested in finding out the hypothetical global equilibrium temperature for the earth's atmosphere, because it is obviously not in a global thermal equilibrium. However, I am interested in a system that tells me the amount of clothes I should wear for any given local temperature, so that I feel neither cold nor hot. If the temperature changes the next day, then I need to put on or take off some clothes in order to maintain the local thermal equilibrium. The fact of the earth's atmosphere not being in a global thermal equilibrium does not prevent me from constantly adjusting the amount of clothes I wear to reach a local thermal equilibrium with my environment.

Chapter 3

Dynamic Derivation

This chapter lies at the heart of this book, as far as the methodology is concerned. You will see a brand new way of deriving option pricing equations, one that is applicable to both complete and incomplete markets alike.

3.1 Static-Dynamic Equivalence Principle

In reality, there are liquid and illiquid options. Let us now extrapolate things to their respective limits, and imagine a model world where European options come with two flavors: The static-flavored options are those that once in your portfolio, you must hold them until maturity, and the dynamic-flavored options are those that are continuously traded without friction on the market. It is obvious that the static-flavored ones represent the illiquid over-the-counter options, and the dynamic-flavored ones model the idealized (zero bid-ask spread and no transaction costs) liquid exchange-traded options.

It is clear that with the same payoff function, a dynamic-flavored option is always worth at least as much as its corresponding static-flavored cousin, because you can always choose to hold a dynamic-flavored option to maturity without trading it, similar to the argument that an American style option is always worth no less than the corresponding European one. The question is how much liquidity premium you should pay for a dynamic-flavored option over that of a corresponding static-flavored one.

The answer is—*zero*. The reason is a fairly trivial arbitrage argument: If a dynamic-flavored option were to trade at a premium over the corresponding static-flavored option in a model world, then you could immediately short the corresponding dynamic-flavored option whenever you put a static-flavored option into your portfolio, which would give you a riskless profit, as the positions of the two flavored options cancel out at the maturity.

Therefore for valuation purposes, an option, no matter how often it is actually traded, can be regarded as if it were continuously traded without friction. I call this the static-dynamic equivalence principle. What do you gain from such a dynamic perspective? The answer is that if an option is continuously traded, then its position in the portfolio is a control variable, you can apply the stochastic control theory to find the optimal strategy.

The following point is *extremely* important: the aim of the dynamic derivation methodology is to discover a fictitious option price process under which your current position remains optimal. Thus by fictitiously making an option continuously traded without friction, I can use the machinery of stochastic control theory to establish a link between an option's position and its fair value, from which the optimal trading size problem can be solved.

It will be shown later in this chapter that in an incomplete market, the optimal trading strategy establishes a link between an option's position and its fair value through a pair of PDEs. But let me first show what the dynamic perspective will provide in the complete market case, which is the topic of the next section.

3.2 The BS Equation: A New Approach

The model market is assumed to consist of one stock s, which pays a continuous dividend yield $\tilde{r}$, and one type of European option f with the payoff function $f(T, s(T))$, where T is the option expiration time. Once again the trading takes place in a zero interest rate margin account with unlimited lending and borrowing.

As before, the stock movement is modeled by a geometric Brownian motion with a constant drift ν and a constant volatility σ,

$$ds = \nu s\, dt + \sigma s\, dB^s \tag{3.1}$$

Applying Ito's lemma to the option equilibrium price $f(t, s)$ gives

$$df = (f_t + \nu s f_s + \frac{1}{2}\sigma^2 s^2 f_{ss})\, dt + \sigma s f_s\, dB^s, \tag{3.2}$$

where the subscripts t and s denote their respective partial derivatives. I emphasize that f is your personal fair value, it does not represent the market price of the option.

I introduce a shorthand notation, which comes handy for various derivations later in this book. An Ito type stochastic differential equation for a dynamic state variable z is written as

$$dz = C_{(z,t)}\, dt + C_{(z,x)}\, dB^x + C_{(z,y)}\, dB^y + \cdots \tag{3.3}$$

where dB^x and dB^y are Brownian motions for the random factors x and y, respectively. Equation (3.3) should be viewed as the definition for the shorthand notation $C_{(*,*)}$ of the coefficients in a stochastic differential equation. For example, from (3.1) $C_{(s,t)} = \nu s$, and from (3.2) $C_{(f,s)} = \sigma s f_s$.

With n_0 shares of stock and n_1 options in the portfolio, the change of wealth dw, or trading P&L, for the margin account during a small time interval dt is composed of three terms: (i) the stock price movement, (ii) the change of the option value, (iii) dividend payout,

$$\begin{aligned} dw &= n_0\, ds + n_1\, df + \tilde{r} n_0 s\, dt \\ &= \left[n_0 C_{(s,t)} + \tilde{r} n_0 s + n_1 C_{(f,t)}\right] dt + \left[n_0 C_{(s,s)} + n_1 C_{(f,s)}\right] dB^s \\ &:= C_{(w,t)}\, dt + C_{(w,s)}\, dB^s \end{aligned} \tag{3.4}$$

The explicit expressions for $C_{(w,t)}$ and $C_{(w,s)}$ can be found by substituting in the coefficients $C_{(s,*)}$ and $C_{(f,*)}$.

The traditional derivation of the BS equation is based on the static point of view, in which a European option is held fixed in the portfolio ($n_1 = 1$), but the number of shares of the underlying stock is continuously adjusted. The BS delta hedging argument—a portfolio that consists of an option and $n_0 = -f_s$ shares of stock is riskless, leads to the famous BS equation.

In the new dynamic derivation approach, both the number of shares of the underlying stock n_0 and the number of options n_1 in the portfolio can change. The aim is to choose the control variables n_0 and n_1 optimally, in the sense that the expected utility function $E[U(w(T_0))]$ is maximized, where the investment horizon T_0 is greater or equal to the expiration time T. Since options do not exist beyond T and that the pure stock investment problem of Section 2.3 has no horizon effect, you use the same strategy whether the investment horizon is T_0 or T. Hence, without loss of generality, I let the investment horizon T_0 equal to the maturity T.

It is time to ask yourself a *crucial* question: Is my current position optimal? In other words, what will the necessary conditions for optimality lead to? I now turn to the standard machinery of the stochastic control theory to answer this question.

The value function J, which is the maximized expected utility conditioned on the current state information, is defined as

$$J(t, w, s) := \sup_{n_0, n_1} E[U(w(T))] \tag{3.5}$$

where t is the current time, w is the current wealth, and s is the current stock price. The HJB equation, which is the necessary condition for optimality, is

$$\sup_{n_0, n_1} \mathcal{L}J = 0 \tag{3.6}$$

with $\mathcal{L}J$ being defined as

$$\begin{aligned} \mathcal{L}J := {} & J_t + C_{(s,t)} J_s + C_{(w,t)} J_w \\ & + \frac{1}{2} C_{(s,s)}^2 J_{ss} + \frac{1}{2} C_{(w,s)}^2 J_{ww} + C_{(s,s)} C_{(w,s)} J_{sw} \end{aligned} \tag{3.7}$$

where the subscripts on J denote partial derivatives.

The expression $\sup_{n_0,n_1} \mathcal{L}J$ means that the first order derivative of $\mathcal{L}J$ with respect to n_0 and n_1 should be zero, respectively.

The condition $\partial(\mathcal{L}J)/\partial n_0 = 0$ leads to the following equation, bearing in mind that the n_0 dependency comes from the coefficients $C_{(w,t)}$ and $C_{(w,s)}$ given by (3.4),

$$\left[C_{(s,t)} + \tilde{r}s\right] J_w + C_{(s,s)} \left[C_{(w,s)} J_{ww} + C_{(s,s)} J_{sw}\right] = 0 \tag{3.8}$$

After substituting all the coefficients $C_{(*,*)}$ into (3.8), the solution for the optimal number of shares $\hat{n}_0$ in the portfolio is

$$\hat{n}_0 = -\frac{\nu + \tilde{r}}{\sigma^2} \frac{1}{s} \frac{J_w}{J_{ww}} - \frac{J_{sw}}{J_{ww}} - n_1 f_s \tag{3.9}$$

By abuse of notation, I denote n_1 both as the control variable, as well as its optimal value, due to the assumption that your current option position n_1 remains optimal until expiration.

Substituting the optimal strategy (3.9) into the HJB equation (3.6) leads to the same equation as (2.19), and the final condition (2.20) remains valid. In light of the optimal stock trading strategy for the pure stock investment problem (cf. (2.18)), the optimal stock holding (3.9) for $\hat{n}_0$ can be written as

$$\hat{n}_0 = \bar{n}_0 - n_1 f_s \tag{3.10}$$

Notice that this decomposition of the optimal stock trading strategy is independent of the utility function. The concept of optimal hedging delta n_0^h is defined to be the difference between the optimal stock trading strategies corresponding to the cases with and without options in the portfolio, *i.e.*, $n_0^h := \hat{n}_0 - \bar{n}_0$. Equation (3.10) says that the optimal hedging scheme is the same as the well-known delta hedging scheme.[1]

The condition $\partial(\mathcal{L}J)/\partial n_1 = 0$ gives rise to the following equation, bearing in mind that the n_1 dependency comes from the coefficients $C_{(w,t)}$ and $C_{(w,s)}$ given by (3.4),

$$C_{(f,t)} J_w + C_{(f,s)} \left[C_{(w,s)} J_{ww} + C_{(s,s)} J_{sw} \right] = 0 \tag{3.11}$$

Combining equations (3.8) and (3.11) immediately leads to

$$\frac{C_{(f,t)}}{C_{(f,s)}} = \frac{C_{(s,t)} + \tilde{r}s}{C_{(s,s)}} \tag{3.12}$$

Substituting the coefficients $C_{(*,*)}$ defined by (3.1) and (3.2) into equation (3.12) produces the famous BS equation under zero interest rates:

$$f_t - \tilde{r} s f_s + \frac{1}{2}\sigma^2 s^2 f_{ss} = 0 \tag{3.13}$$

Let me emphasize that the BS equation is independent of the utility function, because no explicit assumption on its form is made during the derivation.

So far, I have reproduced the classical results of the BS option pricing theory using the dynamic derivation approach. Maybe you are a little underwhelmed by this winding and inelegant way of rederiving the BS equation.[2] The following observation might change your perception about the new approach: In the traditional BS equation derivation, the delta hedging argument is a critical *input*; but in the new dynamic derivation of the BS equation, the delta hedging idea is an *output*! This important fact should not be overlooked. Thus the new dynamic derivation approach is not merely a rehash of some old concepts, but instead offers a different insight.

The concept of perfect replication, which plays a crucial role in the BS option pricing theory, does not enter into the dynamic derivation as an input. In fact the output of the dynamic derivation reveals that the model market is complete. This is how the logic goes: if the market price does not agree with the solution (fair value) of the option pricing equation (the BS equation), then your current position is not optimal, which means you need to adjust your position by trading with the market. Because the BS equation does not depend on the option position n_1 (a feature you will not see when markets are

[1] It will be shown later that this conclusion is false for incomplete market models.

[2] This BS equation derivation based on the stochastic control theory is not in the literature. In addition to the original derivations by Black and Scholes [14] and by Merton [70] (Chapter 8), other perspectives can also lead to the BS equation, see [6, 13, 41].

incomplete), you would keep on trading but never reach a local equilibrium with the market, hence you would end up with an infinite size option position. It seems puzzling that a risk-averse trader would take an infinite option position—unless the option is riskless, which is indeed the case here, as shown by the classical delta hedging argument.

The whole point of the dynamic derivation approach is that it requires little insight, by *blindly* following a well-defined mathematical procedure of the stochastic control theory, the idea of delta hedging reveals itself, together with the BS equation and the completeness of the model market.

The perfect replication argument used in the traditional BS option theory does not apply to incomplete market models, because no trading strategies can make the final wealth distribution be a delta function. Since the perfect replication concept does not enter into the dynamic derivation methodology, it can be applied to incomplete markets as well as complete ones. Indeed you will see later in this book that the dynamic derivation approach for incomplete market models yields fruitful results.

3.3 Restricted Stock: Static Perspective

The BS complete market option theory requires two partners, an option and its underlying stock, to dance together in a specific pattern (delta hedging) to eliminate all the risks. If trading restrictions are applied such that the two partners can no longer dance like they use to, then the model market becomes incomplete. In this section and the next, I will make the market incomplete by forbidding tradings of the underlying stock (restricted stock).

The restricted stock model is not as artificial as it seems to be. If the number of shares in the portfolio is restricted to be zero, then there is no difference between this model and the one of a nontrading risk factor, such as a real option.[3] Another important case is when stock transaction costs are present, it will be shown later in this book that there is a no-transaction region inside which the restricted stock model applies.

The main reason I choose the restricted stock model to illustrate many key concepts related to incomplete markets is that the problem can be viewed from two perspectives. I will address the static perspective in this section, and the dynamic perspective in the next.

Assuming that the stock does not pay any dividend ($\tilde{r} = 0$), and that there are n_0 shares of stock, and n_1 European options with payoff function $f(T, s(T))$ currently in your portfolio, where n_0 and n_1 are viewed not as control variables, but as parameters in the problem, the final wealth at the expiration time T is

$$w(T) = \tilde{w} + n_0 s(T) + n_1 f(T, s(T)) \tag{3.14}$$

where the constant $\tilde{w}$ is related to the initial wealth w, the initial stock price s and the option market price p, *i.e.*, $\tilde{w} = w - n_0 s - n_1 p$. The expected utility at T is simply $E[U(w(T))]$. If you bought an extra small amount of option δn_1 initially at the market price p, then the resulting change of your final wealth

[3] Real options are hard to define mathematically, but are easy to describe, such as whether a company should pursue a capital project like land development.

is

$$\delta w(T) = [f(T, s(T)) - p] \, \delta n_1 \tag{3.15}$$

The change of the expected utility $\delta\mathcal{U}$ due to this change of δn_1 is

$$\begin{aligned}
\delta\mathcal{U} &= E[U(w(T) + \delta w(T))] - E[U(w(T))] \\
&= E[U'(w(T))(f(T, s(T)) - p)]\delta n_1 \\
&:= E[U'(w(T))](f - p)\delta n_1
\end{aligned} \tag{3.16}$$

where the prime denotes taking the first order derivative, and the quantity f on the right-hand side of the last equal sign is defined as follows

$$f := \frac{E[f(T, s(T))U'(w(T))]}{E[U'(w(T))]} \tag{3.17}$$

I now show that f is the equilibrium price.[4] If the market price p of the option is traded less than f, then from equation (3.16) $\delta\mathcal{U}/\delta n_1 > 0$, because $U'(w) > 0$. Conversely, $\delta\mathcal{U}/\delta n_1 < 0$, when $p > f$. Therefore buying the option increases your expected utility when it is traded below f, while selling is required if it is traded above f. If $f = p$, then your current position is optimal, which means you want to hold your current position. To see that changing the current optimal position decreases the expected utility, you need to expand (3.16) to the second order of δn_1, since the first order is zero. Due to the concave nature of a utility function, $U''(w) < 0$, $\delta\mathcal{U} < 0$ if $\delta n_1 \neq 0$. Therefore f is the equilibrium price, and you are in a local equilibrium with the market when $f = p$.

From expression (3.17), it is evident that the equilibrium price f depends on the option position n_1, as $w(T)$ depends on n_1. Thus changing the option position n_1 in the portfolio affects the option's equilibrium price.

It is obvious that the equilibrium price f depends on the option payoff function $f(T, s(T))$, but the functional dependence is not linear, because $w(T)$ in (3.17) depends on $f(T, s(T))$ as well. Therefore an important conclusion is immediately reached that in incomplete markets, an option's equilibrium price, or fair value, cannot be written in the form of $E^Q[f(T, s(T))]$, where Q is a probability measure independent of the payoff function. This is in contrast to the complete market situation where an option's price can always be written in the linear functional form of $E^Q[f(T, s(T))]$, where Q is a risk-neutral measure independent of the payoff function.

So far I have kept the discussions in this section generic without specifying the utility function $U(w)$, the payoff function $f(T, s(T))$ and the distribution of the stock price $s(T)$. To carry out further computations, all three aforementioned quantities must be specified. The exponential utility function (2.5) is used; the payoff function of the European option $f(T, s(T))$ is assumed to be of that of a vanilla call option, *i.e.*, $\max(0, s(T) - k)$; and the probability density function for $s(T)$ is a lognormal distribution with parameters $(\nu - \frac{1}{2}\sigma^2)T$ and $\sigma^2 T$.

[4] This quantity has a few different names associated with it in the literature. It is called the Davis price by some mathematical finance researchers, due to the work of Davis [29]; Carr and Madan [21] call it the personalized price. I think the idea goes way back, see Chapter 7 of Merton [70].

For the exponential utility function, the initial wealth related quantity $\tilde{w}$ factors out form both the numerator and denominator of (3.17), so the equilibrium price f is independent of the initial wealth. The expectations in formula (3.17) are simple one-dimensional integrals, *i.e.*,

$$f = \frac{\displaystyle\int_{\ln(k/s)}^{\infty} \Big[s\exp(Y) - k \Big] \, \exp\left[-\gamma W(Y) - D(Y)\right] \, dY}{\displaystyle\int_{-\infty}^{\infty} \exp\left[-\gamma W(Y) - D(Y)\right] \, dY} \tag{3.18}$$

where

$$W(Y) \;:=\; w(T) = n_0 s\exp(Y) + n_1 \max(0,\, s\exp(Y) - k) \tag{3.19}$$

$$D(Y) \;:=\; \frac{1}{2\sigma^2 T}(Y - \nu T + \frac{1}{2}\sigma^2 T)^2 \tag{3.20}$$

I could not simplify (3.18) further to any known analytical expression. But with all the parameters known, the integrals, and hence the equilibrium price, can be calculated numerically with high precision.[5]

I emphasize that it is rare that an explicit integral expression for the equilibrium price can be found in an incomplete market model, such as the example in this section, because the probability density distribution for $w(T)$ is in general not known. For example, if the stock pays a continuous dividend yield $\tilde{r}$, then there is an extra term, $\tilde{r}n_0 \int_0^T s(t)\,dt$, in expression (3.14) for $w(T)$. To do the quadrature, I must know the joint density distribution of $s(T)$ and $\int_0^T s(t)\,dt$ for a geometric Brownian motion, which does not have a simple analytic expression. Otherwise arithmetic Asian options for a geometric Brownian motion in the BS framework would have simple analytical formulas.[6] Therefore the static perspective for incomplete market problems usually leads to dead ends.

3.4 Restricted Stock: Dynamic Perspective

A different perspective is adopted in analyzing the same problem outlined in the previous section. I once again follow the standard procedure of the stochastic control theory. Since the stock is restricted, the number of shares n_0 in the portfolio is a parameter, not a control variable. There is only one control variable here, namely n_1.

The notation used here is the same as the one in Section 3.2 on the dynamic derivation of the BS equation. The early part of the current derivation mirrors that of the BS equation. The dynamic equations for ds and df are given by (3.1) and (3.2), respectively. The budget equation for dw is (cf. (3.4))

$$\begin{aligned} dw &= n_0\, ds + n_1\, df + \tilde{r} n_0 s\, dt \\ &= \Big[n_0 C_{(s,t)} + \tilde{r} n_0 s + n_1 C_{(f,t)} \Big]\, dt + \Big[n_0 C_{(s,s)} + n_1 C_{(f,s)} \Big]\, dB^s \end{aligned} \tag{3.21}$$

where the shorthand notation $C_{(*,*)}$ is defined by (3.3).

[5] Notice that the condition $n_0 + n_1 \geq 0$ must be satisfied in order for the integrals to converge. This simply means that a buy-and-hold investor who uses the exponential utility function never shorts a lognormally distributed stock.

[6] See the papers by Lewis [60] and by Linetsky [61] and references therein for various formulas of arithmetic Asian options, which involve integrals of special functions.

Once again by assuming the current position being optimal, you try to find out what the necessary conditions for optimality will lead to. The value function J for this problem is defined as

$$J(t, w, s) := \sup_{n_1} E[U(w(T))] \tag{3.22}$$

The HJB equation, which is the necessary condition for optimality, is

$$\sup_{n_1} \mathcal{L}J = 0 \tag{3.23}$$

with $\mathcal{L}J$ being defined by (3.7). The expression $\sup_{n_1} \mathcal{L}J$ means $\partial(\mathcal{L}J)/\partial n_1 = 0$, which leads to (cf. (3.11))

$$C_{(f,t)} J_w + \left[n_0 C_{(s,s)} + n_1 C_{(f,s)} \right] C_{(f,s)} J_{ww} + C_{(s,s)} C_{(f,s)} J_{sw} = 0 \tag{3.24}$$

where I have used definitions (3.7) for $\mathcal{L}J$ and (3.21) for $C_{(w,t)}$ and $C_{(w,s)}$.

This is where the current derivation and the one for the BS equation in Section 3.2 start to differ. Since n_0 is restricted, there is no counterpart to equation (3.8). To proceed, the utility function must be specified. For the exponential utility function, the w dependency of the value function J can be factored out into the following form,

$$J(t, w, s) = -\frac{1}{\gamma} \exp(-\gamma w) \exp(\, -\gamma \phi(t, s) + \gamma n_1 f(t, s) \,) \tag{3.25}$$

which also serves as the definition for $\phi(t, s)$.

Substituting the coefficients $C_{(*,*)}$ defined by (3.1) and (3.2), and expression (3.25) into equation (3.24) leads to the following PDE on $f(t, s)$

$$f_t + (\nu - \gamma n_0 \sigma^2 s - \gamma \sigma^2 s \phi_s) s f_s + \frac{1}{2} \sigma^2 s^2 f_{ss} = 0 \tag{3.26}$$

The final condition for f is simply the payoff function $f(T, s(T))$.

Obviously you cannot solve PDE (3.26) without knowing what ϕ is. Substituting (3.25) into the HJB equation (3.23) leads to the equation for $\phi(t, s)$,

$$\begin{aligned} \phi_t - \tilde{r} s \phi_s + \frac{1}{2} \sigma^2 s^2 \phi_{ss} + \frac{1}{2\gamma} \left(\frac{\nu + \tilde{r}}{\sigma} \right)^2 & \\ - \frac{1}{2} \gamma \sigma^2 \left(s n_0 + s \phi_s - \frac{\nu + \tilde{r}}{\gamma \sigma^2} \right)^2 &= 0 \end{aligned} \tag{3.27}$$

The final condition for ϕ is

$$\phi(T, s(T)) = n_1 f(T, s(T)) \tag{3.28}$$

It is clear from equation (3.27) and final condition (3.28) that if $n_0 = n_1 = 0$ (empty portfolio), then $\phi(t, s) = 0$. Therefore the quantity ϕ is called the position-dependent-term.

To summarize, under the dynamic perspective, PDE (3.27) with final condition (3.28) for the position-dependent-term $\phi(t, s)$ is solved first; then PDE (3.26) with the payoff function is used to obtain the equilibrium price $f(t, s)$.

Despite that PDE (3.26) for f is linear for a given ϕ, the dependency of f on the final payoff function $f(T, s(T))$ is nonlinear, because the position-dependent-term ϕ depends on the payoff function as well. The important point is that the option's equilibrium price f depends on the number of options n_1 in the portfolio. This is because f depends on ϕ, and ϕ is dependent on n_1 through final condition (3.28). Hence, as mentioned before, an optimal trading strategy in an incomplete market establishes a link between an option's position and its equilibrium price. Having position-dependent option pricing equations is the key to solve the optimal trading size problem.

I now make several remarks. The first is to elaborate on the connection between the restricted and unrestricted cases. When there is no stock trading restriction, the stock holding n_0 takes on its optimal value $\hat{n}_0$ (cf. (3.9)). Substituting (3.25) into (3.9) immediately leads to

$$\hat{n}_0 = \frac{\nu + \tilde{r}}{\gamma \sigma^2 s} - \phi_s \tag{3.29}$$

Now replacing n_0 in (3.26) with expression (3.29) for $\hat{n}_0$ recovers the BS equation (3.13).

The second remark is on definition (3.25) for the position-dependent-term ϕ. You may wonder why I do not define the position-dependent-term as $\Phi(t, s) = \phi(t, s) - n_1 f(t, s)$, which seems to be a natural thing to do. In fact the PDEs I initially derived were for f and Φ. The problem of using Φ, instead of ϕ, is that the PDE for Φ depends on f. Therefore the PDEs for f and Φ are mutually dependent. I later discovered that the PDE for ϕ is independent of f, and hence can be solved separately, bearing in mind that the payoff function in the final condition for ϕ is a known function. Therefore by using ϕ instead of Φ, the mutual coupling becomes a one-way influence. Financial interpretation on expression (3.25) will be discussed in the next chapter (around expression (4.15)).

The third remark is that although PDE (3.27) for ϕ is nonlinear, a change of variable

$$\psi(t, s) := \exp(-\gamma n_0 s - \gamma \phi(t, s)) \tag{3.30}$$

leads to the following linear PDE for $\psi(t, s)$

$$\psi_t + \nu s \psi_s + \frac{1}{2}\sigma^2 s^2 \psi_{ss} - \gamma n_0 \tilde{r} s \psi = 0 \tag{3.31}$$

The final condition (3.28) on ϕ gives the corresponding one on ψ

$$\psi(T, s(T)) = \exp(-\gamma n_0 s(T) - \gamma n_1 f(T, s(T))) \tag{3.32}$$

Equivalence to the Static Perspective

So far there are two different ways of obtaining the option equilibrium price associated with the restricted stock model: (i) the static perspective through expression (3.18); (ii) the dynamic perspective by solving PDE (3.27) for ϕ and then PDE (3.26) for f (with $\tilde{r} = 0$). Will these two different ways of computing the equilibrium price give the same answer? The static-dynamic equivalence principle of Section 3.1 says that the results should agree; I am about to explicitly demonstrate that indeed they do agree.

The proof consists of two parts.

The first part is to establish the following relation, which is called the tangent relation

$$\phi'(t,s) = f(t,s) \tag{3.33}$$

where ϕ' is defined to be the first order derivative of ϕ with respect to the option position n_1, as ϕ is dependent on n_1 through the final condition (3.28). The proof of (3.33) is straightforward. Taking derivative with respect to n_1 on equation (3.27) and final condition (3.28) for ϕ gives the PDE and the final condition on ϕ', which are the same as those of f. Therefore ϕ' and f are the same.

The second part uses the famous Feynman-Kac formula, which is stated below without the proof.[7] The Feynman-Kac formula links the solution of a parabolic PDE with the average of a diffusion process. The version I am about to present is for a one-dimensional diffusion, but the multi-dimensional generalization is straightforward. Specifically, if $g(t,x)$ satisfies the following parabolic PDE

$$\frac{\partial g}{\partial t} + \frac{1}{2}a^2(x)\frac{\partial^2 g}{\partial x^2} + b(x)\frac{\partial g}{\partial x} + c(x)g + d(t,x) = 0 \tag{3.34}$$

with the final condition $g(T,x) = G(x)$, then the solution $g(t,x)$ for $t < T$ can be written as

$$\begin{aligned} g(t,x) = E_{t,x}\Bigg[&G(X(T))\exp\left(\int_t^T c(X(u))du\right) \\ &+ \int_t^T d(l,X(l))\exp\left(\int_t^l c(X(u))du\right)dl\Bigg] \end{aligned} \tag{3.35}$$

where the expectation $E_{t,x}[*]$ is taken over the following Ito diffusion process

$$dX(u) = b(X(u))\,du + a(X(u))\,dB_u, \tag{3.36}$$

with the initial condition $X(t) = x$. The Feynman-Kac formula will be used *repeatedly* later in this book.

Now apply the Feynman-Kac formula to equation (3.31) for ψ with $\tilde{r} = 0$, where the Ito diffusion for the average is simply the geometric Brownian motion for the stock (cf. (3.1)). The result is

$$\psi(t,s) = E\left[\exp(-\gamma w(T))\right] \tag{3.37}$$

where final condition (3.32) for ψ has been used, and $w(T)$ is defined as

$$w(T) := n_0 s(T) + n_1 f(T, s(T)) \tag{3.38}$$

Taking derivative with respect to n_1 on definition (3.30) for ψ gives

$$\phi'(t,s) = -\frac{1}{\gamma}\frac{\psi'(t,s)}{\psi(t,s)} = \frac{E[f(T,s(T))\exp(-\gamma w(T))]}{E[\exp(-\gamma w(T))]} \tag{3.39}$$

where expression (3.37) for ψ has been used in the second step.

The left-hand side of (3.39) is f due to the tangent relation (3.33), and the right-hand side of (3.39) is the same as the right-hand side of (3.17) with the utility function $U(w)$ being the exponential function.

[7] See Oksendal [75] or Appendix 1.2 of Lewis [59] for the proof.

Therefore I have proved, at least for the exponential utility function, that the equilibrium price f under the dynamic perspective is the same as the one under the static perspective.

I emphasize that the dynamic perspective is much more powerful than the static one. The static approach often leads to dead ends, because the final wealth distribution is not known *a priori*, *e.g.*, the $\tilde{r} \neq 0$ case of the restricted stock model mentioned at the end of the previous section. On the other hand, the dynamic perspective produces a pair of PDEs, which can always be solved numerically.

Decomposition of ϕ

The position-dependent-term ϕ can be decomposed into the one caused by the underlying stock alone ($n_1 = 0$), plus a remainder,

$$\phi(t,s) = \bar{\phi}(t,s) + h(t,s) \tag{3.40}$$

Since by definition $\bar{\phi}$ is ϕ when $n_1 = 0$, it satisfies the same PDE as the one for ϕ (cf. (3.27)),

$$\begin{aligned} \bar{\phi}_t - \tilde{r}s\bar{\phi}_s + \frac{1}{2}\sigma^2 s^2 \bar{\phi}_{ss} + \frac{1}{2\gamma}\left(\frac{\nu+\tilde{r}}{\sigma}\right)^2 \\ - \frac{1}{2}\gamma\sigma^2\left(sn_0 + s\bar{\phi}_s - \frac{\nu+\tilde{r}}{\gamma\sigma^2}\right)^2 = 0 \end{aligned} \tag{3.41}$$

but with a different final condition for $\bar{\phi}$

$$\bar{\phi}(T, s(T)) = 0 \tag{3.42}$$

Notice that $\bar{\phi} = 0$ if $n_0 = 0$, so $\bar{\phi}$ is caused by taking a position in the underlying stock. Also note that if there is no restriction on the stock holding, then n_0 takes on its optimal value $\hat{n}_0$ given by (3.29), which makes the last term on the left-hand side of (3.41) vanish. Hence equation (2.22) for the pure stock investment problem is recovered.

The nonlinear PDE for the remainder term h is[8]

$$h_t + (\nu - \gamma n_0\sigma^2 s - \gamma\sigma^2 s\bar{\phi}_s - \frac{1}{2}\gamma\sigma^2 s h_s)sh_s + \frac{1}{2}\sigma^2 s^2 h_{ss} = 0 \tag{3.43}$$

with the final condition

$$h(T, s(T)) = n_1 f(T, s(T)) \tag{3.44}$$

It is clear that the solution h for the homogeneous equation (3.43) is zero if the final condition vanishes ($n_1 = 0$), which implies that h is nonzero only when there are options in the portfolio.

To summarize, the position-dependent-term ϕ can be decomposed into two parts, each with a clear financial interpretation. The first part represents the effect of taking a position in the underlying stock alone; and the second part represents the influence of existing option positions in the portfolio. The quantity h has an additional financial meaning, namely it is the portfolio-indifference-price to be discussed in the next chapter.

[8] A somewhat related PDE is derived in the paper by Musiela and Zariphopoulou [73].

Instead of solving ϕ in one step, you can now solve for $\bar{\phi}$ first, which is independent of the option position; you then solve for the portfolio-indifference-price h, which in general depends on $\bar{\phi}$. Finally the equilibrium price f can be obtained by using the tangent relation

$$f(t,s) = h'(t,s) \tag{3.45}$$

which is the result of (3.33) and (3.40), and that $\bar{\phi}$ is independent of n_1.

3.5 DOPE Cooking Recipe

The result of the dynamic derivation is a pair of PDEs, which can be solved, at least numerically, to obtain an option's fair value. This pair of PDEs deserve a special name—DOPE, which stands for dynamic option pricing equations. The word "dynamic" here has two-fold meanings: the first being that the PDEs come from dynamic derivation, and the second being that the PDEs are used to dynamically manage your option positions, because option positions and valuations are linked by the PDEs.

The BS equation, which is obtained from the perfect replication argument, only applies to complete market models. DOPE, which is the result of the dynamic derivation, does not care whether a model market is complete or not. If the underlying model market is complete, then DOPE reduces to the famous BS equation; but for incomplete market models, DOPE offers a natural way to uniquely price an option. Therefore the dynamic derivation methodology can be regarded as an extension of the perfect replication pricing methodology, and DOPE can be viewed as a natural generalization of the BS option pricing equation.

The importance of the dynamic derivation methodology cannot be overstated. Because any small perturbation of a complete market model in general breaks its completeness, thus there are a lot more incomplete market models than complete ones.[9] Hence a systematic option pricing methodology for incomplete market models is a giant step forward comparing to the one that only applies to complete market models.

Different market models produce different types of DOPE, I have already shown you the one for the restricted stock model in this chapter, which is called the RS-DOPE, (cf. (3.26) and (3.27)). I will show you others later in this book, like the one for the stochastic volatility model (SV-DOPE), and the one for the short rate interest model (SR-DOPE). Despite that DOPEs corresponding to different models are not exactly the same, they share many common features, since they are all derived from the same dynamic derivation methodology. I now outline the major steps of obtaining DOPEs:

- choose a model for the state variables and their dynamical equations, *e.g.*, equation (3.1) for ds;
- identify the control variables, which usually are the position n_0 of the underlying asset (assuming only one risk factor), and the N European option positions n_i ($i = 1, 2, \ldots, N$) in the portfolio;
- apply Ito's lemma to the equilibrium price f^i (cf. (3.2));

[9] Using an analogy of real numbers, complete market models are like integers, they are few and far in between; but incomplete market models are like irrational numbers, they are everywhere.

- write down the budget equation, *i.e.*, the equation for dw (cf. (3.4));
- use the HJB equation (cf. (3.6)), which is the necessary conditions of optimality based on the current position;
- set $\partial(HJB)/\partial n_0$ to zero, obviously if the risk factor is nontradable ($n_0 \equiv 0$) or is restricted, then this step does not apply;
- set $\partial(HJB)/\partial n_i$ to zero (cf. (3.24));
- factor the value function $J(t, w, *)$ into

$$-\frac{1}{\gamma}\exp(-\gamma w)\exp\left(-\gamma\phi + \gamma\sum_{i=1}^{N} n_i f^i\right) \tag{3.46}$$

  where $f^i(t, *)$ is the equilibrium price of the ith option, and $\phi(t, *)$ is called the position-dependent-term (cf. (3.25));
- substituting J into $\partial(HJB)/\partial n_0 = 0$ gives the optimal hedging relation (cf. (3.29));
- substituting J into $\partial(HJB)/\partial n_i = 0$ leads to the PDE for f^i (cf. (3.26)), which involves ϕ; notice that all options have the same pricing equation, but with different final payoff functions;
- reduce the HJB equation to a PDE for ϕ, which does not involve f^i (cf. (3.27));
- decompose the position-dependent-term ϕ further based on the influences of various positions (cf. (3.40)), and obtain the PDEs and final conditions for the corresponding parts (cf. (3.41) and (3.43)).

Do not worry if you are unclear about the list, as these steps will be illustrated in detail when I derive the SV-DOPE later in this book.

As mentioned earlier that the optimal trading size problem can be solved naturally when combining risk aversion and market incompleteness. It is now clear why. Risk aversion is modeled by a concave utility function. The expected utility maximization naturally leads to the dynamic derivation methodology, which is a no-questions-asked, just-do-it mathematical procedure of the stochastic control theory. The result of the dynamic derivation is a pair of PDEs called DOPE. In incomplete markets, DOPE establishes a link between an option's equilibrium price and the underlying option portfolio, which is the key to solve the optimal trading size problem. As you can see that each step is a natural consequence of the previous one.

It seems to me that much of the continuous-time finance literature is moving along two separate tracks, namely (i) portfolio optimizations by expected utility maximizations, and (ii) derivatives pricing by the perfect replication argument.[10] The dynamic derivation methodology has merged these two tracks together, by showing that options should be priced in the context of portfolio optimizations, which establishes a link between option values and the underlying option portfolio.

[10] Both the dynamic programming approach to portfolio optimizations and the perfect replication argument approach to derivatives pricing in continuous-time settings were pioneered by Robert Merton in the late sixties and early seventies.

Epilog

I have preached a new paradigm of option pricing in this book, which moves away from the traditional perfect replication argument. Instead it adopts the approach of pricing options in the context of portfolio optimization. The latter approach is equivalent to the former in complete markets, but is much more useful in incomplete markets. Notice that the new approach in incomplete markets makes option values position dependent, which is meaningful only if you take the personal rather than the market perspective.

The central tenet of this book is that an optimal trading strategy establishes a link between the fair value of a risky asset and its position (together with the positions of other related risky assets) in the portfolio. For models governed by diffusion processes, this idea leads to the new dynamic derivation procedure that produces a pair of PDEs called DOPE. Solutions of various types of DOPE provide a theoretically consistent way of making quantitative trading decisions by adjusting the position to satisfy the local equilibrium equation *fair value* = *market price*.

Incomplete markets are just about anywhere and everywhere. Why stop at the topics mentioned in this book? Clearly the new methodology can be applied to other areas of quantitative finance such as commodities and insurance mathematics. In addition, theories for non-European options including many exotic derivatives are waiting for development.

Are you excited? Do you feel the rush? That is what using DOPE will do to you. Maybe The Street will be filled with DOPE addicts in the not-too-distant future. Perhaps this is just the beginning of a new era for derivatives pricing and trading...

The end?

Bibliography

[1] Lucia Albarran Aguilar and Wim Schoutens. Do you know how much you pay? *Wilmott Magazine*, pages 48–56, July-August 2005.

[2] Dong-Hyun Ahn and Bin Gao. A parametric nonlinear model of term structure dynamics. *Review of Financial Studies*, 12(4):721–762, 1999.

[3] Yacine Aït-Sahalia, Yubo Wang, and Francis Yared. Do option markets correctly price the probabilities of movement of the underlying asset? *Journal of Econometrics*, 102:67–110, 2001.

[4] Erling D. Andersen and Anders Damgaard. Utility based option pricing with proportional transaction costs and diversification problems: an interior-point optimization approach. *Applied Numerical Mathematics*, 29:395–422, 1999.

[5] Leif B. G. Andersen and Vladimir Piterbarg. Moment explosions in stochastic volatility models. preprint, 2005.

[6] Jesper Andreasen, Bjarke Jensen, and Rolf Poulsen. Eight valuation methods in financial mathematics: The black-scholes formula as an example. *Mathematical Scientist*, 23(1):18–40, 1998.

[7] Thierry Ané and Hélyette Geman. Order flow, transaction clock and normality of asset returns. *Journal of Finance*, 55(5):2259–2284, 2000.

[8] C. Atkinson and S. Mokkhavesa. Intertemporal portfolio optimization with small transaction costs and stochastic variance. *Applied Mathematical Finance*, 10:267–302, 2003.

[9] Marco avellaneda and Michael Lipkin. A market-induced mechanism for stock pinning. *Quantitative Finance*, 3:417–425, 2003.

[10] Gurdip Bakshi, Charles Cao, and Zhiwu Chen. Empirical performance of alternative option pricing models. *Journal of Finance*, 52(5):2003–2049, 1997.

[11] Carl M. Bender and Steven A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers*. Springer, 1999.

[12] Fred Espen Benth and Kenneth Hvistendahl Karlsen. A PDE representation of the density of the minimal entropy martingale measure in stochastic volatility markets. *Stochastics: An International Journal of Probability and Stochastic Processes*, 77(2):109–137, 2005.

[13] Yaacov Bergman. A characterization of self-financing portfolio strategies. http://www.haas.berkeley.edu/finance/WP/rpf113.pdf, 1981.

[14] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–659, May-June 1973.

[15] Nicolas P. B. Bollen and Robert E. Whaley. Does net buying pressure affect the shape of implied volatility functions? *Journal of Finance*, 59(2):711–753, 2004.

[16] Andrei Borodin and Paavo Salminen. *Handbook of Brownian Motion: Facts and Formulae*. Birkhauser, Boston, 2nd edition, 2002.

[17] Alberto Bressan. Viscosity solutions of Hamilton-Jacobi equations and optimal control problems: an illustrated tutorial. www.sissa.it/~bressan/hj.pdf, 2001.

[18] Sid Browne. Beating a moving target: Optimal portfolio strategies for outperforming a stochastic benchmark. *Finance and Stochastics*, 3:275–294, 1999.

[19] Sid Browne. Risk-constraint dynamic active portfolio management. *Management Science*, 46(9):1188–1199, 2000.

[20] Peter Carr, Xing Jin, and Dilip Madan. Optimal investment in derivative securities. *Finance and Stochastics*, 5:33–59, 2001.

[21] Peter Carr and Dilip Madan. Optimal positioning in derivative securities. *Quantitative Finance*, 1:19–37, 2001.

[22] Peter K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1):135–155, 1973.

[23] Les Clewlow and Stewart Hodges. Optimal delta-hedging under transaction costs. *Journal of Economic Dynamics and Control*, 21:1353–1376, 1997.

[24] Rama Cont. Modelling term structure dynamics: An infinite dimensional approach. *International Journal of Theoretical and Applied Finance*, 8(3):357–380, 2005.

[25] John C. Cox, Jonathan E. Ingersoll, Jr., and Stephen A. Ross. An intertemporal general equilibrium model of asset prices. *Econometrica*, 53:363–384, 1985.

[26] J. Cvitanic and I. Karatzas. On portfolio optimization under drawdown constraints. *IMA Volumes in Mathematics and its Application*, 65:35–46, 1995.

[27] R. A. Jarrow D. Heath and A. Morton. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, 60:77C105, 1992.

[28] Anders Damgaard. Utility based option evaluation with transaction costs. *Journal of Economic Dynamics and Control*, 27:667–700, 2003.

[29] M. H. A. Davis. Option pricing in incomplete markets. In Michael A. H. Dempster and Stanley R. Pliska, editors, *Mathematics of Derivative Securities*, pages 216–226. Cambridge University Press, Cambridge, UK, 1997.

[30] M. H. A. Davis. Complete-market models of stochastic volatility. *Proc. R. Soc. Lond. A*, 460:11–26, 2004.

[31] M. H. A. Davis, V. G. Panas, and T. Zariphopoulou. European option pricing with transaction costs. *SIAM Journal of Control and Optimization*, 31:470–493, 1993.

[32] Mark H.A. Davis and David G. Hobson. The range of traded option prices. preprint, 2004.

[33] Kresimir Demeterfi, Emanuel Derman, Michael Kamal, and Joseph Zou. A guide to volatility and variance swaps. *Journal of Derivatives*, 6:9–32, Summer 1999.

[34] Bernard Dumas. Super contact and related optimality conditions. *Journal of Economic Dynamics and Control*, 15:675–685, 1991.

[35] Nicolae Gârleanu, Lasse Heje Pedersen, and Allen M. Poteshman. Demand-based option pricing. preprint, 2005.

[36] Viatcheslav Gorovoi and Vadim Linetsky. Black's model of interest rates as options, eigenfunction expansions and japanese interest rates. *Mathematical Finance*, 14(1):49–78, 2004.

[37] M. R. Grasselli and T. R. Hurd. A monte carlo method for exponential hedging of contingent claims. preprint, 2003.

[38] M. R. Grasselli and T. R. Hurd. Indifference pricing and hedging in stochastic volatility models. preprint, 2004.

[39] S. J. Grossman and Z. Zhou. Optimal investment strategies for controlling drawdowns. *Mathematical Finance*, 3(3):241–276, 1993.

[40] Martin B. Haugh and Andrew W. Lo. Asset allocation and derivatives. *Quantitative Finance*, 1:45–72, 2001.

[41] Vicky Henderson. Black-scholes model. In Josef L. Teugels and Bjørn Sundt, editors, *Encyclopedia Of Actuarial Science*, volume 1, pages 175–184. John Wiley & Sons, Chichester, UK, 2004.

[42] Vicky Henderson and David Hobson. Utility indifference pricing–an overview. forthcoming.

[43] Steven Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343, 1993.

[44] Alan F. Ho. Optimal trading strategy for european options with transaction costs. *Advances in Mathematics*, 177:1–65, 2003.

[45] David Hobson. Stochastic volatility. In D. Hand and S. Jacka, editors, *Statistics in Finance*, pages 283–306. Arnold Publishers, London, UK, 1998.

[46] David G. Hobson and L. C. G. Rogers. Complete models with stochastic volatility. *Mathematical Finance*, 8(1):27–48, 1998.

[47] Stewart D. Hodges and Anthony Neuberger. Optimal replication of contingent claims under transaction costs. *Review of Futures Markets*, 8(2):222–239, 1989.

[48] Chi-fu Huang and Robert H. Litzenberger. *Foundations for Financial Economics*. Princeton-Hall, Englewood Cliffs, New Jersey, 1988.

[49] Aytac İlhan, Mattias Jonsson, and Ronnie Sircar. Portfolio optimization with derivatives and indifference pricing. preprint, 2004.

[50] Aytac İlhan, Mattias Jonsson, and Ronnie Sircar. Optimal investment with derivative securities. *Finance and Stochastics*, 9:585–595, 2005.

[51] Karel Janeček and Steven E. Shreve. Asymptotic analysis for optimal investment and consumption with transaction costs. *Finance and Stochastics*, 8:181–206, 2004.

[52] Alireza Javaheri. *Inside Volatility Arbitrage*. John Wiley & Sons, New Jersey, 2005.

[53] Ioannis Karatzas. *Lectures on the Mathematics of Finance*. American Mathematical Society, Providence, RI, 1997.

[54] Tong Suk Kim and Edward Omberg. Dynamic nonmyopic portfolio behavior. *Review of Financial Studies*, 9(1):141–161, 1996.

[55] Robert Kohn and Oana Papazoglu. On the equivalence of the static and dynamic asset allocation problems. *Quantitative Finance*, 6(2):173–183, 2006.

[56] Ralf Korn. The martingale optimality principle: The best you can is good enough. *Wilmott Magazine*, pages 61–67, July-August 2003.

[57] Ralf Korn and Holger Kraft. On the stability of continuous-time portfolio problems with stochastic opportunity set. *Mathematical Finance*, 14(3):403–414, 2004.

[58] H. E. Leland. Option pricing and replication with transaction costs. *Journal of Finance*, 40:1283–1301, 1985.

[59] Alan L. Lewis. *Option Valuation under Stochastic Volatility.* Financial Press, Newport Beach, CA, 2000.

[60] Alan L. Lewis. Asian connections. *Wilmott Magazine*, pages 57–63, September 2002.

[61] Vadim Linetsky. Spectral expansions for asian (average price) options. *Operations Research*, 52:856–867, 2004.

[62] Vadim Linetsky. The spectral representation of bessel processes with constant drift: Applications in queueing and finance. *Journal of Applied Probability*, 41(2):327–344, 2004.

[63] Jun Liu and Francis A. Longstaff. Losing money on arbitrage: Optimal dynamic portfolio choice in markets with arbitrage opportunities. *Review of Financial Studies*, 17(3):611–641, 2003.

[64] Jun Liu and Jun Pan. Dynamic derivative strategies. *Journal of Financial Economics*, 69:401–430, 2003.

[65] L. MacLean and W. T. Ziemba. Growth versus security tradeoffs in dynamic investment analysis. *Annals of Operations Research*, 85:193–225, 1999.

[66] Leonard. MacLean, William. T. Ziemba, and Yuming Li. Time to wealth goals in capital accumulation. *Quantitative Finance*, 5:343–355, 2005.

[67] Pascal J. Maenhout. Robust portfolio rules and asset pricing. *Review of Financial Studies*, 17(4):951–983, 2004.

[68] Malik Magdon-Ismail and Amir Atiya. Maximum drawdown. *Risk Magazine*, 17:99–102, October 2004.

[69] Sergei Maslov and Yi-Cheng Zhang. Optimal investment strategy for risky assets. *International Journal of Theoretical and Applied Finance*, 1(3):377–387, 1998.

[70] Robert C. Merton. *Continuous Time Finance*. Basil Blackwell, Cambridge, MA, 1992.

[71] S. Mokkhavesa and C. Atkinson. Perturbation solution of optimal portfolio theory with transaction costs for any utility function. *IMA Journal of Management Mathematics*, 13:131–151, 2002.

[72] Michael Monoyios. Option pricing with transaction costs using a markov chain approximation. *Journal of Economic Dynamics and Control*, 28:889–913, 2004.

[73] Marek Musiela and Thaleia Zariphopoulou. An example of indifference prices under exponential preferences. *Finance and Stochastics*, 8:229–239, 2004.

[74] Anthony Neuberger. The log contract. *Journal of Portfolio Management*, 20(2):74–80, 1994.

[75] Bernt Oksendal. *Stochastic Differential Equations: an introduction with applications.* Springer, New York, 6th edition, 2003.

[76] Hervé Roche. Optimal consumption and investment under a drawdown constraint. preprint, 2006.

[77] L. C. G. Rogers. The relaxed investor and parameter uncertainty. *Finance and Stochastics*, 5:131–154, 2001.

[78] L. C. G. Rogers. Why is the effect of proportional transaction costs $O(\delta^{2/3})$? In G. Yin and Q. Zhang, editors, *Mathematics of Finance, AMS Contemporary Mathematics Series*, pages 303–308. American Mathematical Society, 2004.

[79] L. C. G. Rogers and Surbjeet Singh. Modelling liquidity and its effects on price. preprint, 2006.

[80] Martin Schweizer and Johannes Wissel. Term structure of implied volatilities: Absence of arbitrage and existence results. preprint, 2005.

[81] Neil Shephard. Stochastic volatility. In Steven Durlauf and Lawrence Blume, editors, *New Palgrave Dictionary of Economics*. Palgrave Macmillan, 2nd edition, 2006. forthcoming.

[82] Ronnie Sircar and Thaleia Zariphopoulou. Bounds & asymptotic approximations for utility prices when volatility is random. *SIAM Journal of Control and Optimization*, 43(4):1328–1353, 2005.

[83] Sasha Stoikov. Pricing options from the point of view of a trader. preprint, 2005.

[84] Edward Thorp. The kelly criterion in blackjack, sports betting, and the stock market. www.bjmath.com/bjmath/thorp/paper.htm, 1997.

[85] Robert Tompkins. Stock index futures markets: Stochastic volatility models and smiles. *Journal of Futures Markets*, 21(1):43–78, 2001.

[86] Scott Weiner. Should stochastic volatility matter to the cost-constrained investor? *Mathematical Finance*, 14(1):131–139, 2004.

[87] A. E. Whalley and P. Wilmott. An asymptotic analysis of an optimal hedging model for option pricing with transaction costs. *Mathematical Finance*, 7(3):307–324, 1997.

[88] A. E. Whalley and P. Wilmott. Optimal hedging of options with small but arbitrary transaction cost structure. *Euro. Jnl of Applied Mathematics*, 10:117–139, 1999.

[89] Paul Wilmott. *Derivatives*. John Wiley & Sons, New York, 1998.

[90] Valeri Zakamouline. European option pricing and hedging with both fixed and proportional transaction costs. *Journal of Economic Dynamics and Control*, 30:1–25, 2006.

[91] Thaleia Zariphopoulou. A solution approach to valuation with unhedgable risks. *Finance and Stochastics*, 5:61–82, 2001.