Applications of Optimization to Mathematical Finance

by

David Saunders

A thesis submitted in conformance with the requirements of the degree of Master's of Science, Graduate Department of Mathematics, University of Toronto.

©Copyright by David Saunders 1997
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.
Applications of Optimization to Mathematical Finance

Master’s of Science

David Saunders

Department of Mathematics

University of Toronto

Abstract

A number of issues of interest in modern mathematical finance are examined, including arbitrage pricing, portfolio selection, and implied distributions. Attention is focused on the use of optimization theory in solving problems of interest to both practitioners and academics. Various topics of current research are surveyed, including a new technique, using a portfolio selection model to estimate implied distributions.
0.1 Preface: Risk and Reward

A great deal of the research into the mathematics of derivative securities has been motivated by the need to formalize the conventional wisdom of market practitioners. Mathematical models enhance our understanding of colloquialisms such as "there's no such thing as a free lunch", both by reflecting these intuitions in the structure, and by providing tools to analyze their empirical consequences.

The basic intuition behind this thesis, and indeed most applications of mathematics to finance, is that all market participants seek to maximize reward, and minimize risk. Furthermore, these goals are contradictory. Financial market participants are subject to an inherent tradeoff between the risk they assume and the reward that they can obtain. A great deal of research has been devoted to characterizing and studying the concepts of risk and reward in a mathematical framework.

The thesis is divided into three chapters. The first presents the modern theory of arbitrage pricing. In essence, this describes mathematically how the principles contained in the previous paragraph determine the prices of market securities. The fundamental result is that under minimal economic assumptions, the presence of a tradeoff between risk and reward is enough to determine all market prices. The second chapter studies the portfolio selection problem. There prices are taken as given, and the problem is to determine the optimal investments that should be made by a given market participant. A new technique for portfolio selection, originated by Dembo [19] and referred to as Scenario Optimization is presented, and developed in a general setting for the first time. The third chapter turns the pricing problem around, and studies the inverse problem of determining information about the probability distributions of market variables based on observations of the prices of derivative securities. This has important applications to hedging, and pricing over-the-counter instruments (those which are not sold on the open market). A special section in the third chapter examines how duality in optimization theory relates to the solution of this inverse problem.

The computations for this thesis were performed using the RiskWatch program for risk management developed by Algorithmics Inc. I thank Algorithmics for the opportunity to participate in its research group, and to use its
software. Alex Kreinin and Ben DePrisco provided code that aided in the computations. I would like to thank my colleagues at RiskLab, and Algorithmics Inc., who made writing this thesis an interesting and pleasurable experience. In particular, I would like to thank Stathis Tompaidis, Juan Gonzalez, and Prof. Claudio Albanese for many interesting discussions and helpful comments. My supervisory team of Prof. Luis Seco and Dr. Dan Rosen have been a constant source of encouragement and inspiration, and deserve special thanks. All remaining errors in the thesis are my own.

David Saunders
# Contents

0.1 Preface: Risk and Reward ........................................... 2

1 Arbitrage Pricing ................................................. 7
  1.1 Introduction .................................................. 7
  1.2 Pricing Methodologies ......................................... 7
  1.3 The Central Concepts of Arbitrage Pricing ....................... 8
    1.3.1 Arbitrage Opportunities: Free Lunch ..................... 9
    1.3.2 Completeness of Markets ................................ 9
  1.4 A Single Period Model ........................................ 10
  1.5 Multi-Period Model ........................................... 14
  1.6 Continuous Time Model ...................................... 21
    1.6.1 The Black-Scholes Equation ............................ 26

2 Portfolio Optimization ............................................ 28
  2.1 Introduction .................................................. 28
  2.2 Traditional Portfolio Selection Models ......................... 29
    2.2.1 Markowitz’ Mean-Variance Model ....................... 29
    2.2.2 An Application of Stochastic Control Theory to Port-
         folio Selection ........................................ 33
  2.3 Scenario Optimization: A New Technique for Portfolio Selection 37
    2.3.1 Regret: Measuring Risk Against a Benchmark .......... 37
    2.3.2 Scenario Optimization: The General Case .......... 39
    2.3.3 Scenario Optimization Formulation of Portfolio Selec-
          tion .................................................. 43
    2.3.4 Parametric Formulation: The Regret-Reward Model .... 47
    2.3.5 The Regret-Reward Efficient Frontier ............... 48
3 Implied Probabilities in Finance

I The Probabilistic Implications of Option Prices 62
3.1 Introduction ......................................... 63
3.2 The Black-Scholes Model ................................. 63
   3.2.1 Market Deviations from the Black-Scholes Model ... 63
   3.2.2 The Implied Volatility Surface .......................... 64
   3.2.3 Historical Behaviour of the Volatility Smile ........... 65
3.3 Alternative Specifications for the Underlying Process ...... 66
3.4 Evolution of Stochastic Interest Rate Models ............. 68
3.5 Equilibrium Models ...................................... 69
   3.5.1 The Capital Asset Pricing Model ...................... 69
   3.5.2 Mixed Jump-Diffusion Model ............................ 70
   3.5.3 Stochastic Volatility Model ............................ 72
3.6 No-Arbitrage Models ...................................... 74
   3.6.1 The Single Factor Diffusion Setting .................... 74
3.7 Interpolation Techniques .................................. 80
   3.7.1 Shimko’s Method ...................................... 80
   3.7.2 Smiling Trees ......................................... 80
3.8 Optimization Methods ..................................... 87
   3.8.1 Discrepancy from a Prior Distribution ................ 88
   3.8.2 The Maximum Entropy Criterion ......................... 92
3.9 A Statistical Technique: Nonparametric Regression .......... 97
3.10 Use of a Prior Distribution ............................... 100
3.11 Stability ............................................... 101
3.12 Empirical Tests ......................................... 103
3.13 Directions for Future Research ............................ 104

II Implied Scenario Probabilities 105
3.14 Introduction ............................................. 106
3.15 Duality, Reflection and Inversion ........................ 106
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.16 Recapitulation: The Regret-Reward Model</td>
<td>108</td>
</tr>
<tr>
<td>3.16.1 The Regret-Reward Frontier</td>
<td>108</td>
</tr>
<tr>
<td>3.16.2 Notation</td>
<td>108</td>
</tr>
<tr>
<td>3.16.3 Implied Probabilities</td>
<td>109</td>
</tr>
<tr>
<td>3.16.4 Role of the Model Inputs</td>
<td>110</td>
</tr>
<tr>
<td>3.17 Examples</td>
<td>110</td>
</tr>
<tr>
<td>3.17.1 Perfect Markets</td>
<td>110</td>
</tr>
<tr>
<td>3.17.2 Varying the Prior Distribution: Lognormality</td>
<td>117</td>
</tr>
<tr>
<td>3.18 Performance on Market Data</td>
<td>122</td>
</tr>
<tr>
<td>3.18.1 Regarding the Data</td>
<td>124</td>
</tr>
<tr>
<td>3.18.2 Group Constraints</td>
<td>125</td>
</tr>
<tr>
<td>3.19 Concluding Remarks and Directions for Future Research</td>
<td>125</td>
</tr>
</tbody>
</table>
Chapter 1

Arbitrage Pricing

1.1 Introduction

In this chapter, we describe the pricing of financial securities by the arbitrage methodology. We illustrate the principles of this technique with a series of examples, ranging from a simple, single period model of the market to the case where assets obey a multi-factor diffusion process (the famous Black-Scholes model follows as a special case). The more general situation where assets can be arbitrary semimartingales is omitted. The interested reader is referred to the papers of Harrison and Pliska [35] and Delbaen and Schachermeyer [18].

1.2 Pricing Methodologies

There are two main approaches to the pricing of financial instruments. These are commonly referred to as the equilibrium and arbitrage-free methods. In this section, we briefly describe the intuition behind each of these methods, and then explain our preference for the arbitrage-free approach as a technique for pricing financial derivatives. The remainder of the chapter then presents the the arbitrage-free methodology by employing a series of illustrative examples.

Equilibrium methods focus on investor preferences, and assume that the economy tends to gravitate towards a state where all investors have allocated their resources optimally. In any other circumstance, investors with
suboptimal allocations will attempt to improve their positions, thus creating an instability. This instability will only disappear once the economy enters into a steady-state where no market participants are motivated to cause disturbances (i.e. when all investors have achieved optimal wealth allocation). In contrast to the equilibrium approach, the arbitrage-free methodology begins by assuming that the prices of a small number securities are given, and then deduces prices of other instruments by attempting to match their behaviour with these “basic” securities. The main assumption employed is that markets are free from opportunities to earn riskless profits (i.e. arbitrage). This leads to the result that any two portfolios producing identical payoffs under all scenarios should have the same price (otherwise, riskless profits are possible by purchasing the cheaper portfolio and selling the more expensive one).

There are two main advantages possessed by the arbitrage-free pricing methodology over its equilibrium counterpart that account for the preference given it in this chapter (and in the thesis in general). The first is that arbitrage pricing does not require any assumptions regarding investor preferences, aside from the basic axiom that market participants will always prefer more wealth to less (referred to in economics as insatiability). While equilibrium models are suitable for economic theory, in applications to derivative securities pricing it is not often justifiable to assume a specific form of preference function for a given investor. The second advantage of arbitrage pricing is that it provides an explicit algorithm whereby the violation of the arbitrage-free results will lead (at least in theory) to a risk-free profit.

1.3 The Central Concepts of Arbitrage Pricing

In this section, we describe the two fundamental concepts that underly derivative security pricing by arbitrage. The first is the notion of an arbitrage opportunity, i.e. a guarantee to make a profit without risk. The second is the completeness of a market, which relates to whether a certain set of fundamental securities “spans” the space of possible payouts.
1.3.1 Arbitrage Opportunities: Free Lunch

An arbitrage opportunity consists of the ability to earn a profit without assuming any risk. An arbitrage opportunity guarantees that an investor will be able to make a profit independent of the random behaviour of market variables. Since arbitrage offers reward with no risk, it represents a violation of the assumption that reward can only be achieved by incurring risk. It is therefore standard to assume in finance theory that arbitrage opportunities do not exist in markets. This is justified on practical grounds since arbitrage opportunities are so attractive that financial agents move to take advantage of them as soon as they arise (in fact a large amount of money is spent looking for arbitrage violations in order to capitalize on just such a chance). The actions of these market participants, along with the fundamental laws of supply and demand, force the prices into a state where arbitrage is not possible, so the opportunities to make riskless profits vanish as quickly as they appear. Essentially, arbitrage pricing reduces to a simple principle: if two securities yield an identical payoff in the future, and they must have the same current price. Otherwise, arbitrage opportunities would be possible by buying the cheaper security and selling the more expensive one.

Suppose the prices of a small set of securities are observed in the market. Then any security whose future behaviour can be reproduced exactly by a trading strategy involving the observed securities must have the same price as the initial cost of establishing the strategy. Otherwise there would exist two portfolios with identical payoffs, but different prices, and hence an arbitrage opportunity. These intuitive arguments will be formalized in the examples considered below. The above strategy of reproducing the payoff of a security with the observable securities leads directly to the notion of completeness of markets.

1.3.2 Completeness of Markets

In general, a market consists of a set of tradeable instruments whose prices are known, and a set of payoffs. The market is complete if every payoff can be replicated perfectly by holding portfolio consisting of only the tradeable instruments. (This definition allows for the weights of the various securities to change over time, so long as the value of the portfolio is not affected by
such rebalancing). Thus, markets are complete if every payoff is equivalent to (can be attained by trading in) a portfolio of instruments whose current prices are known. As time passes the amount of each tradeable instrument in the portfolio can be altered, so long as no cash enters or leaves the portfolio. The arbitrage methodology states that the price of the payoff must be equal to the cost of setting up the replicating portfolio. Otherwise, an arbitrage opportunity would arise (the uniqueness of the price of any payoff is automatic from the assumption of the absence of arbitrage). Thus, \textit{if markets are complete, and free of arbitrage opportunities, then each payoff has a unique price.}

We now present a series of illustrative examples that highlights the relation between the notions of the absence of arbitrage, completeness of markets, and derivative securities pricing.

\subsection{A Single Period Model}

In this section we describe a simple single period model of financial markets. This model will be important to us for two main reasons. First, it provides much of the intuition that is necessary for more general models of financial markets. Second, many of the optimization models that are discussed in the following chapters are based on analysis in the single period framework.

The market consists of a finite number, \( n \) of securities. The current date is time 0, and all investors are taken to construct a portfolio today and then hold it without rebalancing until a terminal date \( T \). Uncertainty at time \( T \) is defined by a finite number \( S \) of scenarios. The \( i \)th scenario occurs with probability \( p_i \) at time \( T \). A security is uniquely defined by its payout at the terminal date, and its current price. We assume that the \( j \)th security has current price \( q_j \), and pays out the amount \( d_{ij} \) if scenario \( i \) prevails at the terminal date. Thus the market is completely characterized by the price vector \( q \in \mathbb{R}^n \), the payout matrix \( D = \{d_{ij}\} \in \mathbb{R}^{S \times n} \) and the vector \( p \) of scenario probabilities. One of the fundamental results of the theory of derivative securities pricing yields the irrelevance of the probabilities \( p \).

A \textit{portfolio} is a vector \( \theta \in \mathbb{R}^N \). This is to be interpreted as an investor
holding $\theta_j$ of security $j$. The price of constructing the portfolio $\theta$ is

$$q^T \theta$$

(1.1)

and the payout from holding the portfolio $\theta$ is

$$\sum_{j=1}^{N} d_{ij} \theta_j$$

(1.2)

if state $i$ occurs at the terminal date. Therefore, the payoff of the portfolio $\theta$ is given by the vector (random variable)

$$D^T \theta$$

(1.3)

A portfolio $\theta$ is called an arbitrage if either $q^T \theta < 0$ and $D^T \theta > 0$ or $q^T \theta \leq 0$ and $D^T \theta > 0$. Here, all inequalities are to be interpreted componentwise. This corresponds to our intuitive notion of arbitrage as an opportunity to obtain profit without risk. In the first case, an immediate cash inflow is obtained, with no possibility of cash outflow in the future. In the second, a strictly positive future inflow of cash is secured with certainty for zero cost.

A price deflator is a vector $\pi \in \mathbb{R}^S$, such that $\pi > 0$, and

$$D\pi = q$$

(1.4)

Thus, for each of the elementary securities (those available for trade) in our market, its price is given by the sum of its payouts, weighted by the elements of the price deflator. This shall have an important interpretation for us later in terms of expectations with respect to probability measures. For now, we state the fundamental result for the single period case:

**Theorem 1.4.1.** There is no arbitrage if and only if there exists a price deflator.

**Proof** Consider the set $A \subseteq \mathbb{R} \times \mathbb{R}^S$

$$A = \{ (-q^T \theta, D^T \theta) \}$$

(1.5)

and denote by $B$ the positive orthant of $L = \mathbb{R} \times \mathbb{R}^S$ (all vectors with each entry positive). If there is no arbitrage, then $A \cap B = \{0\}$. Since $A$
and $B$ are both closed and convex subsets of $L$, the Hyperplane Separation Theorem (a consequence of the Hahn-Banach theorem, see Luenberger [55]) yields the existence of a linear functional $\phi$ such that

$$\phi(x) < \phi(y)$$

(1.6)

for all $x \in A$, $y \in B$. Since $A$ is a linear subspace, this implies that $\phi(x) = 0$ for all $x \in A$, and $\phi(y) > 0$ for all $y \in B$. Since any linear functional on $L$ is given by an inner product with respect to one of its elements ($L$ is a Hilbert space), and using the strict positivity of $\phi$ on $B$, we have that there must exist $\alpha > 0$, and $\psi \in \mathbb{R}_+^S$ (i.e. the subset of $\mathbb{R}^S$, consisting of all vectors with each component strictly positive) such that

$$\phi(w, z) = \alpha w + \psi^T z$$

(1.7)

for all $(w, z) \in \mathbb{R} \times \mathbb{R}^S$. Thus, since $\phi = 0$ on $A$, we have that for each $\theta \in \mathbb{R}^n$,

$$-\alpha q^T \theta + \psi D^T \theta = 0$$

(1.8)

It follows the $\psi/\alpha$ is a state price vector. On the other hand, if there exists a state price vector, $\pi$, such that $D\pi = q$ then $q^T \theta \leq 0$ implies $\pi^T D^T \theta \leq 0$, whence the strict positivity of $\pi$ implies that $D^T \theta$ cannot be strictly positive. Hence there is no arbitrage. Q.E.D.

We now examine the economic concept of completeness of markets in our single period setting. Let $\psi$ be an arbitrary vector in $\mathbb{R}^S$. This represents a payout structure for a hypothetical security in the market. The fundamental problem in derivative securities pricing is to determine an appropriate price for $\psi$ based on the prices of our fundamental securities. As we shall see, if markets are complete, and there is no arbitrage, then simple economic reasoning allows us to obtain the price of $\psi$ from the vector $q$. The question of completeness amounts to whether or not there exists a portfolio of fundamental securities that has the same payout as $\psi$. Mathematically, this simply asks whether there exists a vector $\theta \in \mathbb{R}^N$ such that

$$D^T \theta = \psi$$

(1.9)
And thus, we see that in the single period model, the market is complete if and only if $D$ has rank $S$. Now suppose that the security with payoff $\psi$ is also available for trade in the market. This amounts to augmenting the payout matrix $D$ to $D^*$, which is the old $D$, with an extra row containing $\psi^T$. Let $c$ be the price of $\psi$. Then by the theorem, if the market is to remain free from arbitrage then we must have $\psi^T \pi = c$. Now suppose that the payoff of $\psi$ can be replicated using only the fundamental securities, i.e. suppose that there exists a portfolio $\theta$ such that $D^T \theta = \psi$. Then

$$c = \psi^T \pi = (D^T \theta)^T \pi = \theta^T (D \pi) = \theta \cdot q$$  \hspace{1cm} (1.10)

That is, under the no-arbitrage assumption, the portfolio $\theta$ has current price $c$. This provides us with a methodology for pricing any claim whose payoff is in the linear span of the fundamental instruments. If the market is complete, this linear span consists of all possible contingent claims.

More intuitively, if the security $\psi$ cost more then the portfolio $\theta$, i.e. if $c > q^T \theta$ then arbitrage profits would be possible by selling $\psi$ at $c$, and purchasing $\theta$ for $q^T \theta$. The total portfolio would have a deterministic payout of 0 at the terminal date, but would provide an initial inflow of cash. Similarly, if $c < q^T \theta$, then the security $\psi$ would be underpriced, and arbitrage profits would be attainable by buying $\psi$ and selling the replicating portfolio. Here we employ the assumption that there are no restrictions on short selling, and that the activities of any single investor do not affect prices.

In short, when markets are complete, we have the following simple algorithm for pricing any random payout $\psi$ by arbitrage:

- Determine the portfolio $\theta$ that exactly replicates the payoff, i.e. obeys the equation $D^T \theta = \psi$.

- Set the price of $\psi$ to be equal to $q^T \theta$.

This algorithm shall become more mathematically complex as the market model obtains a more complicated structure, but the fundamental intuitions of pricing by replication and arbitrage will remain.

From the fundamental theorem presented above, we know that the absence of arbitrage is equivalent to the existence of a price deflator $\pi > 0$ such that:

$$D \pi = q$$  \hspace{1cm} (1.11)
The positivity of $\pi$ provides an interesting interpretation in terms of probability. Consider the vector

$$\alpha = \frac{\pi}{\|\pi\|_1}$$

(1.12)

Then $\alpha > 0$, and $\|\alpha\|_1 = 1$, so $\alpha$ gives a vector of probabilities on the possible future scenarios. The price of any payoff $\psi$ is given by $\pi \cdot \psi$. If we denote by $E_\alpha$ expectation with respect to the probabilities $\alpha$, then we have

$$\pi \cdot \psi = \|\pi\|_1 E_\alpha(\psi)$$

(1.13)

This yields another, equivalent methodology for pricing any payoff $\psi$:

- Find the price deflator $\pi$
- Price the derivative security at $\|\pi\|_1 E_\alpha(\psi)$

Thus the price of any given payout is (a constant multiplied by) its expectation under a certain special set of probabilities. These probabilities shall assume greater prominence once we consider the multi-period setting.

### 1.5 Multi-Period Model

The single period model presented above had one major flaw. Each investor selected a portfolio $\theta$ at the initial date, and then held it until the terminal date $T$. The weights on the portfolio remained fixed through time (i.e. there was no trading in the market). We now generalize the above single period model to a multi-period setting where investors can rebalance their holdings as time passes. This will require a more complicated mathematical machinery, and we will have to be more formal regarding the probabilistic setting.

We assume that market participants are allowed to rebalance their portfolios at times $0 = t_0, t_1, \ldots, t_K = T$. Let $(\Omega, \mathcal{F}, P)$ be a probability space, with $\Omega$ finite, and let $\{\mathcal{F}_t\}_{t=t_0}^T$ be a filtration on $\mathcal{F}$. The natural interpretation of the filtration is that it represents the way in which information is revealed through time, i.e. $\mathcal{F}_{t_k}$ represents all the information available to investors at time $t_k$. So if an event $A$ is in $\mathcal{F}_{t_k}$, then at time $t_k$ investors will know whether or not event $A$ has occurred based on their current information.
As in the previous section, we assume that there are a finite number, \( N+1 \), of securities available for trade in the market. Each security price is described by a stochastic process \( S^n_t \), \( n = 0, \ldots, N \). Thus, \( S^n_{t_k} \) represents the price of security \( n \) at time \( t_k \). We further assume that each of the processes \( S^n \) is adapted to the filtration \( \mathcal{F} \). If we take the filtration \( \mathcal{F}_t \) to be that generated by the price processes \( S^n \), then the time \( t_k \) prices of each security are known at \( t_k \).

We isolate a security to serve as a discount factor (in most market models, this is taken to be a risk free bond). For our purposes, it is enough to choose \( S^0 \) as the discount process, or numeraire for the model (we further require that \( S^0_0 = 1 \)). Denote the discounted security process \( Z \) by

\[
Z = \beta S = \frac{1}{S^0_0} S
\]

A trading strategy \( \theta \) is a predictable \( \mathbb{R}^{N+1} \) valued stochastic process. The interpretation is that the trading strategy \( \theta \) represents holdings of \( \theta^n_{t_k} \) units of security \( n \) at time \( t_k \) (here the superscript denotes the component process, and the subscript denotes the time index). This convention will be followed throughout. To ease the notational burden, we shall also denote \( S^n_{t_k} \) by \( S^n_k \), with a similar convention for all other processes. The value of the trading strategy \( \theta \) at time \( t \) is therefore given by:

\[
\sum_{n=0}^{N} \theta^n_t S^n_t = \theta_t \cdot S_t
\]

where \( S \) is the \( N+1 \) dimensional process with component \( n \) equal to \( S^n \). The point of making the trading strategy a predictable process is that investors should decide on their time \( t \) holdings prior to knowing the prices at time \( t \). At time \( t_k \) the investor is assumed to rebalance the portfolio from the weights \( \theta_k \) to the weights \( \theta_{k+1} \). For arbitrage pricing, we are chiefly interested in portfolios where this rebalancing can be done without an infusion or withdrawal of funds (i.e. the total value of the portfolio remains constant). This leads to the notion of a self-financing trading strategy.

A trading strategy \( \theta \) is said to be self-financing if for every \( k = 0, \ldots, K-1 \)

\[
\theta_k \cdot S_k = \theta_{k+1} S_k
\]
Thus at time $t_k$ the investor has the portfolio $\theta_k$, worth $\theta_k \cdot S_k$. This portfolio is rebalanced into the equally valued portfolio $\theta_{k+1} \cdot S_k$, which is then held until time $t_{k+1}$, when it becomes worth $\theta_{k+1} \cdot S_{k+1}$. The self-financing condition states that the rebalancing of the portfolio weights is performed without an infusion or withdrawal of cash from the portfolio. The change in value of the holdings in the interval $[t_k, t_{k+1}]$ is given by

$$\theta_{k+1} \cdot S_{k+1} - \theta_{k+1} \cdot S_k$$

The total gain from time 0 to time $t_k$ is thus

$$G_k = \sum_{m=1}^{k} \theta_m \cdot (S_m - S_{m-1})$$

We also denote the total value of the trading strategy $\theta$ at time $t_k$ by $V_k$, i.e.

$$V_k = \theta_k \cdot S_k$$

Elementary manipulations show that at any time, the value of a portfolio is equal to its initial cost plus its cumulative gains.

$$V_k = \theta_k \cdot S_k = \theta_1 \cdot S_0 + G_k = \theta_1 \cdot S_0 + \sum_{m=1}^{k} \theta_m (S_m - S_{m-1})$$

(In the continuous time limit, the last term will become a stochastic integral, yielding the value at any given time as the stochastic integral of the trading strategy with respect to the stock processes). For simplicity, we consider only trading strategies $\theta$ such that the corresponding value processes $V$ are always positive (i.e. investors never go into debt).

A *contingent claim* is an $\mathcal{F}_T$ measurable random variable $X$. This represents a security that pays out the stochastic amount $X$ at the terminal date $T$. Once again, the fundamental problem of derivative securities pricing asks whether it is possible to determine the price of each contingent claim given the prices $S^\circ_n$ of the fundamental securities. Our approach will be the same as in the first section. We seek a strategy involving only the fundamental securities that exactly replicates the payoff of the contingent claim. If this
strategy only requires a single initial expenditure of cash (i.e. if it is self-financing), then the cost of the replicating strategy must be identical to that of the derivative. To state these conditions a little more formally, we must define what is meant by an arbitrage in the multi-period model.

A self-financing trading strategy \( \theta \) is said to be an arbitrage if \( \theta_1 \cdot S_0 = 0 \) and \( P[V_T > 0] > 0 \).

That is, the strategy costs nothing to set up, but yields a strictly positive profit at the terminal date with strictly positive probability. This definition may seem to destroy the independence of the empirical measure \( P \) that arose in the single period model. This is not the case. The arbitrage definition presented here depends on \( P \) only through its null sets, i.e. it is invariant under the substitution of equivalent measures. This is actually the identical situation to the single period case, where the measure prescribed the null sets (in that section, the “possible future scenarios” for the market were the atoms generating the \( \sigma \)-algebra of measurable sets).

The market model is said to be complete if for every contingent claim \( X \geq 0 \) there exists a self-financing trading strategy \( \theta \), which is admissible in the sense that \( V_k \geq 0 \) for all \( k \), and replicates \( X \),

\[
V_T = \theta_K \cdot S_K = X \quad P \text{ a.s.} \tag{1.21}
\]

The theorem in the following section highlights the irrelevancy of the empirical measure \( P \). The equivalent measure \( Q \) is the multi-period analogue of the price deflator of the single period model.

**The Fundamental Theorem in Discrete Time**

In this section, we present a discrete time analogue of the theorem presented in the single period setting. The intuitive content of the theorem is that each derivative security (contingent claim) has a unique price if and only if markets are complete and there is a price deflator. The multiperiod (and continuous time) counterpart to the single period price deflator is the equivalent martingale measure, i.e. a measure \( Q \) on \( F \), equivalent to \( P \) (in the sense that they have the same null sets), under which each of the discounted security processes \( \beta S \) is a martingale. The financial interpretation
of this notion is that under $Q$ each security is expected to grow at the same rate as the numeraire ($S^0$).

A functional $\pi$ on the class of $\mathcal{F}_K$ measurable random variables is called a price system if:

- $\pi(X) \geq 0$, $\pi(X) = 0$ iff $X = 0$ a.s. $P$.
- $\pi(aX + bX') = a\pi(X) + b\pi(X') \forall a, b \geq 0$
- $\pi(V_K(\theta)) = V_0(\theta)$ for all admissible trading strategies $\theta$

**Lemma** There is a one to one correspondence between price systems $\pi$ and equivalent martingale measures $Q$ given by:

- $\pi(X) = E_Q(\beta_K X)$ and
- $Q(A) = \pi(S^0_K 1_A) A \in \mathcal{F}$

**Proof:** Let $Q$ be an equivalent martingale measure, and $\pi(X) = E_Q(\beta_K X)$. Then $\pi$ clearly satisfies the first two axioms for a price system. Let $\theta$ be an admissible trading strategy. By self-financing, the predictability of $\theta$, and the fact that $\beta S$ is a martingale under $Q$:

$$E_Q(\theta_k \beta_k S_k | \mathcal{F}_{k-1}) = E_Q(\theta_k E_Q(\beta_k S_k | \mathcal{F}_{k-1})) = E_Q(\theta_k \beta_{k-1} S_{k-1}) = E_Q(\theta_{k-1} \beta_{k-1} S_{k-1})$$

$K$ applications of the above equality then yield

$$\pi(V_K(\theta)) = E(\beta_0 \theta_K S_K) = E(\beta_0 \theta_0 S_0) = V_0(\theta) \quad (1.23)$$

Conversely, if $\pi$ is a price system, and $Q$ is a measure on $\mathcal{F}$ defined by $Q(A) = \pi(S^0_K 1_A) A \in \mathcal{F}$, the finiteness of $\Omega$, together with the axioms for a price system ensure that $Q$ is a measure on $\mathcal{F}$ equivalent to $P$. For the martingale condition, let $\tau \leq K$ be an $\mathcal{F}_t$ stopping time. Fix $n \geq 1$ and consider the admissible, self-financing trading strategy defined by

$$\theta^n_t = (\beta^n \tau^n S^n_{\tau^t}) 1_{t > \tau} \quad \theta^n_t = 1_{t \leq \tau} \quad (1.24)$$
with $\theta^m = 0$ for all other $m$. We have $V_0(\theta) = S_0^n$, and $V_T(\theta) = S_T^n \beta, S_T^n$.

Since $\pi$ is a price system:

$$S_0^n = \pi(S_K^n, \beta, S_T^n) = E_Q(\beta_T S_T^n, \beta, S_T^n) = E_Q(\beta, S_T^n) \quad (1.25)$$

Since the component $n$, and the stopping time $\tau$ are arbitrary, we have that $\beta S$ is a $(Q, F)$ martingale. Q.E.D.

**Theorem 1.5.1.** The market is free from arbitrage opportunities if and only if there is an equivalent martingale measure $Q$.

**Proof:** The proof is similar to the discrete time result, and is adapted from Harrison and Pliska [35], Duffie [26], and Baxter [7].

Suppose that there is no arbitrage. Let $X^+$ denote the set of all $\mathcal{F}_K$ measurable random variables $X$ such that $E_P(X) \geq 1$. Let $X^0$ denote the set of all $\mathcal{F}_K$ random variables $X$ such that $X = V_T(\theta)$ for some self-financing trading strategy $\theta$ ($\theta$ is not necessarily admissible), with $V_0(\theta) = 0$. By the absence of arbitrage opportunities, we have that $X$ and $X^+$ are disjoint (see Harrison and Pliska [35] p. 228). Since $X^0$ is a linear subspace, and $X^+$ is a closed convex cone in $\mathbb{R}^n$, the Separating Hyperplane Theorem yields the existence of a linear functional $F$ on $\mathbb{R}^n$ such that $F(X) > F(Y)$ for all $X \in X^+$, $Y \in X^0$. Since $X^0$ is a linear subspace, we must have $F(Y) = 0 \forall Y \in X^0$ and $F(X) > 0 \forall X \in X^+$.

Consider the functional $\pi(X)$ defined by

$$\pi(X) = \frac{F(X)}{F(S_K^n)} \quad (1.26)$$

Then $\pi$ clearly satisfies the first two axioms for a price system. To verify the third, let $\theta$ be any self-financing, admissible trading strategy, and consider the strategy $\psi$ defined by:

$$\psi^k_n = \begin{cases} \theta^0_k - V_0(\theta) & n = 0 \\ \theta^1_k & \text{otherwise}. \end{cases} \quad (1.27)$$

It follows that $\psi$ is a self-financing trading strategy with $V_0(\psi) = 0$, and $V_T(\psi) = V_T(\theta) - V_0(\theta) S_K^n$, i.e. an element of $X^0$. Therefore, it must be that
\[ \pi(V_K(\psi)) = 0, \text{ and} \]
\[ \pi(V_K(\theta)) - V_0(\theta) = \pi(V_K(\theta) - V_0(\theta)S_K^0) = \pi(V_K(\theta)) = 0 \]  
(1.28)

So \( \pi \) is a price system, and by the lemma there must exist an equivalent martingale measure.

Conversely, let \( Q \) be an equivalent martingale measure, and suppose that \( \theta \) is an arbitrage. Then \( P[V_T(\theta) > 0] > 0 \), and hence \( Q[V_T(\theta) > 0] > 0 \), so \( V_0(\theta) = E_Q(\beta_T V_T(\theta)) > 0 \), contradicting the fact that \( \theta \) is an arbitrage. Q.E.D.

Harrison and Kreps [34] have shown that the market is complete (every contingent claim \( X \geq 0 \) can be written as \( V_T(\theta) \) for an admissible, self-financing trading strategy \( \theta \), if and only if the equivalent martingale measure is unique. This financial content of this statement is that each contingent claim has a unique (arbitrage-free) price if and only if markets are complete and there are no arbitrage opportunities.

Before proceeding to the continuous case, we take a moment to obtain an intuitive interpretation for the above theorem. First, notice that since the discounted process \( \beta S \) is a martingale under \( Q \), we have

\[ E_Q(\beta_T S_T) = \beta_0 S_0 = S_0 \]  
(1.29)

so that the price of any security is its discounted expected payout under the measure \( Q \) (\( Q \) is often referred to as risk-neutral since it has the affect of changing the drift of each component of process \( S \) to that of the risk-free security \( S^0 \), see the continuous time formulation for an explicit calculation). Similarly, for any value process \( V \),

\[ E_Q(\beta V_T(\theta)) = V_0(\theta) \]  
(1.30)

If the market is complete, then we once again have an algorithm for pricing any contingent claim:

- Determine the self-financing trading strategy \( \theta \) such that \( V_T(\theta) = X \) almost surely.
The arbitrage free price for the contingent claim is $V_0(\theta)$.

Again, the main theorem of this section provides an alternative statement of this methodology.

- Determine the measure $Q$, equivalent to $P$ such that the discounted stock process $\beta S$ is a martingale under $Q$.
- Set the price of the claim $X$ to be equal to $E_Q(\beta_T X)$.

### 1.6 Continuous Time Model

In this section, we consider the formulation of the market model when trading is allowed to occur continuously through time. We shall restrict the mathematical framework to the case where the stock prices are defined by Itô diffusions. The motivation for this restriction is twofold. First, in the diffusion setting, the connections between completeness of the market, the existence of an equivalent martingale measure, and the absence of arbitrage remain as clear as in the discrete time case. This is not the case in the more general setting. In fact, while there exist a large number of proofs of the “fundamental theorem of asset pricing”, in discrete time, the problem in the continuous time setting can be said to remain open (although, see the work of Delbaen and Schachermeyer [18]). Secondly, by working in the diffusion setting, we can take advantage of the close connections between Markov processes and partial differential equations to obtain explicit valuation formulas for many contingent claims. Our final section presents an important example of this, the Black-Scholes valuation formula for a European call option.

We now construct a market model where there exist $N$ risky stocks, and a single risk-free bond (notice that in the continuous time setting, we explicitly single out a security to act as numeraire). The evolution of the risk-free bond through time is described by the equation

$$dS^0 = r_t S^0 dt \quad (1.31)$$
while the stochastic behaviour of the traded risky securities follow an $N$ dimensional Itô process

$$dS = \mu Sdt + \sigma SdW \tag{1.32}$$

where $S = (S^1, \ldots, S^N)$, $\mu \in \mathbb{R}^N$, $\sigma \in \mathbb{R}^{N \times N}$, $W$ is an $N$ dimensional Brownian motion on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$, $\{\mathcal{F}_t\}$ is right continuous, and contains all the $P$-null sets in $\mathcal{F}$ (i.e. $\{\mathcal{F}_t\}$ satisfies the “usual conditions”), and $\mu$ and $\sigma_i$, $i = 1, \ldots, n$ are $\{\mathcal{F}_t\}$ predictable processes. We consider a finite time interval $[0, T]$, where $T$ is once again referred to as the terminal date.

A trading strategy is a $\mathbb{R}^{N+1}$ valued stochastic process, predictable with respect to the filtration $\{\mathcal{F}_t\}$ (for the definition of predictability for continuous processes, see e.g. Rogers and Williams [70]).

Once again, we denote $\beta = 1/S^0$, and consider the discounted stock process $\beta S$. The discrete time definitions for the value and gains processes for the market model suggest an extension to the continuous time limit by stochastic integrals. This yields the value process $V$, such that $V_0(\theta) = \theta_0 S_0$ and

$$V_t(\theta) = \sum_{n=0}^{N} \theta^n S^n_t \tag{1.33}$$

The gains process for a given trading strategy $\theta$, $G_t(\theta)$, is defined by

$$G_t(\theta) = \int_0^t \theta_s \cdot dS_s = \sum_{n=0}^{N} \int_0^t \theta^n_s dS_s \tag{1.34}$$

A trading strategy $\theta$ is said to self-financing if, for every $t \in [0, T]$,

$$V_t(\theta) = V_0(\theta) + G_t(\theta) \tag{1.35}$$

It should be noted that the it is by no means self-evident that the above definitions are the appropriate ones for the continuous time model. The restriction of trading strategies to predictable processes, and the above stated definitions of the value and gains processes are usually defended on their formal resemblance to the analogous notions in discrete time. A similar argument results in the following definition of an arbitrage opportunity.
A trading strategy \( \theta \) is called an arbitrage opportunity if \( V_0(\theta) = 0 \) and 
\[
P[V_T(\theta) > 0] > 0.
\]

Motivated by the discrete time case, we are interested in determining a measure \( Q \) such that the discounted value process is a \( Q \)-martingale. We then examine the relationship between the existence and uniqueness of such a measure and the economic notions of market completeness and the absence of arbitrage.

In the setting of Ito diffusions, the existence of an equivalent martingale measure is a consequence of Girsanov's theorem on the behaviour of processes under change of measure. In order to apply Girsanov's theorem we require the following condition on the covariance matrix process \( a(t) = \sigma(t)\sigma^T(t) \) to hold. There must exist \( \varepsilon > 0 \) such that
\[
\xi^T a(t) \xi \geq \varepsilon ||\xi||^2 \quad (1.36)
\]
for all \( \xi \in \mathbb{R}^N \), and \( (t, \omega) \in [0, T] \times \Omega \). This nondegeneracy assumption ensures that \( \sigma^{-1}(t, \omega) \) exists and is bounded (see Karatzas and Shreve [50]).

Denote by \( \gamma(t) \), the \( \mathbb{R}^N \) valued transition function
\[
\gamma(t) = \sigma^{-1}(\mu(t) - r(t)1) \quad (1.37)
\]
where \( 1 = (1, \ldots, 1)^T \in \mathbb{R}^N \).

In addition to the nondegeneracy assumption (1.36), we require that Novikov's condition:
\[
E \left[ \exp \left( \frac{1}{2} \int_0^T \gamma_s^2 ds \right) \right] < \infty \quad (1.38)
\]
holds. This implies that the exponential local martingale:
\[
Z_t = \exp \left[ - \int_0^t \gamma_s dW_s - \frac{1}{2} \int_0^t \gamma_s^2 ds \right] \quad (1.39)
\]
is in fact a martingale (with respect to \( (\mathcal{F}_t, P) \)). See, for example Øksendal [67]. The equivalent martingale measure \( Q \), for the model can then be defined by the equation
\[
Q(A) = E_P[Z(T)^1_A] \quad (1.40)
\]
In other words, $Q$ is the measure such that

$$\frac{dQ}{dP} = Z(T) \quad (1.41)$$

(the reason for choosing this measure is that the process $\gamma$ is selected so as to obtain the change in the drift coefficient in the Ito process from $\mu$ to $r1$).

The strict positivity of $Z$ automatically yields that $Q$ is equivalent to $P$. By Girsanov's theorem, we have that:

$$\tilde{W}_t = W_t + \int_0^t \gamma(s) ds \quad (1.42)$$

is a $(\mathcal{F}_t, Q)$ Brownian motion (recall the remark above regarding the importance of the choice of measure in specifying the Brownian motion).

Under the new measure, the stock price assumes the form

$$dS = r(t)S_t dt + \sigma_t S_t d\tilde{W} \quad (1.43)$$

We also have that $Y_t = \beta_t S_t$ is a martingale under $Q$. A number of important observations are in order. The first is that the volatility coefficient of the process is invariant under change of measure. This is of fundamental importance to option pricing. In the arbitrage pricing methodology the value of a contingent claim $X$ does not depend on its expected payout $E_P(X)$ with respect to the measure $P$ governing stock price movements. In the Black-Scholes model, this will reveal itself through the independence of the valuation formula from the drift coefficient $\mu$. On the other hand, the volatility $\sigma$, which is invariant under the change of measure will have an important role in future valuation formula. The equivalent martingale measure $Q$ is often referred to as the risk-neutral measure, because it equates the drifts of all risky securities to that of the risk-free bond $S^0$.

A positive $\mathcal{F}_T$ measurable random variable is referred to as a contingent claim. Again, we are interested in determining which contingent claims can have their payoffs replicated by self-financing trading strategies involving the fundamental securities.
For any contingent claim, the conditional expectations

\[ P_t = E_Q(\beta_T X | \mathcal{F}_t) \]  

are a martingale (see Rogers and Williams [70]). By the nonsingularity of \( \sigma \), and the martingale representation theorem (see Rogers and Williams [70]) can be represented as an integral with respect to the \( N \) dimensional martingale \( Y \).

\[ P_t = P_0 + \sum_{k=1}^{N} \int_0^t \phi_u^k dY_u^k \]  

(1.45)

where \( \theta \in \mathbb{R}^N \) is a predictable process. Now consider the trading strategy \( \theta \in \mathbb{R}^{N+1} \) with \( \theta^k = \phi^k \) for \( k = 1, \ldots, N \) and

\[ \theta_0 = P_t - \sum_{k=1}^{N} \theta_t^k Y_t^k \]  

(1.46)

By the above equation, we have that

\[ V_t(\theta) = (\theta_t^0 S_t^0) + \sum_{k=1}^{N} \theta_t^k S_t^k) = V_0 + G_t(\theta) \]  

(1.47)

so that \( \theta \) is self-financing. If there is to be no arbitrage, the price of the contingent claim \( X \) must be equal to

\[ V_0(\theta) = E_Q(\beta_T S_T) \]  

(1.48)

To illustrate this, suppose \( \theta \) is such that \( \theta_0 \cdot S_0 = 0 \). Consider the process \( \beta_t V_t(\theta) \), which is a martingale under \( Q \). Then

\[ E_Q(\beta_T V_T(\theta)) = \beta_0 V_0(\theta) = 0 \]  

(1.49)

so that the positivity of \( \beta \) gives that \( V_T(\theta) = 0 \). We conclude that the existence of an equivalent martingale measure implies that there is no arbitrage.

The above discussion shows that the market is complete, and under sufficient regularity conditions there exists an equivalent martingale measure (and hence there is no arbitrage). As in the discrete time case, this leads to two related algorithms for pricing the claim \( X \). To use replication directly:
• Determine the self-financing trading strategy \( \theta \) such that \( V_T(\theta) = X \) almost surely.

• Set the claim’s price to be equal to \( V_0(\theta) \).

The same price is also provided by the “risk-neutral valuation” methodology:

• Determine the measure \( Q \) such that \( \beta_t S_t \) is a martingale under \( Q \).

• Set the claim’s price to be equal to \( E_Q(\beta_T X) \).

We have had to assume certain conditions in order to ensure the existence of an equivalent martingale measure. In the general case, where assets are allowed to be arbitrary semimartingales (stochastic integrators), and trading strategies are predictable processes (integrands), the situation is even more complex. See the papers of Harrison and Pliska [35], and Delbaen and Schachermayer [18] for attempts at developing the continuous time framework in full generality.

1.6.1 The Black-Scholes Equation

As a special case of the above diffusion model, we consider a market containing only a risk-free bond \( B \) and a risky stock \( S \) governed by the set of equations

\[
\begin{align*}
    dB &= rBdt \\
    dS &= \mu Sdt + \sigma SdW
\end{align*}
\]

where \( \mu, \sigma, \) and \( r \) are real constants referred to as the expected stock return, stock volatility, and risk-free interest rate respectively. Here \( W \) denotes a real-valued Brownian motion, defined on a probability space \((\Omega, \mathcal{F}, P)\). The discount process is given by \( \beta_t = 1/B_t \). We again employ the Girsanov formula to change measures to \( Q \), under which the drift constant becomes \( r \).

\[
\begin{align*}
    dB &= rBdt \\
    dS &= \mu Sdt + \sigma SdW
\end{align*}
\]

where \( \mu, \sigma, \) and \( r \) are real constants.
Let $X$ be a positive $\mathcal{F}_T$ measurable random variable. Then, according to the arbitrage pricing methodology developed in the previous section, we have that the unique arbitrage free price of the contingent claim $X$ is

$$E_Q(\beta T X) = \exp(-rT)E_Q(X) \quad (1.54)$$

An application of the *Feynman-Kac formula* (see for example Øksendal [67]) then yields that the price of the claim $X$, when viewed as a function of the current stock price $S_0 = S$, and the time to expiration $T - t$, will obey the partial differential equation:

$$rf = \frac{\partial f}{\partial t} + rS \frac{\partial f}{\partial S} + \frac{1}{2} \sigma^2 \frac{\partial^2 f}{\partial S^2} \quad (1.55)$$

with terminal condition

$$f(S, T) = X(S) \quad (1.56)$$

Notice that the partial differential equation explicitly involves the risk-free interest rate $r$, and volatility $\sigma$, but is independent of the expected return $\mu$.

In the case where the payoff $X$ is that of a standard European call option, with strike price $K$ i.e. $X(S) = [S - K]_+$, the expectation (1.54) can be evaluated explicitly to yield the following analytic solution to the initial value problem (where $N$ represents the standard normal distribution function)

$$f(S, t) = SN(d_1) - e^{-r(T-t)}KN(d_2) \quad (1.57)$$

where

$$d_1 = \ln(S/K) + (r + \sigma^2/2)(T - t) \quad \sqrt{T - t} \quad (1.58)$$

$$d_2 = d_1 - \sigma - \sqrt{T_t} \quad (1.59)$$

This is the famous Black-Scholes formula for the price of a European call option on a non-dividend paying stock (see Black and Scholes [9]).
Chapter 2

Portfolio Optimization

2.1 Introduction

The recurring theme in the applications of mathematics to finance is that there is a basic tradeoff between risk and reward. Reward is only obtained by those market participants who are willing to incur some risk. An arbitrage violates this tradeoff, since it allows unlimited reward for zero risk. The first chapter presented the theory of derivative security pricing using arbitrage arguments. The main results derived therein demonstrated that, given the completeness of markets, the assumption of the absence of arbitrage is sufficient to determine a unique price for every derivative security. In this chapter, we go beyond simply pricing securities. Taking prices as exogenously specified, we analyze the problem of determining the optimal strategy for a given investor.

The problem of determining how a market participant should allocate wealth among the securities that are available for trade is referred to as the portfolio selection problem. Portfolio selection requires the investor to choose a preferred wealth allocation from a number of feasible alternatives. This has two significant consequences, one conceptual and the other mathematical. First, since an investor must make a choice between eligible portfolios, preferences enter into the model explicitly. Second, since the investor will seek to employ the best strategy, the mathematical solution to the portfolio selection problem will require tools from the theory of optimization. More precisely, since economic phenomena exhibit random behaviour, the portfolio selection
problem falls within the domain of stochastic optimization.

The chapter is structured as follows. We begin by examining historically popular models for the portfolio selection problem. The first is the venerable mean-variance portfolio selection model of Harry Markowitz, which assumes that the investor evaluates a portfolio's performance in terms of its expected return, and returns variance. Next, we outline Merton's application of stochastic control theory to portfolio choice. The main body of the chapter is devoted to a development of the Scenario Optimization model. Scenario Optimization extends the Markowitz mean-variance model by introducing a more general notion of risk, known as regret. We examine a parametrization of the basic Scenario Optimization model, arrived at by the inclusion of an additional constraint, representing a prescribed level of required overperformance of a benchmark portfolio. This provides a mechanism for explicitly analysing the tradeoff between risk and reward in the model. This leads to the risk-reward frontier a generalization of the mean-variance frontier of traditional portfolio theory. We present a number of examples illustrating how the risk-reward frontier provides information about the characteristics of financial markets. The chapter concludes by indicating future directions for the application of the mathematical theory of optimization to the problem of portfolio choice.

2.2 Traditional Portfolio Selection Models

2.2.1 Markowitz' Mean-Variance Model

The mean-variance model was introduced in 1952 by H. Markowitz, who later won the Nobel prize in Economics for his work in portfolio theory. The model represented one of the first attempts to present mathematically the intuitive axiom that markets exhibit a tradeoff between risk and reward.

The main assumption of the mean-variance framework is that all the relevant information for an investor to quantify the performance of a given portfolio of assets is given in terms of two parameters: its expected return, and returns variance. Expected return is used as a measure of the reward attached to holding a given portfolio. The variance of the returns distribution is taken as a measure of risk. Investors are assumed to seek the portfolio
that simultaneously maximizes expected returns and minimizes risk (returns variance). In other words, for a fixed level of expected return, investors will select the portfolio with minimum variance, and similarly, for a given variance investors will choose to hold the portfolio with the highest expected return.

**Variance as a Measure of Risk**

The use of variance as a measure of risk assumes that investors regard variability as undesirable. It is more often the probability of a certain level of loss that interests the market practitioner. Under the assumption that returns are normally distributed, the probability of a given level of loss increases with the standard deviation of the distribution (for a fixed mean), and hence variance is an appropriate measure of risk. For asymmetric probability distributions, it is not at all clear that variability provides an appropriate representation for the risk of a portfolio. This has led to the suggestion of many alternative risk measures, which typically use some more direct measurement of the potential for loss. For example, Markowitz [58] considered the standard deviation of the negative returns as an alternative risk measure. Later in this chapter, we shall examine *regret*, which measures risk against a specific benchmark portfolio.

**The Mean-Variance Model**

We now present the mathematical formulation of the basic mean-variance portfolio selection model. We consider a market in which there are $n$ assets among which wealth will be invested. Denote the return of the $i$th asset over the next period by $R_i$. Since the behaviour of the assets in the future is unknown, $R_i$ is a random variable. Denote the expected level of return on the $i$th asset by $\mu_i$,

$$E(R_i) = \mu_i \quad i = 1, \ldots, n$$  \hspace{1cm} (2.1)

Let the variance of $R_i$ be $\sigma_i^2$, so

$$E[(R_i - \mu_i)^2] = \sigma_i^2 \quad i = 1, \ldots, n$$  \hspace{1cm} (2.2)

Finally, denote by $\sigma_{ij}$ the covariance between $R_i$ and $R_j$

$$E[(R_i - \mu_i)(R_j - \mu_j)] = \sigma_{ij}$$  \hspace{1cm} (2.3)
Let $\Sigma$ be the matrix with $\Sigma_{ij} = \sigma_{ij}$, if $i \neq j$ and $\Sigma_{ii} = \sigma_i^2$. We consider portfolios defined by vectors of weights, $w \in \mathbb{R}^n$, where $w_i$ represents the fraction of total wealth invested in the $i$th asset. If all wealth must be invested in the assets, and negative positions cannot be taken, we must have

$$||w||_1 = \sum_{i=1}^n w_i = 1, \quad w_i \geq 0, i = 1, \ldots, n$$  \hspace{1cm} (2.4)

The expected return of the portfolio with defined by the weights $w$ is given by

$$\mu_w = \sum_{i=0}^n w_i \mu_i$$  \hspace{1cm} (2.5)

and its variance

$$\sigma^2 - w = w^T \Sigma w$$  \hspace{1cm} (2.6)

In the introduction to this section, it was stated for a given level of return, investors would seek to minimize risk (represented here by variance).

$$\min_w w^T \Sigma w$$  \hspace{1cm} (2.7)

subject to

$$\sum_{i=1}^n w_i \mu_i \geq \alpha$$  \hspace{1cm} (2.8)

$$\sum_{i=1}^n w_i = 1$$  \hspace{1cm} (2.9)

$$w_i \geq 0 \quad i = 1, \ldots, n$$  \hspace{1cm} (2.10)

Similarly, the problem of maximizing the return of a given portfolio while only assuming a certain level of risk becomes

$$\max_w \sum_{i=1}^n w_i \mu_i$$  \hspace{1cm} (2.11)
subject to

\[ w^T \Sigma w \leq \alpha \]  \hspace{1cm} (2.12)

\[ \sum_{i=1}^{n} w_i = 1 \]  \hspace{1cm} (2.13)

\[ w_i \geq 0 \quad i = 1, \ldots, n \]  \hspace{1cm} (2.14)

By the use of penalty functions, either of the above problems can assume the

\[ \max w \mu_w - \lambda w^T \Sigma w \]  \hspace{1cm} (2.15)

subject to

\[ \sum_{i=1}^{n} w_i = 1 \]  \hspace{1cm} (2.16)

\[ w_i \geq 0 \quad i = 1, \ldots, n \]  \hspace{1cm} (2.17)

form

where \( \lambda > 0 \) is a parameter representing the relative importance of risk (variance) and reward (expected return). The above problem is a parametric quadratic programming problem (with parameter \( \lambda \)), and forms the essence of the Markowitz mean variance model. As \( \lambda \) varies, the tradeoff between risk and reward in the model is revealed.

**The Mean-Variance Efficient Frontier**

The mean-variance model assumes that among all portfolios with a fixed level of risk, investors will choose the one with the maximum expected return. Similarly, among all portfolios with a given expected return, investors will select the one with the minimum risk (variance). The relation plotting the minimum variance attainable for each level of return (or alternatively, the maximum return achievable for each level of risk) is referred to as the *risk-return frontier*. Portfolios whose mean and variance lie on the efficient frontier are referred to as efficient portfolios. The set of all efficient portfolios is therefore determined by the solutions to the program (2.15) as the
parameter $\lambda$ is allowed to vary on $(0, \infty)$. According to mean-variance theory, any investor will choose to hold an efficient (various utility assumptions can be employed to determine the optimal efficient portfolio), and hence any investor's portfolio selection problem is given by (2.15) for a specific level of $\lambda$. Many alternative methods have been suggested for comparing portfolios that lie on the efficient frontier, see e.g. [74] [65].

Two observations are of special importance to contrast the present model with the arbitrage pricing theory of the first chapter. First, investor preferences must explicitly come into play in any model for portfolio selection (above by the selection of risk measure, and of aversion parameter $\lambda$). Second, the highly touted "risk-neutral" measure, which was vital to the solution of the pricing problem of the first chapter, is entirely irrelevant to the problem of optimal portfolio selection. In this chapter, where prices are assumed to be given exogenously, it is the previously discarded empirical measure that is important.

2.2.2 An Application of Stochastic Control Theory to Portfolio Selection

In the mean-variance model presented above, the investor selected a portfolio based on the properties of its returns distribution, and then held onto this portfolio without rebalancing until an expiration date. Due to the single time step, and the absence of dynamic trading in securities, the mean-variance formulation bears a certain resemblance to the single period pricing model presented in the chapter on derivative security valuation by arbitrage. When considering arbitrage pricing, after examining a static single period market we progressed to the case where prices followed continuous time diffusion processes. We now present a similar extension for the portfolio selection problem.

We consider the original formulation of the portfolio selection problem in continuous time, due to Merton [61] (our exposition closely resembles those in Duffie [26] and Øksendal [67]). The solution employs results from the theory of stochastic control for diffusion processes (see Fleming and Rishel [32]). For further work on the portfolio selection problem in continuous time, see Karatzas, Lehozcky, Sethi, and Shreve [48], and the survey paper by
We consider a market model with \( n + 1 \) securities, \( X_i, i = 0, \ldots, n \). The first security, \( X_0 \), is taken to be risk-free. The evolution of its value through time is governed by the following equation
\[
dX_0 = rX_0 dt
\]
for simplicity of exposition, we take \( r > 0 \) to be constant. The prices of the remaining \( n \) assets are assumed to be described by an \( n \) dimensional Ito diffusion:
\[
dx = \mu X dt + \sigma X dW
\]
where \( \mu \in \mathbb{R}^n \), \( \sigma \in \mathbb{R}^{n \times m} \) and \( W \in \mathbb{R}^m \) is a standard \( m \)-dimensional Brownian motion on a probability space \( (\Omega, \mathcal{F}, P) \) (as usual, we employ the natural filtration, i.e. \( \mathcal{F}_t = \sigma(W_s; s \leq t) \) augmented by the \( P \) null sets of \( \mathcal{F} \)).

Suppose that the investor holds the fraction \( w_i \) of total wealth in the risky asset \( i, i = 1, \ldots, n \). Therefore the fraction
\[
w_0 = 1 - \sum_{i=1}^n w_i
\]
of total wealth will be held in the riskless asset. The above constraint simply says that the total available wealth (and no more) should be allocated among the tradeable securities.

Investors may wish to rebalance their holdings as time progresses, based on the random behaviour of the market prices. This implies that the mathematical description of an investment strategy determining holdings in the market securities should be a stochastic process, i.e. \( w_i = w_i(\omega, t), i = 1, \ldots, n \). The total wealth process \( Y \) of the strategy investing \( w_i \) in risky asset \( i \) at time \( t \) is governed by the stochastic differential equation
\[
dY_t^w = Y_t^w \left( rw_0 + \sum_{i=1}^n w_i \mu_i dt + \sum_{i=1}^n \sigma_i w_i dW_t^i \right)
\]
where \( W_t^i \) is the \( i \)-th component of the process \( W_t \). In order to outlaw clairvoyance, it is typically assumed that the \( w_i \) are adapted processes (notice
that the total wealth of the investor at any time $t$ depends explicitly on the portfolio weightings $w_i$. The model is greatly simplified if the $w_i$ are assumed to be random only through their dependence on the securities prices, i.e.

$$w_i(\omega, t) = v_i(X(\omega, t), t) 
(2.22)$$

In the language of stochastic control theory, this means that the weights $w_i$ are restricted to be Markov controls (the terminology deriving from the observation that given (2.22) the stochastic process for $Y$ becomes an Ito diffusion, and therefore a Markov process).

In order to determine the optimal investment strategy, it is essential to develop a criterion for assessing the performance of a given portfolio. We assume that the investor under consideration will evaluate the portfolio by considering its total wealth at a terminal date $t_0$. Let the value of the wealth $Y$ be given by $U(Y)$ where $U : [0, \infty) \rightarrow [0, \infty)$. For obvious reasons $U(\cdot)$ should be increasing. Also, in economic applications of stochastic control, $U(\cdot)$ is usually assumed to be concave.

The expected value of the terminal wealth for the portfolio with holdings governed by the weight process $w$, as seen at time $s < t_0$, is given by

$$J(w, Y_s, s) = E\left[U(Y_{t_0}) \mid \mathcal{F}_s\right] 
(2.23)$$

If we impose the constraint that investors can never go into debt (i.e. the total wealth $Y$ must be positive or else the portfolio is liquidated), then the expected performance at time $s$ becomes

$$J(w, Y_s, s) = E\left[U(Y_T^w) \mid \mathcal{F}_s\right] 
(2.24)$$

where $T$ is the first exit time from the region $Q = \{(t, y) \mid t < t_0, y > 0\}$. (This first inequality deriving from the imposition of a terminal time, the second from the no-debt condition).

The differential generator corresponding to the process $Y_t^w$ is

$$A(w)(t, y) = \frac{\partial}{\partial t} + y \left( w_0 + \sum_{i=1}^{n} w_i \mu_i \right) \frac{\partial}{\partial y} + \frac{1}{2} y^2 \left( \sum_{i=1}^{n} w_i^2 \left( \sum_{i=1}^{n} \sigma_i^2 \right) \right) \frac{\partial^2}{\partial y^2} 
(2.25)$$
For this problem, the Hamilton-Jacobi-Bellman equation therefore assumes the form:

$$\sup_w A^w f(t, y) = 0 \quad (t, y) \in Q$$ (2.26)

$$f(y, t_0) = U(t_0) \quad f(0, t) = U(0)$$ (2.27)

(the second condition formalizes the intuitively obvious restriction that there is no utility to holding zero wealth).

For each \(w_i, i = 1, \ldots, n\), we apply the first order conditions for optimality to the equation (2.26) yielding

$$w_i(t, y) = (\mu_i - r) - \frac{f_y(t, y)}{y f_{yy}(t, y)} \left( \sum_{i=1}^{n} \sigma_i^2 \right)^{-1}$$ (2.28)

Substituting this into equation (2.26) yields the following nonlinear boundary value problem for the function \(f\) (thus also yielding the optimal portfolio weights \(w_i\))

$$A(w)(t, y) = \frac{\partial f}{\partial t} + y \left( w_0 r + \sum_{i=1}^{n} w_i \mu_i \right) \frac{\partial f}{\partial y} + \frac{1}{2} y^2 \left( \sum_{i=1}^{n} w_i^2 \right) \frac{\partial^2 f}{\partial y^2} = 0$$ (2.29)

$$w_i(t, y) = (\mu_i - r) - \frac{f_y(t, y)}{y f_{yy}(t, y)} \left( \sum_{i=1}^{n} \sigma_i^2 \right)^{-1}$$ (2.30)

$$f(y, t_0) = U(t_0) \quad f(0, t) = U(0)$$ (2.31)

While the equation (2.29) is in general quite difficult to solve, analytic solutions have been given for many different forms of the utility function \(U\). The problem has been generalized by a number of authors to deal with portfolio selection under constraints, transactions costs, and other violations of the perfect markets assumptions of the first chapter.
At this point, we reiterate the important differences between the portfolio selection problem and the pricing problem, as they are illustrated by the above stochastic control formulation. First, the function $U$ can be investor dependent, so preferences other than nonsatiability form an integral part of the model (notice the appearance of the drift parameter $\mu$ in the Hamilton-Jacobi-Bellman equation (2.29), and compare to its absence in the Black-Scholes equation (1.55)). Second, the risk neutral measure does not appear in the above exposition. The expectation operator $E$ denotes integration with respect to the empirical measure $P$. While $P$ is nearly irrelevant for derivatives pricing, it is vital to portfolio selection.

2.3 Scenario Optimization: A New Technique for Portfolio Selection

The remainder of this chapter is devoted to explaining the mathematical and financial properties of a recent model for portfolio selection, known as Scenario Optimization. The model was first introduced by Dembo [19].

Scenario Optimization introduces two main innovations to the theory of portfolio selection. The first is a new measure of risk, known as regret, that reflects the market practice of measuring performance in relation to a standard benchmark instrument. The second is a new algorithm for solving problems in stochastic optimization. We shall examine each of these innovations in turn, beginning with the measurement of risk in financial markets.

2.3.1 Regret: Measuring Risk Against a Benchmark

We have stated that a fundamental paradigm in the application of mathematics to finance that there is an inherent tradeoff in markets between risk and reward. The intuitive content of this assertion is much clearer than the suitability of any particular mathematical interpretation of the terms risk and reward. Optimization models quantify these notions, revealing how they counterbalance in theoretical markets. The Markowitz model assumes that reward is given by the expected return of a portfolio, while its risk is given by its standard deviation. In the stochastic control formulation introduced by Merton, measures of risk and reward are described implicitly
by the shape of the utility function $U$. In this section, we explain a new measure of risk, known as \textit{regret}, that underlies the financial applications of Scenario Optimization.

The notion of regret is motivated by the methods for performance measurement used by many market participants. In practice, financial agents are often evaluated by how well (or poorly) they perform in comparison to a given \textit{benchmark} portfolio. For example, mutual fund managers are required to outperform a certain financial index, and their performance is not evaluated on the basis of gross return, but instead on the return of their portfolio relative to the benchmark index. A similar situation arises in hedging applications, where one seeks a minimal cost portfolio that must do at least as well as a given derivative security under all possible future scenarios. For many complex calculations, the behaviour of a large portfolio of financial instruments must be mimicked by a much smaller portfolio. The determination of such a reduced portfolio is referred to as the \textit{portfolio compression} problem. When solving a compression problem, any discrepancy between the behaviour of the large portfolio and that of the compressed portfolio is considered to be a source of risk. In order to reflect the above practical considerations, Dembo [20] has introduced a new notion of risk, referred to as \textit{regret} which measures risk with respect to a benchmark random variable.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Select a random variable $\tau$ to be the \textit{benchmark} against which performance will be measured. Let $\mathcal{M}$ be a subspace of the space of all random variables (real valued, Borel measurable functions on $\Omega$). Then a \textit{regret functional} corresponding to the benchmark $\tau$ is a mapping $\rho(\cdot; \tau): \mathcal{M} \to \mathbb{R}$ obeying the following axioms:

- $\rho(\pi; \tau) \geq 0 \ \forall \pi \in \mathcal{M}$
- $\rho(\tau; \tau) = 0$
- $\rho(\pi_1 + \pi_2; \tau) \leq \rho(\pi_1; \tau) + \rho(\pi_2; \tau)$

The first condition says that regret should be positive for all portfolios (random variables). The second says that there should be no risk associated with holding the benchmark. The third condition (subadditivity) states that the total risk of a portfolio of assets cannot be reduced simply by dividing it
into component subportfolios (alternatively it can be viewed as expressing the financial axiom that diversification reduces risk). Examples of regret functionals meeting the above axioms include:

- The norm of the discrepancy: $\rho(\pi; \tau) = \|\pi - \tau\|_p$, where $1 \leq p \leq \infty$, $\mathcal{M} = L^p(P)$ and $\tau \in L^p(P)$.

- The expected underperformance with respect to the benchmark: $\rho(\pi; \tau) = \|\pi - \tau\|_p$, where $\mathcal{M}, \tau$ and $p$ are as above and $X_-$ denotes the negative part of the random variable $X$.

- If the benchmark is taken to be the constant random variable $0$, then $\rho(X; 0) = \sigma^2(X)$ is a regret functional, where $\mathcal{M} = L^2(P)$.

- If $\delta(\cdot, \cdot)$ is any pseudo-metric on $\mathcal{M}$, and $\tau \in \mathcal{M}$, then $\rho(\pi) = \delta(\pi, \tau)$ is a regret functional.

The definition of regret given above is general enough to include many of the traditional definitions of risk (e.g. standard deviation, and expected loss). Among the methods of risk assessment that do not meet the above description of a regret functional is the Value at Risk standard, introduced by J.P. Morgan [42], which may fail to satisfy the subadditivity requirement.

The main contribution of the regret definition is that it takes into account the tendency of market participants to evaluate their performance with respect to standard benchmark instruments (whose behaviour is often random itself). As we shall see in the following, this makes the regret measure a natural candidate for the objective function in a minimization problem.

2.3.2 Scenario Optimization: The General Case

In this section, we introduce the basic Scenario Optimization paradigm. In essence, the algorithm is as follows. The optimization is stated as the minimization of a regret functional (possibly time dependent), over a constrained set. The regret functional (or alternatively the underlying probability space), is then discretized to achieve an efficient numerical approximation for the problem. As we shall see below, a specific choice of discretization will allow the problem to assume a linear programming form.
The Scenario Optimization Formulation of a Stochastic Optimization Problem

The general form of a problem in mathematical optimization is

$$\min_{x} f(x)$$

subject to

$$x \in \Theta$$

where the optimization occurs over all $$x \in \Theta$$, where $$\Theta \subseteq X$$, and typically $$X$$ is a Banach space. Often, the feasible set is defined by a system of inequalities, so that the problem assumes the form:

$$\min_{x} f(x)$$

subject to

$$g_k(x) \leq 0 \quad k = 1, \ldots, n$$

$$h_j(x) = 0 \quad j = 1, \ldots, m$$

In optimization, randomness can enter into the problem framework in two conceptually different ways. The first is for the data to be observed with uncertainty (random perturbations due to observational or experimental error). This occurs, for example, in many engineering applications, when physical quantities can only be measured to within a given tolerance. Randomness can also enter into an optimization problem when the decision variables are assumed to be stochastic processes (i.e. the underlying variables are of a fundamentally stochastic nature). Many such problems from economics are considered in this thesis (e.g. the stochastic control problem presented above). The first situation is amenable to treatment either by stochastic optimization, or by perturbation methods (see, for example Zlobec [83], or Levitin [53]). The second demands methods from stochastic optimization, i.e. where certain functions in the optimization framework (2.32), are allowed to depend on random parameters.
Consider a probability space \((\Omega, \mathcal{F}, P)\). Then modifying the problem (2.34) to reflect the influence of stochastic factors results in the following \textit{stochastic optimization} problem:

\[
\min_x f(x(\omega), \omega) \tag{2.37}
\]

subject to

\[
h_k(x(\omega), \omega) \leq 0 \quad k = 1, \ldots, n \tag{2.38}
\]

\[
g_j(x(\omega), \omega) = 0 \quad j = 1, \ldots, m \tag{2.39}
\]

where the functions \(x : \Omega \to X, g_j, h_k : (X \times \Omega) \to \mathbb{R}\) are taken to be Borel measurable. Of course, the above problem can be subsumed in the general form (2.32), however, the explicit formulation (2.37) is useful since since it exhibits the distinct intuition and flavour of stochastic optimization. For example, when (2.37) is regarded as a standard optimization problem on \(X \times \Omega\), then the a solution \(x\) is feasible only if it obeys the constraints for all \(\omega \in \Omega\). The presence of the probability measure \(P\) allows for various alternative criteria for feasibility (for example, \(x\) could be \textit{almost-feasible} if it satisfied the constraints almost-surely \(P\)). Often the objective function \(f\), is in the form of an expectation with respect to the measure \(P\). Thus the measure \(P\) often plays a dual role, both accounting for the structure of the feasible set, and the selection of its optimal points (as an aside, we note that there is a parallel here to the role played by the empirical measure in portfolio selection in finance- which first determines the sets of measure zero for the pricing measure \(Q\), and then determines the objective function in many selection problems).

When the objective function is a regret functional with respect to the random variable \(\tau\), then the problem assumes the form:

\[
\min_x \rho(x; \tau) \tag{2.40}
\]

subject to

\[
h_k(x(\omega), \omega) \leq 0 \quad k = 1, \ldots, n \tag{2.41}
\]
Often the regret function can be expressed as an expectation with respect to the probability measure \( P \). This results in an objective function of the form:

\[
E[R(x(\omega), \tau(\omega))]
\]  

(2.43)

(e.g. \( R(x, y) = ||x - y|| \)). For stochastic processes with a time component, (i.e. multi-step optimization), the objective assumes the form

\[
\int w(t) E(R(x_i(\omega), \tau(\omega))) \, dt
\]  

(2.44)

where \( w \) is a weighting function, assumed to depend on time \( t \), but not the random parameter \( \omega \) (in financial applications, this often bears the interpretation of a discount factor).

The expectation operator \( E \) represents integration with respect to the probability measure \( P \). It is extremely unusual for the integral to be analytically tractable (yielding an explicit solution to the stochastic optimization problem via Lagrange multipliers). Typically, the integral will have to be approximated by a finite sum. Thus, the space \( \Omega \) is discretized into a finite number of representative “scenarios” \( \omega_s, s = 1, \ldots, S \), with probability weights \( p_s \). It is this discretization process that gives Scenario Optimization its name.

Under discretization, the Scenario Optimization problem assumes the form:

\[
\min_x \sum_{k=1}^{T} \sum_{i=1}^{S} p_{ik} R(x(\omega_i); \tau(\omega_i)) \]  

(2.45)

subject to

\[
h_k(x(\omega_i), \omega_i) \leq 0 \quad k = 1, \ldots, n \quad i = 1, \ldots, S
\]  

(2.46)

\[
g_j(x(\omega_i), \omega_i) = 0 \quad j = 1, \ldots, m \quad i = 1, \ldots, S
\]  

(2.47)

where the variable \( i \) indexes the \( S \) scenarios in the discretization of the probability space, and \( k \) indexes time, and \( p_{ik} \) is the probability that the \( i \)th future scenario will occur at time \( t_k \).
As an example to illustrate the efficiency of the method, we assume that regret is given by $E(R(x(\omega); \tau(\omega)))$, where $R(x(\omega) - \tau(\omega)) = |x(\omega) - \tau(\omega)|$, and all the constraint functions are linear, then the Scenario Optimization problem becomes:

$$\min_x \sum_{k=1}^{T} w(t_k) \sum_{i=1}^{S} p_{ik} |x_i - \tau_i|$$  \hspace{1cm} (2.48)$$

subject to

$$C_i x_i \leq 0 \quad i = 1, \ldots, S$$  \hspace{1cm} (2.49)$$

$$A_i x_i = 0 \quad i = 1, \ldots, S$$  \hspace{1cm} (2.50)$$

where $x_i = x(\omega_i)$, $\tau_i = \tau(\omega_i)$, $C_i$ is the $n \times S$ matrix representing the functions $h_k$ under scenario $i$, and $A_i$ is the $m \times S$ matrix representing the functions $g_j$ under scenario $i$. Thus the stochastic optimization problem has been reduced to a linear programming form. Hence, the goal of formulating a fundamentally stochastic problem in a computationally efficient manner has been achieved.

We now examine the applications of Scenario Optimization to various problems within mathematical finance.

### 2.3.3 Scenario Optimization Formulation of Portfolio Selection

In this section, we apply the general methodology for stochastic optimization described in the preceding sections to the problem of portfolio selection.

Consider an economy with a finite number, $N$, of instruments available for trade. We take today to be time 0 and seek to maximize wealth at a future time $T$. The price of each security today is known, given by $c_j$, $j = 1, \ldots, N$. Uncertainty is modeled by a measurable space $(\Omega, \mathcal{F})$ upon which a probability measure $P$ is placed. Let the price of security $j$ at the rollover date $T$ be given by the random variable $f_j(\omega)$. Thus the value of a portfolio containing $x_j$ units of security $j$ is worth:

$$\sum_{j=1}^{N} f_j(\omega) x_j$$  \hspace{1cm} (2.51)$$

43
at the rollover date. We also assume that the price of a security is independent of the amount $x_j$ of the security held in the portfolio.

According to the procedure outlined above, a portfolio is selected to serve as the market benchmark, against which performance will be measured. We shall denote the current price of this benchmark by $q$, and the random variable governing its value at the rollover date by $\tau$.

**Minimal Cost Replication: Optimal Hedging**

To find the minimal cost portfolio that exactly matches the behaviour of the benchmark, or target portfolio, one must solve:

$$\min_{x} c^T x$$

subject to

$$\sum_{j=1}^{n} x_j f_j(\omega) = \tau(\omega)$$

The linearity in the cost function suggests that an appropriate discretization scheme may lead to a linear programming formulation for the replication problem.

Liquidity constraints may also be used to limit the amount of each instrument one can hold. If the limits for the $j$-th instrument are $l_j$ and $u_j$ (recalling that short selling may be permitted), then the replication problem becomes:

$$\min_{x} c^T x$$

subject to

$$\sum_{j=1}^{n} x_j f_j(\omega) = \tau(\omega) \quad \forall \omega \in \Omega$$

$$l_j \leq x_j \leq u_j \quad j = 1, \ldots, m$$
Solving this problem yields a constructive strategy for replicating the benchmark. This is important in hedging applications. If the portfolio to be hedged is selected as the benchmark in the above problem, it provides the minimum cost hedging strategy. The strategy can be made increasingly realistic since the constraints can be used to model market imperfections such as liquidity constraints and bid-ask spreads.

The Optimal Investment Problem: Minimizing Regret

Assuming that regret can be written as an expectation, the problem to determine a minimum regret portfolio with respect to a benchmark \( \tau \) is given by

\[
\min_x \mathbb{E}[R(\sum_{j=1}^{n} x_j f_j(\omega); \tau(\omega))] \tag{2.57}
\]

subject to

\[
l_j \leq x_j \leq u_j \quad j = 1, \ldots, m \tag{2.58}
\]

The problem can be constrained to restrict the amount spent on the optimal portfolio to be no more than a given level \( C_0 \).

\[
\min_x \mathbb{E}[R(\sum_{j=1}^{n} x_j f_j(\omega); \tau(\omega))] \tag{2.59}
\]

subject to

\[
l_j \leq x_j \leq u_j \quad j = 1, \ldots, m \tag{2.60}
\]

\[
c^T x \leq C_0 \tag{2.61}
\]

This can be regarded as a parametric optimization model with parameter \( C_0 \). By varying the level of \( C_0 \), one can determine how regret is reduced as the optimal portfolio is allowed to be more expensive. Repeatedly solving the program for various levels of \( C_0 \) reveals the marginal decrease in regret with respect to portfolio cost.
Linear Programming Formulation of a Regret Minimization

Suppose that uncertainty at the rollover date $T$ is expressed by a finite number $S$ of possible future scenarios. Let $D$ be the matrix with entries $d_{ij}$, where $d_{ij}$ is the payoff of instrument $j$ under scenario $i$. Then a portfolio containing $x_j$ units of security $j$ will be worth

$$\sum_{j=1}^{n} d_{ij} x_j$$

if state $i$ prevails at the rollover date.

The vector giving the value of the portfolio $x$ under each scenario is therefore:

$$Dx$$

Let $\tau \in \mathbb{R}^S$ be the vector whose entries are the values of the benchmark at the rollover date under each of the prespecified scenarios. Then if $p_s$ is the probability of scenario $s$ occurring at time $T$ ($s = 1, \ldots, S$) the optimization problem becomes:

$$\min_{x_j, y^+_i, y^-_i} \sum_{i=1}^{S} p_i (y^+_i + y^-_i)$$

subject to

$$Dx - \tau = y^+ - y^-$$

$$l_j \leq x_j \leq u_j$$

$$y^+ , y^- \geq 0$$

The above model determines the minimum regret portfolio under a given set of constraints. By the inclusion of a parametric constraint representing a minimum required level of reward, it will reveal the interaction between risk and reward in the model.
2.3.4 Parametric Formulation: The Regret-Reward Model

In this section, we consider an extension of the problem (2.64) to include a constraint on the reward achieved by the optimal portfolio. The problem assumes a form similar to Markowitz model (2.15).

We consider the portfolio selection problem (2.59), where expected regret is minimized subject to feasibility restrictions on the portfolio. Let \( c \) be the vector of current prices of the market instruments. As before, we denote \( f = \sum x_j f_j \), and let \( q \) denote the price of the target portfolio.

To this point, we have focused mainly on how risk is formulated in the Scenario Optimization model through the notion of regret. Since portfolio selection models illustrate the dynamic relationship between risk and reward in markets, the model must also quantify the intuitive notion of reward. The program will retain its linear structure if reward is taken to be the *expected excess profit* achieved by the portfolio over the benchmark.

The expected profit of the target realized at the rollover date is given by:

\[
E[(\tau(\omega)) - q]
\]  

(2.68)

Whereas the expected profit realized by the replicating portfolio is:

\[
\left( \sum_{j=1}^{n} E[f_j(\omega)] - c^T x \right) - (E[\tau(\omega)] - q)
\]

(2.69)

Investors seek to outperform a benchmark by at least a certain amount \( K \), at minimum regret. This amounts to the inclusion of an additional constraint in the optimization model. Specifically, problem (2.59) becomes:

\[
MR(K) = \min_{x} \rho(f; \tau)
\]  

(2.70)

subject to

\[
f \in X
\]

(2.71)

\[
\left( \sum_{j=1}^{n} x_j E[f_j(\omega)] - c^T x \right) - (E[\tau(\omega)] - q \geq K)
\]

(2.72)
where MR stands for the minimum regret attainable under the given constraints.

The solution of the above problem gives, for each level of required expected excess profit $K$, the portfolio with minimum regret. The construction is similar to the mean-variance formulation of Markowitz, where the efficient frontier gives the minimum attainable variance for a specified level of return.

Denote by $\pi(f; \tau)$, the excess profit of a portfolio $f$, with respect to the benchmark $\tau$

$$\pi(f; \tau) = \sum_{j=1}^{n} x_j E[f_j(\omega)] - c^T x - E[\tau(\omega)]$$  \hspace{1cm} (2.73)

Then, by using a penalty function method, the optimization problem (2.70) can be written as

$$MR(K) = \min_{x} \rho(f; \tau) - \lambda \pi(f; \tau)$$  \hspace{1cm} (2.74)

subject to

$$f \in X$$  \hspace{1cm} (2.75)

which is analogous to the problem (2.15) in the mean-variance portfolio theory. The parameter $\lambda > 0$ represents the relative importance of regret and reward, and can be seen as a measure of an investor's risk aversion.

2.3.5 The Regret-Reward Efficient Frontier

The parametric optimization problem (2.70) defines the minimum regret $MR$ as a function of required over-performance $K$. This function provides the minimum regret attainable for each specified level of excess over performance. It is the analogue in the regret-reward model to the Markowitz mean-variance frontier, and is therefore referred to as the regret-reward efficient frontier. We now examine how the qualitative properties of the frontier reveal quantitative facts about the underlying market structure.

Recalling from the first chapter the fundamental role played by completeness of the market in finance, it is not surprising that the $x$-intercept (level of
perfect replication, or zero regret) of the frontier is of the utmost importance. This point gives the smallest possible level of excess profit at which the target can be perfectly replicated under all scenarios. This has a close relationship to the concept of arbitrage. Here, we introduce a generalization of the standard arbitrage concept given in the first chapter, and explain how this relates to the completeness of the market as illustrated by the risk-return frontier in financial markets.

**Benchmark Arbitrage**

A portfolio \( x \) is an arbitrage with respect to the benchmark portfolio \( \tau \), if

\[
\rho(x; \tau) = 0, \quad \pi(x; \tau) > 0
\]  

A zero value for the objective function in (2.70) indicates that there exists a feasible portfolio \( x \) that performs at least as well as the target under each of the scenarios. By the standard definition, \( x \) will be an arbitrage if its price is strictly less than that of \( \tau \), or if \( x \) has a price equal to \( \tau \), but strictly outperforms \( \tau \) in at least one of the scenarios. The portfolio is a benchmark arbitrage with respect to \( \tau \) if it outperforms \( \tau \) under every scenario and has a higher expected profit. We shall now examine the relation between these two notions of arbitrage in the discrete market considered above.

First of all, suppose that there exists an arbitrage in the \( n + 1 \) instruments \( x_0, \ldots, x_n \). Then there is an arbitrage with respect to the null portfolio, that costs 0 and pays off nothing under all scenarios. The arbitrage portfolio \( x^* \) yields:

\[
D^T x^* \geq 0
\]  

and

\[
c^T x^* < 0
\]  

The first of these equations yields that \( x^* \) has zero regret. Substituting the second into the return constraint in the optimization problem (2.70) gives:

\[
p^T D^T x - c^T x \geq K > 0
\]  

49
which holds for all sufficiently small $K$. Thus, the existence of an arbitrage in the market of tradeable instruments implies the existence of an arbitrage with respect to the null benchmark.

Ideally, we would like to show that the two definitions of arbitrage are equivalent. We now suppose that there exists an arbitrage with respect to the benchmark $\tau$. That is, there is a portfolio $x^*$ such that

$$D^T x^* - \tau \geq 0$$

and

$$(p^T D^T x^* - c^T x^*) - (p^T \tau - q) \geq K > 0$$

For the portfolio to be an arbitrage, it is required that $c^T x^* - q < 0$. But by (2.81)

$$c^T x^* - q < (p^T D^T x^* - p^T \tau) - K$$

Let $K^*$ be the maximal $K$ for which the above inequality holds. Then a necessary and sufficient condition for $x^*$ to be an arbitrage with respect to $\tau$ is that

$$p^T D^T x^* - p^T \tau \leq K^*$$

Graphically, benchmark arbitrage opportunities are easy to identify. The regret-reward frontier generated by the optimization with benchmark $\tau$ indicates that there exists an arbitrage with respect to the portfolio $\tau$ if and only if it crosses the at a level of strictly positive reward ($K_i > 0$). The completeness of markets arguments of the first chapter yield that if the regret-reward frontier passes through the origin, and is strictly increasing at the origin (i.e. $MR(0) = 0$ and $MR(K) > 0$ for all $K > 0$), then there is no arbitrage in the market. A final possibility is that perfect replication of the target may not be achievable without expected underperformance of the benchmark. We shall illustrate each of these situations with a simple example.
Example: Perfect Replication of the Benchmark

To illustrate characteristics of a complete market with no benchmark arbitrage, we use the standard financial relationship of put-call parity. That is, we replicate a put option on a stock with a call option (on the same underlying and having the same strike price and time to expiration), the underlying, and cash. The put-call parity equation

\[
P + S = Ke^{-r(T-t)} + C
\]  
(2.84)

indicates that a perfect replication will always be possible in this market. In this section, we solve the following risk-reward model:

\[
\min \| D^T x - \tau \|_1
\]  
(2.85)

subject to

\[
p^T (y^+ - y^-) - q^T x \geq K - c
\]  
(2.86)

Consider a market where the investor can trade in only securities, a stock, a European call option on that stock, and a riskless bond. Assume that the stock has current price 100, the call option has a strike price of 100. Furthermore, both the bond and the option mature in one year. The current price of the bond is 90, and it pays off 100 with certainty. The current price of the option is 12. The only stochastic variable in the model is the price of the stock. Since the payoff of the option is derived from the stock price, and the bond is riskless, the stock price will completely specify the payout structure for the market. We assume that the stock can assume the values 80, 100, 120 each with probability 1/3 in one year from today. The observant reader will notice that this provides us with three independent instruments under three scenarios, and hence that the market will be complete. There will be no arbitrage if and only if there exists a price deflator (risk-neutral measure). More formally, the payoff matrix for our model has the form:

\[
\begin{pmatrix}
100 & 100 & 100 \\
80 & 100 & 120 \\
0 & 0 & 20
\end{pmatrix}
\]  
(2.87)
where the first row represents the payoff of the bond, the second row the payoff of the stock, and the third the payoff of the call option. Notice that the payoff matrix $D$ has full rank. From the above information, we also know that the price vector for the market is given by

\[
\begin{pmatrix}
90 \\
100 \\
12
\end{pmatrix}
\]  

(2.88)

The theory presented in the first chapter tells that there is no arbitrage in the model if and only if there exists a vector $\psi$ in $\mathbb{R}^3_+$ such that $D\psi = q$. Since $D$ is of full rank, we simply solve $\psi = D^{-1}q$ to obtain:

\[
\psi = \begin{pmatrix}
0.1 \\
0.2 \\
0.6
\end{pmatrix}
\]  

(2.89)

This leads to a vector of risk neutral probabilities

\[
\alpha = \begin{pmatrix}
1/9 \\
2/9 \\
2/3
\end{pmatrix}
\]  

(2.90)

Suppose now that we are interested in pricing a simple put option on the stock, with strike price 100, expiring one year from today. The incorporation of this security into the market will add the row $\beta = (20, 0, 0)$ to the payoff matrix $D$. Let the matrix obtained by adding this additional row to $D$ be denoted by $\tilde{D}$. Let $c$ be the current price of the put option. We know by the put call parity relationship (as well as by the completeness of the market $D$) that the payoff of the put option can be reconstructed by trading only in the call, the stock and the riskfree bond. In order to preclude arbitrage opportunities, the cost of constructing this replicating strategy must be equal to the current cost of the put option. Alternatively, the price of the put option must be its expected payoff under the risk-neutral measure discounted to the current time (i.e. multiplication by the price deflator). Mathematically, this simply amounts to

\[
\beta \psi = c
\]  

(2.91)

Therefore we have that the arbitrage free price of the put option is $c = 2$.

We now consider a small set of examples that show how the regret-reward model reveals the structure of the simple market described above.
First, we solve the regret-reward model (2.70) with the put at its arbitrage free price of 2. This yields the following regret-reward frontier:

![Graph](image)

The fact that the regret-reward frontier passes through the origin shows that the payoff of the put can be replicated in the market. Since the slope of the regret-reward frontier is positive at the origin, there is no arbitrage with respect to the benchmark put option.

**Example: Existence of Benchmark Arbitrage**

Now let us suppose that the put option is overpriced. More specifically, we assume that we are currently able to sell the put option in the market for 4 units of value. From the discussion above, it is clear that the put can be replicated with only the bond, the stock and the call option, at a price of 2. Thus, arbitrage profits are possible by selling the put option in the market for 4 and then purchasing the replicating portfolio in the market. This provides an immediate cash payout of 2, and as deterministic payoff of zero in the future. The regret-reward frontier for the model where the put option has price 4 is simply a shift of the frontier for the model when the
option has price 2:

Notice that the regret-reward frontier crosses the x-axis (line of zero regret) at a positive level of reward. Thus positive reward is attainable at a level of zero regret. That is, there is a benchmark arbitrage in the market with respect to the put option.

**Example: UnderPriced BenchMark**

We now consider the final case, where the benchmark put option is priced less than its arbitrage free price of 2. For the sake of concreteness, suppose that the put option has current price 1. At first glance, the situation may seem exactly the same as above, i.e. arbitrage profits should be possible by shorting the replicating portfolio and purchasing the benchmark. However, this is not exactly the case, since in the regret-reward model the objective is to outperform the benchmark with the replicating instruments. The situation is clarified by examining the regret-reward frontier (again, a shift of the previous frontiers, owing to the simple nature of our example).
Using the tradable instruments, it is not possible to replicate the payoff of the benchmark with zero regret! Since the frontier crosses the zero regret axis, we know that replication of the put’s payoff is possible (recall that our market is complete), however, since the put is selling for a discount in the market, this replication is not possible unless reward is strictly negative. In a sense, the underpriced benchmark dominates the tradeable instruments in the market. In other words, once investor are allowed to assume long positions in the benchmark portfolio, arbitrage opportunities are available in the market.

2.3.6 Duality, Sensitivity Analysis and Implied Scenario Probabilities

One of the benefits provided by the linear programming formulation of the Scenario Optimization model is that it allows us to take advantage of the well established duality theory of linear programming. Let us consider the regret-reward model:
subject to
\[ Y^+, Y^- \leq 0 \] (2.93)
\[ \begin{align*}
&(p^T y^+ - q^T x) - (p^T y^- - c) \\
&\geq K 
\end{align*} \] (2.94)
\[ y^+, y^- \geq 0 \] (2.95)

The dual of the above linear program is:
\[
\max_{\pi, \lambda} \pi^T x + (K - c)\lambda 
\] (2.96)
subject to
\[ D\pi - \lambda q = 0 \] (2.97)
\[ 0 \leq \pi - \lambda p \leq p \] (2.98)
\[ \lambda \geq 0 \] (2.99)

where the variables in brackets are the primal variables corresponding to the dual constraints. One of the fundamental results in the theory of linear programming states that the primal problem (2.92) has an optimal solution if and only if the dual problem (2.96) has an optimal solution, and in this case the optimal values coincide. When the linear program is solved by the simplex method, the solution of the primal yields the solution of the dual, and vice-versa. The optimal dual variables provide important sensitivity information on the primal solution.

Specifically, let us consider the marginal risk that an investor must assume in order to obtain a given level of reward. Let \( x^*, (y^+)^*, (y^-)^* \) be the optimal solution to the problem (2.92) with reward level \( K \), and let \( \lambda^*, \pi^* \) be the optimal solution to the corresponding dual problem. How much additional
regret is it necessary to assume in order to obtain a profit level of \( K + \Delta K \). From the duality theory of linear programming, we find that for small enough \( \Delta K \), this amount is

\[
\pi^* \Delta K
\]

(2.100)

The following observation has been made by Dembo [20]. Based on the fact that the assumed probabilities \( p \) only appear in the dual constraint:

\[
0 \leq \pi - \lambda p \leq p
\]

(2.101)

corresponding to the "tracking errors", \( y^+, y^- \), we see that the sensitivity of the linear program to changes in the input probabilities is intimately tied up with the behaviour of the tracking errors. A brief recollection of the material from the first chapter shows that this is not surprising. The vector \( p \) represents the empirical measure. The values \( y^+, y^- \) relate to whether or not the contingent claim \( \tau \) can be replicated, i.e. to completeness of the market. Given the economic connection between completeness of markets, and the redundancy of the empirical measure (aside from determining the null sets), we should seek a similar relationship in our optimization framework.

Suppose that the dual constraint (2.98) is not binding. Then small changes in the input probabilities will not affect the solution of the linear programming problem (2.96). By the complementary slackness conditions of linear programming, if the constraint (2.98) is not binding at the optimal solution, then it must be the case that \((y^+)^* = (y^-)^* = 0\), i.e. it is possible to obtain a perfect replication of the benchmark. If constraint (2.98) is not binding at the optimal solution for any \( K \), then the regret-reward results will be independent of the probability weights \( p \). Let \((\lambda, \pi)\) be any feasible solution to the dual problem (2.96), with \( \pi > 0 \). Then by the dual feasibility, and the constraint (2.97), we have

\[
D(\pi/\lambda) = q
\]

(2.102)

That is, \( \pi/\lambda \) is a price deflator for the market consisting of the replicating instruments with payoff matrix \( D \), and the price vector \( q \). Thus, if we denote \( \alpha = \pi/\lambda \), then the vector:

\[
\rho = \frac{\alpha}{\|\alpha\|_1}
\]

(2.103)
bears the interpretation of state probabilities. Notice that if \( \pi > 0, \lambda > 0 \) means that \( \rho > 0 \), i.e. \( \rho \) and \( p \) are equivalent (in the measure theoretic sense). The standard result from the first chapter tells us such a vector exists if and only if there is no arbitrage in the market. We now illustrate this result in the Scenario Optimization regret-reward framework.

According to the theory of the first chapter, there will be no arbitrage among the replicating instruments and the benchmark \( \tau \) iff there exists a vector \( \psi > 0 \) such that

\[
\begin{bmatrix} D & \tau \end{bmatrix}^T \psi = [q \ c]^T \quad (2.104)
\]

Since any feasible solution to the dual problem with \( \pi > 0 \) and \( \lambda > 0 \) (henceforth referred to as a strictly positive dual solution) yields the the vector \( \alpha = \lambda/\pi \) such that

\[
D\alpha = q \quad (2.105)
\]

the only question that remains is if for any of these \( \alpha \), it is the case that

\[
\tau^T \alpha = c \quad (2.106)
\]

Thus, there is no arbitrage in the market iff there exists a strictly positive dual feasible solution \( [\lambda \ \pi] \) such that \( \tau^T (\lambda/\pi) = c \).

As a preview to the (more in depth) study of the implied scenario probabilities \( \alpha \) in the third chapter, we present the probabilities that arise from the dual variables in the model that we examined in the previous section:
Observe that these correspond exactly to the probabilities that we obtained from the single period theory that was used to introduce the model.

2.4 Directions For Future Research

We identify the following as important directions of future research on Scenario Optimization and portfolio selection:

- Analysis of various scenario generation techniques, both through extensive numerical testing, and the determination of error bounds.

- Analysis of the sensitivities of the model to perturbations in its various inputs.

- The inverse problem of determining investor preferences based on portfolio holdings, in various optimization settings (although see the work of Dybvig and Rogers [31])

- Resolution of the difficulties arising from transactions costs in the continuous time framework (stochastic control) setting (although see the recent work of Shreve and Soner [77]).
• Rigorous analysis of the errors occurring in the Scenario Optimization approximation (2.45). Examination of the convergence of the discrete to continuous case.
Chapter 3

Implied Probabilities in Finance
Part I

The Probabilistic Implications of Option Prices
3.1 Introduction

The principal theme of this thesis is the relationship between risk and reward in financial markets, and how this relationship manifests itself in the mathematical models used for derivative securities pricing and portfolio selection. The first chapter examined how derivative securities prices can be determined using only arbitrage considerations, and assumptions on the stochastic behaviour of the underlying securities. In the second chapter, we outlined the problem of selecting the optimal portfolio for a specific investor given prices, and the stochastic processes for the tradeable securities, exogeneously. In this chapter, we examine the problem of inferring the stochastic process followed by market instruments based on derivatives prices. This is the inverse of the arbitrage pricing problem. In arbitrage pricing, stochastic processes are assumed, and prices are then derived based on these processes. In reality, prices are known, while probabilistic laws are unobservable, i.e. the situation faced by market participants is exactly the opposite of that assumed by arbitrage pricing. Many derivative securities prices are readily observable, but little is known about the stochastic law governing the behaviour of market instruments. This chapter examines the attempts to ascertain information about this process based on the information available from market prices.

3.2 The Black-Scholes Model

The Black-Scholes partial differential equation, provides a technique for pricing any derivative security whose value depends on the level of a single underlying market variable. Considering the popularity of the model, it is important to determine how well it corresponds to the behaviour of market instruments. A number of empirical studies testing the validity of the Black-Scholes model have been performed, and significant departures from the model predictions have been observed (see for example Rubinstein [71] [72]).

3.2.1 Market Deviations from the Black-Scholes Model

The most obvious explanation for the deviation of the behaviour of market instruments from the Black-Scholes price would be that the assumptions
regarding financial markets employed in construction of the riskless portfolio are too restrictive. Actual markets have transactions costs and taxes, which are not accounted for in the replication methodologies described in the first chapter. Trading occurs at discrete intervals in time, and is subject to significant liquidity and divisibility constraints. Despite these objections, most research has focussed on replacing the assumption that stock prices follow a geometric Brownian motion:

\[ dS = \mu S dt + \sigma dW \]  

(3.1)

with some alternative stochastic process. This new stochastic process should be chosen to provide a mathematical description of the random behaviour of the underlying that leads to derivatives prices in closer correspondence with those observed in the market.

3.2.2 The Implied Volatility Surface

The Black-Scholes formula for pricing European options takes as inputs the current stock price, the risk free interest rate, stock volatility, and option strike price and exercise date, providing as output a single option price. In practice, all the input variables to the Black-Scholes model are observable in markets except for the stock volatility. What is observable in markets is the actual option price. By solving a nonlinear equation, it is possible to determine the value \( \sigma^* \) for the volatility parameter that, when used as input to the Black-Scholes formula, will reproduce the observed market price. The value \( \sigma^* \) is referred to as the option’s Black-Scholes implied volatility (or simply its implied volatility, if use of the Black-Scholes model is understood). If the stock follows the stochastic process (3.1), then according to the Black-Scholes model, each option on the same underlying should yield the same implied volatility. In reality, options markets depart significantly from this prediction. Black-Scholes implied volatilities for options on the same underlying vary with both strike price and time to expiration. The graph of the Black-Scholes implied volatility against option strike price for any fixed time to maturity is referred to as the volatility smile. The relation between time to expiration and implied volatility is referred to as the volatility term structure. The collection of smiles for all maturities (or,
equivalently term structures for all strike prices), gives implied volatility as a function of two variables: time to expiration and strike price, and is referred to as the implied volatility surface.

The Black-Scholes analysis can be extended to allow both the risk-free interest rate and the volatility to be known functions of time. This was first demonstrated by Merton [62], and leads to the slightly more general stochastic process for the stock level:

\[ dS = r(t)dt + \sigma(t)dW \]  \hspace{1cm} (3.2)

If implied volatility does not vary with strike price (i.e. the smile is flat), such a stochastic process can be chosen so as to fit any specific volatility term structure.

### 3.2.3 Historical Behaviour of the Volatility Smile

During the more than twenty years since its introduction, the Black-Scholes model has been the subject of extensive empirical testing. The tests have generally shown that, though markets never behaved exactly in accordance with the predictions of the Black-Scholes framework, in recent years deviations from the model have become much more pronounced.

From the perspective of understanding the implied volatility surface, the most interesting studies have been carried out by Rubinstein [71] [72]. In the earlier study, a non-parametric mini-max estimator was used to test the null hypothesis that the Black-Scholes model holds on an extensive database of equity options traded on the Chicago Board Options Exchange during the late 1970's. Rubinstein found that the data conveniently separated into two periods. During the first period (August 13, 1976 to October 21, 1977) the volatility smile showed Black-Scholes implied volatility as a decreasing function of the strike price, whereas in the second period (October 24, 1977 to August 31, 1978) the smile had a positive slope, yielding implied volatility as an increasing function of strike. The general shapes for the smile, and the pronounced change occurring between the two periods held for nearly all options on the 30 equities considered in the study.
Rubinstein [72] carried out a similar study on options on the S&P 500 index. The index is an aggregation of the values of various equities, representing a cross section of the market. As such, it is less likely to experience jumps than any individual stock (there is a smoothing effect from averaging). It has been postulated that stock jumps are caused by the arrival of new information regarding a particular company or industry on the market (consider the recent Bre-X fiasco). Rubinstein's mini-max statistic showed that in 1986 only small deviations from the Black-Scholes model (as measured by parity of the implied volatilities of options on the same underlying and with the same maturity date) were present in the index options. However, during 1987 the departure from the Black-Scholes flat volatility smile became much larger. This deterioration accelerated from 1988 through 1992. Recent index options data consistently exhibits a significant negative slope in the volatility smile. Rubinstein suggests that this might be explained as a reaction of traders to the stock market crash of October 19, 1987. Fear of a future crash leads to much higher prices for out-of-money put options (those with relatively low strike prices—these options stand to make significant profits in the event of a market crash). The consistency of recent deviations from the predictions of the Black-Scholes model, is the prime motivation for researchers to question its lognormality assumptions, seeking alternative probability measures more in accordance with market data. In this chapter we examine attempts to forego the exogenous specification of a stochastic process for the underlying, in favour of an approach that seeks to infer information on the underlying distribution from market prices.

3.3 Alternative Specifications for the Underlying Process

Rubinstein [72] identifies a number of ways in which the Black-Scholes model can be extended in order to better fit empirical data. We present a slightly modified version of this list here, examining briefly the progress that has been made in each type of modification.

- Volatility can be made to be a function dependent on time. $\sigma = \sigma(t)$
- Volatility can be made to depend on both time and the level of the underlying. \( \sigma = \sigma(S,t) \)

- An untraded source of risk can be incorporated into the model. Volatility can be taken to be stochastic \( d\sigma^2 = \alpha \sigma^2 dt + \xi \sigma^2 dZ \), or the process be assumed to have a jump component.

- Market imperfections such as transaction costs, liquidity constraints, or taxes can be introduced.

Almost immediately following the publication of the original Black-Scholes result, Merton [62] was able to show that volatility could be made time dependent without altering any of the essential arguments. If volatility is a known function of time, then the correct option price is obtained by using the Black-Scholes formula with \( \sigma \) replaced by the mean volatility \( \bar{\sigma} \) over the life of the derivative:

\[
\bar{\sigma}^2 = \frac{1}{T-t} \int_t^T \sigma^2(s) \, ds
\]  
(3.3)

The second proposed modification involves allowing volatility to be both time and state dependent. Models of this type are an area of active research, and the main subject of this survey. In fact, we shall show that allowing volatility to be both time and state dependent allows the model to fit perfectly the prices of all traded options. Furthermore, if the complete implied volatility surface is known, then there is a unique function \( \sigma(S,t) \) such that the process

\[
dS = r(t)Sdt + \sigma(S,t)SdW
\]  
(3.4)

prices correctly all observed options. Such extensions are important in practice, because standard option pricing models allow only the choice of a single volatility parameter. When pricing a derivative whose value is not observable in the market, (for example, certain exotic options), it is not obvious which of the many volatilities implied from traded options to use in the Black-Scholes equation. Allowing for more general forms of volatility function enables us to construct discrete approximations to the process (3.4) which then incorporate all the information in the implied volatility surface, and make the question of which implied volatility to use as the coefficient in the Black-Scholes equation redundant.
The existence of the volatility smile in options data, together with other intuitions of the general behaviour of stock prices led researchers to consider alternative probabilistic assumptions to the lognormality given by the process (3.1). Traditional research began by assuming a stochastic process for the underlying variable and then determining prices based on that process. Recent work has focussed on determining the probabilistic information that can be construed from market data. This follows closely the lines of a similar development in modelling the process for the stochastic behaviour of interest rates.

3.4 Evolution of Stochastic Interest Rate Models

The traditional approach to option pricing described above involves specifying a stochastic process followed by the underlying, and then using arbitrage or equilibrium arguments to determine the prices implied by that stochastic process. Until 1986, a similar approach was taken to modelling the stochastic behaviour of interest rates. Specifications of the stochastic process for the short rate were made endogenously within the model. Examples of such processes, with \( r \) denoting the short rate are:

\[
dr = a(b - r)dt + \sigma dW
\]  

(3.5)

and

\[
dr = a(b - r)dt + \sigma \sqrt{r}dW
\]  

(3.6)

the first model was developed by Vasicek, the second by Cox, Ingersoll and Ross [15]. The above models have many nice equilibrium properties, but because they endogenize the term structure, it was found that they could not reproduce the prices of liquid bonds observed in bond markets. This is analogous to the situation in index options markets encountered above, where the prescribed stochastic processes could not reproduce the observed prices of European call options. If the purpose of the model is to determine appropriate prices for interest rate derivatives, then it should at least price bonds so that there are no arbitrage opportunities in the most liquid markets. Ho and Lee [38] made the fundamental observation that it is possible
to model the short rate in a way consistent with the observed term structure of interest rates. Ho and Lee presented their model as a discrete time binomial tree. In the continuous time limit, their binomial process assumes the form

\[ dr = \theta(t)adt + \sigma dW \]  

where the function \( \theta(t) \) is time dependent, and chosen in such a way that the model fits the observed yield curve. A great deal of research in interest rate derivatives over the past 11 years has gone into generalizing the Ho-Lee model, and determining effective numerical methods for such discrete time approximations (see, for example, Hull [39]).

Recently, analogous developments have taken place in the pricing of index options. Researchers have sought single factor diffusion processes that correctly reproduce the prices of traded market instruments. In order to do so it proves necessary to make the volatility depend on both time and the level of the underlying security.

### 3.5 Equilibrium Models

We shall illustrate the traditional alternatives to the Black-Scholes formulation with two of the most popular models: Merton's mixed jump-diffusion process, and the Hull-White stochastic volatility model. Both these approaches have important intuitive advantages over the process assumed by Black and Scholes, and both have the drawback of introducing an untraded source of risk in the model. The existence of an untraded source of risk is an important disadvantage in any theoretical pricing framework. This is because untradeable risk cannot be "hedged" away in order to eliminate uncertainty (in the language of the first chapter, it cannot form part of a self-financing replicating portfolio).

#### 3.5.1 The Capital Asset Pricing Model

According to the CAPM, the risk of a portfolio can be divided into two distinct types. The first type is the risk which comes from the correlation of the portfolio with the market and is referred to as systematic risk. The
second type of risk, referred to as nonsystematic, cannot be diversified away by trading in the market. An example would be the risk in oil prices from political events. The Capital Asset Pricing Model posits that investors only require compensation for systematic risk. The systematic risk of a portfolio \( p \) is summarized by its "beta", defined to be its correlation coefficient with the market portfolio \( m \):

\[
\beta_p = \frac{\sigma_{pm}}{\sigma_p \sigma_m}
\]  (3.8)

Thus, according to the CAPM, a portfolio with no correlation to the market portfolio (zero beta) should earn an expected return equal to the risk free interest rate. It is this feature of the model that is employed to derive prices in the Mixed Jump-Diffusion and Hull-White stochastic volatility models.

### 3.5.2 Mixed Jump-Diffusion Model

Merton [63] examined option pricing in a framework where the underlying follows a mixture of a diffusion and a Poison process. In this situation, the stock typically varies according to a standard diffusion process (geometric Brownian motion), but occasionally takes a jump, corresponding to the occurrence of a "Poisson event". Denote the random variable governing the size of the jump conditional on the occurrence of a Poisson event by \( Y \). If a Poisson event occurs in the interval \([t, t + dt]\) then the stock jumps from \( S(t) \) to

\[
S(t + dt) = S(t)Y
\]  (3.9)

Let \( k = E(Y - 1) \). The stochastic differential equation describing the jump diffusion process is:

\[
dS = S(\mu - \lambda k)dt + \sigma SdW + Sdq
\]  (3.10)

where \( \mu \) is the instantaneous expected return on the stock, \( \sigma^2 \) is its instantaneous variance, conditional on the nonoccurrence of a Poisson event, \( \lambda \) is the Poisson parameter, and \( q \) is a Poisson process. It is assumed that \( q \) and \( Z \) are independent.
The probability of a single Poisson event occurring in the period \((t, t + dt)\) is equal to \(\lambda dt + o(dt)\), the probability of no Poisson events occurring in \((t, t + dt)\) is \(1 - \lambda dt + o(dt)\) and the probability of more than one Poisson event occurring in \((t, t + dt)\) is \(o(dt)\).

According to Merton [63], Poisson events can be interpreted as representing the arrival of new information regarding the underlying, which can cause a sudden jump in its price. The possibility of Poisson jumps means that the delta-hedged portfolio used in the Black-Scholes analysis is no longer instantaneously riskless. In general, this means that the value of the option will no longer be independent of the expected return of the underlying, or individual investor’s risk preferences. As an alternative to the direct use of risk preferences, Merton employs the Capital Asset Pricing Model (CAPM) to derive an equation for the value of a derivative. Under the interpretation of the Poisson events as the arrival of new information regarding the underlying, and the assumption that \(q\) and \(Z\) are independent, the Poisson events represent only nonsystematic risk. Thus, for the delta-hedged Black-Scholes portfolio, it is still the case that \(\beta_p = 0\). Using an argument similar to the Black-Scholes analysis, it is possible to show the price \(f\) of any derivative security whose value depends only on the level of \(S\) must obey:

\[
rf = \frac{\sigma^2}{2} S^2 \frac{\partial^2 f}{\partial S^2} + (r - \lambda k) S \frac{\partial f}{\partial S} + \frac{\partial f}{\partial t} \lambda + E(f(SY, t) - f(S, t)) \tag{3.11}
\]

subject to the boundary conditions \(F(0, t) = 0\), \(F(S, T) = (S - K)_+\). Notice that in this case the value of the derivative security is independent of the expected return \(\mu\), but does depend on the Poisson probability \(\lambda\) and the random variable \(Y\).

The existence of a positive probability for Poisson events means that the prices of deep in the money and deep out of the money options are higher than their corresponding Black-Scholes values. Thus, prices obtained using the mixed jump-diffusion model will produce a volatility smile where the tails are higher than the middle. The appearance of this shape in actual market data gave the volatility smile its name.

Merton gives a useful intuitive explanation for why the existence of positive probabilities for Poisson events may give rise to smiles of this shape op-
tions markets. For deep out of the money options, the probability of the option expiring in the money near expiration is very small. The existence of Poisson events gives rise to a positive probability that the stock could “jump” up (in the case of a call option) making the option expire in the money (the analogous probability of a downward jump increasing the value of a put). Similarly, consider an option that is deep in the money. Since exercise is almost certain, holding the option is virtually equivalent to holding the stock (in the case of a call option, for a put, the option is almost the same as “holding” -1 units of the stock). However, the option provides some “insurance value” in the sense that if there is an extreme movement in the level of the underlying, option holders can only lose their initial investment (which could be significantly less than the change in the value of the underlying itself). The existence of the Poisson portion of the stochastic process increases the probability of such a calamitous stock move. Thus, the insurance value of a deep in the money option is increased if we assume that the stock follows the stochastic process (3.10) rather than a standard diffusion process.

3.5.3 Stochastic Volatility Model

The Black-Scholes model assumes that volatility is constant, or at least a known function of time. Practical experience in options markets (see Hull [39]) led to the hypothesis that the volatility follows a stochastic process of its own. Hull and White [40] consider a model where the instantaneous variance of the process follows a Geometric Brownian Motion:

\[ dS = \mu S dt + \sigma S dW \]  
\[ dV = \alpha V dt + \xi V dZ \]

where \( V = \sigma^2 \), and \( Z \) is a standard Wiener process. If \( Z \) and \( W \) are uncorrelated, and the volatility has no systematic risk, then under the assumptions of the Capital Asset Pricing Model, it can be shown that any derivative security on \( S \) will follow the equation:

\[ rf = \frac{\partial f}{\partial t} + \frac{1}{2} \left[ \sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} + \xi^2 \frac{\partial^2 f}{\partial V^2} \right] + rS \frac{\partial f}{\partial S} + \alpha \sigma^2 \frac{\partial f}{\partial V} \]  

(3.14)
which resembles the Black-Scholes equation with some “adjustment” terms coming from the stochastic nature of the volatility of the underlying. Notice that the above equation is independent of the instantaneous expected return on the underlying, \( \mu \), but that it does depend on the instantaneous drift of the volatility \( \alpha \), and its instantaneous volatility, \( \xi \). These parameters are not directly observable in the market, and must be estimated from empirical data, or implied by market prices. Hull and White use the CAPM to derive the price of a European call option on \( S \) where \( S \) follows (3.1). The solution assumes the form of a series expansion in terms of the mean of the underlying variance, \( \bar{V} \) over the life of the derivative security.

\[
\bar{V} = \frac{1}{T-t} \int_t^T \sigma^2 \, d\tau 
\]  

(3.15)

In the case where \( Z \) and \( W \) are correlated, with coefficient of correlation \( \rho \), equation (3.14) assumes a slightly more complicated form:

\[
r_f = \frac{\partial f}{\partial t} + rS \frac{\partial f}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} + \alpha \sigma^2 \frac{\partial f}{\partial V} + \rho \sigma^2 \xi \frac{\partial^2 f}{\partial S \partial V} + \frac{1}{2} \xi^2 V^2 \frac{\partial^2 f}{\partial V^2}
\]  

(3.16)

When the correlation coefficient \( \rho \) is positive, volatility tends to be higher at higher levels of \( S \), and lower at lower levels of \( S \). This leads to a volatility smile with a negative slope, i.e. implied volatility as a decreasing function of strike price. On the contrary, when \( \rho < 0 \), implied volatility will tend to vary negatively with the level of the underlying. Notice that the implied volatilities discussed above are Black-Scholes implied volatilities, obtained by assuming that \( \sigma \) is constant rather than stochastic. The convention typically employed, even when volatility is assumed to be a more general function \( \sigma = \sigma(S,t) \) is that the “implied volatility” of a given option is still the constant volatility that would make its market price equal to the Black-Scholes price.

The main drawback of models which consider the volatility to be stochastic is that they can no longer employ arbitrage pricing, and must rely on the CAPM, or some other assumptions regarding investor preferences in order to price derivatives. Recent research in stochastic volatility models has
attempted to extend pricing to an arbitrage based methodology, using the concept of forward volatility (see e.g. Dupire [28], Derman and Kani [22]) and the area remains one of active research.

3.6 No-Arbitrage Models

3.6.1 The Single Factor Diffusion Setting

In this section, we examine attempts to generalize the stochastic process \((\cdot)\), while remaining in a single factor diffusion setting. Since the drift of the process is irrelevant (see chapter 1) for pricing applications, it is the volatility coefficient that is of prime importance for option pricing. The importance of remaining in a single factor diffusion setting comes from the fact that the arguments in the Black-Scholes section of the first chapter remain valid for any one dimensional diffusion. That is, the Feynman-Kac formula can still be applied to obtain a partial differential equation that must be satisfied by any derivative security whose value depends upon the level of a single underlying security. The only difference is that the constant \(\sigma\) in the Black-Scholes equation is replaced by a function \(\sigma(S, t)\).

The Breeden-Litzenberger Result

Breeden and Litzenberger [10] demonstrated that if the prices for options with a continuum of strike prices and a single time to expiration are known, then there exists a unique probability distribution for the value of the underlying at expiration, conditional on its current level, that prices all options correctly. This is equivalent to saying that if the volatility smile is known at any time (remember that there is a one-one correspondence between option price and implied volatility), then the probability distribution for the underlying is also known. All options in the following discussion are taken to be European. In order to prove this result, consider the risk-neutral valuation technique discussed in the introductory chapter. We also require the technical assumption that the value of the call option be twice continuously differentiable as a function of the strike price. If we discount at the constant rate \(r\) then we obtain, for an option with exercise price \(K\), expiring at time
\( T \), (where the current time is \( t \)):

\[
C(S_t, K) = D(T)E[(S_T - K)_+ | \mathcal{F}_t] 
\]

\[
= D(T) \int (S_T - K)_+ dQ 
\]  

(3.17)

where \( D(T) = \exp[- \int_t^T r(\tau) d\tau] \) \( Q \) is the risk-neutral probability measure (martingale measure) governing the behaviour of \( S_T \) conditional on \( S_t \) assuming its current value. Assume that \( Q \) has a density representation, \( f(S_t) \):

\[
C(S_t, K) = D(T) \int_{-\infty}^{\infty} (S_T - K)_+ f(S_T) dS_T 
\]  

(3.19)

differentiating both sides of the above equation with respect to \( K \), we obtain

\[
\frac{\partial C}{\partial K} = -D(T) \int_K^\infty f(S_T) dS_T 
\]  

(3.20)

differentiating with respect to \( K \) once more gives

\[
\frac{\partial^2 C}{\partial K^2} \bigg|_{K=S_T} = D(T) f(S_T) 
\]  

(3.21)

or

\[
f(S_T) = D(T)^{-1} \frac{\partial^2 C}{\partial K^2} \bigg|_{K=S_T} 
\]  

(3.22)

So the probability distribution for the value of the underlying at time \( T \) conditional on the current level \( S_t \) is determined entirely by the prices of European call options expiring at \( T \). It is important to highlight that the distribution obtained using this procedure (and indeed all techniques discussed in this chapter) is the risk-neutral distribution for the underlying (i.e. the distribution corresponding to the equivalent martingale measure). This distribution is appropriate for pricing applications (see e.g. Cox and Ross [16], Hull [39]), but should not be used to predict the future behaviour of the underlying.
Existence of a Unique Diffusion Process Determined by Option Prices

The Breeden-Litzenberger result, proved in the previous section shows that there is a unique probability distribution

$$f(S, T; S_0)$$  \hspace{1cm} (3.23)

that is consistent with a complete specification of the volatility smile (equivalently, the prices of call options at every possible strike price, at a fixed expiration date $T$). In this section, we present a generalization of the above result to single factor diffusion processes. The result is originally due to Dupire [28], although our derivation more closely resembles that of Derman and Kani [21] [22]. From the theory presented in the first chapter, we know that arbitrage pricing requires that the discounted stock price be a martingale. This uniquely specifies the drift of the diffusion:

$$dS = r(t)Sdt + \sigma(S, t)SdW$$  \hspace{1cm} (3.24)

In this section, we prove that an appropriate choice of the diffusion coefficient will yield the correct prices for a surface of call option prices (i.e. prices for each strike price and each time to expiration).

The derivation in the previous section gave the following two results

$$\frac{\partial C}{\partial K} = -D(T) \int^\infty_K f(S, T; S_0) dS$$  \hspace{1cm} (3.25)

$$\left. \frac{\partial^2 C}{\partial K^2} \right|_{K=S} = D(T) f(S, T; S_0)$$  \hspace{1cm} (3.26)

where $D(T) = \exp \left( - \int^T_t r(\tau) d\tau \right)$.

The relationship between the distributions $f(S, T; S_0)$, and the process (3.24) can be determined using the infinitesimal generator $A$ of (3.24)

$$Af = rS \frac{\partial f}{\partial S} + \frac{1}{2} \sigma^2(S, T)S^2 \frac{\partial^2 f}{\partial S^2}$$  \hspace{1cm} (3.27)
The Fokker-Planck (forward Kolmogorov) equation provides the relationship between the generator $A$ and the distributions $f$, and is given by

$$\frac{\partial f}{\partial T} = A^* f$$

(3.28)

where $A^*$ is the adjoint of the generator $A$

$$A^* f = \frac{\partial^2}{\partial S^2} \left( \frac{1}{2} \sigma^2(S, T) S^2 f \right) - \frac{\partial}{\partial S} (r S f)$$

(3.29)

In order to obtain a closed form solution for the volatility coefficient $\sigma(S, t)$, we require that the call option pricing function $C(K, T; S_0)$ be differentiable with respect to $T$. We also need the following restrictions on the boundary (at $\infty$) behaviour of the distributions $f$

$$\lim_{S \to \infty} \left( (S - K) \left( \frac{\partial}{\partial S} \left( \frac{1}{2} \sigma^2(S, T) S^2 f(S, T; S_0) \right) - r S f(S, T; S_0) \right) \right) = 0$$

(3.30)

$$\lim_{S \to \infty} \sigma^2(S, T) S^2 f(S, T; S_0) = 0$$

(3.31)

**Theorem 3.6.1.** There is a unique diffusion process

$$dS = r(t) S dt + \sigma(S, t) S dW$$

(3.32)

with call option pricing function $C(K, T, S_0)$.

**Proof** As in the derivation of the Breeden-Litzenberger result, we begin with the definition of the call option pricing as an expectation with respect to the equivalent martingale measure

$$C(K, T) = D(T) \int_K^\infty (S - K) f(S, T) dS$$

(3.33)

(in the call pricing function $C$, and the distributions $f$, we suppress some dependencies for ease of notation). Differentiating the equation with respect to $T$ (and recalling the definition of $D(T)$) yields

$$\frac{\partial C}{\partial T} = -r(T) D(T) C(K, T) + D(T) \int_K^\infty (S - K) \frac{\partial}{\partial T} f(S, T) dS$$

(3.34)
\[ \frac{\partial C}{\partial T} = -r(T)D(T)C(K, T) \]
\[ + D(T) \int_{K}^{\infty} (S - K) \left( \frac{\partial^2}{\partial S^2} \left( \frac{1}{2} \sigma^2(S,T)S^2f(S,T) \right) - \frac{\partial}{\partial S}(rSf(S,T)) \right) dS \]

(3.35)

by the Fokker-Planck equation (3.28). Integrating by parts
\[ \frac{\partial C}{\partial T} = -r(T)C(K, T) \]
\[ + D(T) \left[ (S - K) \left( \frac{\partial}{\partial S} \left( \frac{1}{2} \sigma^2(S,T)S^2f(S,T) \right) - rSf(S,T) \right) \right]_K^{\infty} \]
\[ - D(T) \int_{K}^{\infty} \frac{\partial}{\partial S} \left( \frac{1}{2} \sigma^2(S,T)f(S,T) \right) - rSf(S,T) dS \]

(3.36)

So that the boundary assumption (3.30) gives
\[ \frac{\partial C}{\partial T} = -r(T)D(T)C(K, T) \]
\[ - D(T) \int_{K}^{\infty} \frac{\partial}{\partial S} \left( \frac{1}{2} \sigma^2(S,T)S^2f(S,T) \right) - rSf(S,T) dS \]

(3.37)

and another integration, together with boundary assumption (3.31)
\[ \frac{\partial C}{\partial T} = -r(T)D(T)C(K, T) \]
\[ + \frac{1}{2} D(T)\sigma^2(S,T)K^2 f(K,T) + D(T)r(T) \int_{K}^{\infty} S f(S,T) dS \]

(3.38)

Now, by the definition of the call option pricing function we have
\[ \frac{\partial C}{\partial T} = r(T)D(T)K \int_{K}^{\infty} f(S,T) dS + \frac{1}{2} \sigma^2(S,T)K^2 f(K,T) \]

(3.39)

so that the properties (3.25) (3.26) give
\[ \frac{\partial C}{\partial T} = -r(T)K \frac{\partial C}{\partial K} + \frac{\sigma^2(S,T)K^2}{2} \frac{\partial^2 C}{\partial K^2} \]

(3.40)

and we can solve for the diffusion coefficient
\[ \sigma^2(S,T) = 2 \cdot \frac{\frac{\partial C}{\partial T} + r(T)\frac{\partial C}{\partial K}}{\frac{\partial^2 C}{\partial K^2}} \]

(3.41)
where all derivatives are understood to be evaluated at $K = S$ Q.E.D.

Of course, in practice option prices are only observed at discrete intervals in time and only at a finite number of strike prices. Hence, interpolation or curve fitting techniques are required to determine a diffusion. Once the diffusion process is determined, there are a number of methods for constructing binomial approximations which are consistent in the sense that they converge to the process in the continuous time limit. We consider a few such methods below, and describe how they have been specifically applied to the construction of implied binomial trees.

**The Local Volatility Surface**

We now have two distinct volatility surfaces. The first is the traditional Black-Scholes implied volatility surface, and the second is the surface of instantaneous volatilities $\sigma(S, t)$ consistent with the market prices of liquid options, referred to as the *local volatility surface*. We have just demonstrated that it is possible to determine the local volatility surface uniquely from a knowledge of the entire implied surface. Similarly, given $\sigma(S, t)$, one could price all European options with the diffusion process (3.24), and then determine the Black-Scholes implied volatility surface from these prices. Knowledge of the two surfaces is thus theoretically equivalent. There are many advanced statistical techniques used to predict market volatility (e.g. the GARCH model and all of its variants). It is important to point out that the models presented in this survey in no way attempt to predict the future behaviour of market volatility. Implied probability distributions are the risk-neutral (martingale) distributions consistent with the observed prices of European options. The main application of the implied distributions, implied processes, and the discrete trees that approximate them is to provide methods for pricing exotic derivatives consistently with the observed implied volatility surface. It has also been suggested that implied distributions could be used to calculate better hedge ratios (see Shimko [75], Derman and Kani [21], Dupire [30]).
3.7 Interpolation Techniques

3.7.1 Shimko's Method

Based on the Breeden-Litzenberger result, Shimko [75] [76] developed a method for determining the risk-neutral distribution implied by option prices. In light of equation (3.22) any method of interpolation, or curve fitting, should provide an implied distribution for the future values of the underlying. Instead of interpolation option prices directly, Shimko determines a best quadratic fit to the volatility smile. Over the range of observed strike prices, a least squares parabola is fit to the volatility smile is fit. This gives implied volatility (and hence option price) as a function of strike price. The pricing relation is then twice numerically differentiated order to obtain the risk-neutral distribution. Quadratic fitting and numerical differentiation are only used inside the range of traded strike prices. Outside of this range, Shimko prescribes lognormal tails in a way such that the density and its cumulative distribution function are both continuous. Any invertible function of the strike price could be used instead of the implied volatilities. Shimko's method does not in any way assumes that the Black-Scholes model is valid. The Black-Scholes formula is only used in order to transform prices into implied volatilities, to fit the curve in the space of Black-Scholes implied volatilities, rather than option prices. This reflects the behaviour of market makers, who give quotes in terms of volatilities and then use computers in order to convert these volatilities into prices for market instruments. It is also important to observe that while curve fitting may produce smoother results with more aesthetic appeal, unlike interpolation it does not guarantee that the prices of all the original options will be accurately reproduced by the model.

3.7.2 Smiling Trees

We now turn our attention to constructing binomial and trinomial option pricing trees consistent with the market prices of liquid European options. The general problem of which diffusion processes can be approximated by computationally simple trees (i.e. those for which the number of nodes grows linearly with the time step) has been considered by Nelson and Ramaswamy [66]. In this section, we present the methods that have been
described explicitly for the purpose of constructing trees consistent with an interpolated Black-Scholes implied volatility surface (and the resulting call option pricing function). Throughout this section, we will employ the following notation:

- \((n, i)\) node of the tree at time step \(n\) and level \(i\).
- \(S_i^n\) Value of the variable \(S\) at time step \(n\) and level \(i\) in a given binomial or trinomial tree.

In a binomial tree a stock at \(S_i^n\) can go up to \(S_{i+1}^{n+1}\) with probability \(p^n_i\), and down to \(S_{i+1}^{n+1}\) with probability \(1 - p^n_i\). For a trinomial tree, a stock at \(S_i^n\) can make an up move to node \(S_{i+1}^{n+1}\) with probability \(p^n_i\), a down move to node \(S_{i+1}^{n+1}\) with probability \(q^n_i\), and a middle move to \(S_{i+1}^{n+1}\) with probability \(1 - p^n_i - q^n_i\).

### Arrow-Debreu Prices

An Arrow-Debreu security pays off 1 if a given event occurs in the future, and 0 otherwise. While these securities are not traded on the market, they have important theoretical interpretations in terms of risk-neutral probabilities. Specifically, the price of the Arrow-Debreu security for a given scenario is the discounted risk-neutral probability that this scenario will occur (e.g. in the discrete time setting described in the first chapter the elements of the price deflator were the prices of the corresponding Arrow-Debreu securities). For simplicity of notation, we assume that the risk-free interest rate is constant over the period \([t, T]\). Let \(\psi(S)\) denote the payoff at time \(T\) of the Arrow-Debreu security corresponding to the event that \(S_T = S^*\). Thus:

\[
\psi(S_T) = \begin{cases} 
1 & \text{if } S_T = S^* \\
0 & \text{otherwise.}
\end{cases}
\] (3.42)

Denote the price of the Arrow-Debreu security with payout \(\psi\) by \(g\). Using risk-neutral valuation to price the Arrow-Debreu security, we obtain:

\[
g = e^{-r(T-t)}E[\psi(S_T)]
\] (3.43)

Given the probabilistic interpretation of the prices of Arrow-Debreu securities, it is not surprising that they are very convenient for describing the
construction of stock pricing trees. The Arrow-Debreu price for node \((n, i)\) in a tree, denoted by \(\lambda^n_i\) is the price of the Arrow-Debreu security corresponding to \(S = S^n_i\) at time step \(n\). Trees are often easiest to construct using the equations for their Arrow-Debreu prices. Here we give the recursive formulas for \(\lambda^n_i\) for binomial trees:

\[
\lambda^{n+1}_i e^{r\Delta t} = \begin{cases} 
  p^n_{n-1} \lambda^n_{n-1} & i = n \\
  p^n_i \lambda^n_{i-1} + (1 - p^n_i) \lambda^n_i & 1 \leq i \leq n - 1 \\
  (1 - p^n_0) \lambda^n_0 & i = 0 
\end{cases} 
\]  

(3.44)

and for trinomial trees:

\[
\lambda^{n+1}_i e^{r\Delta t} = \begin{cases} 
  p^{2n+1}_{2n+1} & i = 2n + 3 \\
  (1 - p^{2n+1}_{2n+1} - q^{2n+1}_{2n+1}) \lambda^n_{2n+1} + p^{2n}_{2n} \lambda^n_{2n} & i = 2n + 2 \\
  p^n_i \lambda^n_{i-2} + (1 - p^n_{i-1} - q^n_{i-1}) \lambda^n_{i-1} + q^n_i \lambda^n_i & 1 \leq i \leq 2n + 1 \\
  (1 - p^n_0 - q^n_0) \lambda^n_0 + q^n_0 \lambda^n_1 & i = 1 \\
  q^n_0 \lambda^n_0 & i = 0 
\end{cases} 
\]  

(3.45)

**Binomial Trees**

Derman and Kani [21] show how to construct a binomial tree consistent with the implied volatility surface. We assume a constant time step \(\Delta t\). Suppose that we have built the tree up to the \(n\)th time step, and are interested in determining the values for the underlying at time \(n + 1\), and the transition probabilities from time \(n\) to time \(n + 1\). For simplicity of exposition, we assume that the underlying pays out no dividends. We have \(2n + 1\) unknown variables at time \(n + 1\). Specifically, the \(n + 1\) values for the stock:

\[
S^{n+1}_i = 0, \ldots, n 
\]  

(3.46)

and the \(n\) transition probabilities:

\[
p^n_i = 0, \ldots, n - 2 
\]  

(3.47)

(note that the probabilities are constrained to have sum 1). The absence of arbitrage for forward contracts yields the following conditions:

\[
F^n_i = e^{r\Delta t} S^n_i = p^n_i S^{n+1}_i + (1 - p^n_i) S^{n+1}_i 
\]  

(3.48)
where \( F^n_i \) the value of the forward contract expiring at time \( t_{n+1} \) given that the current level of the underlying is \( S^n_i \). This gives a total of \( n \) constraints.

Next, we use the values of call and put options determined by interpolating the implied volatility surface. The value of a European call option expiring at time \( t_{n+1} = (n + 1)\Delta t \), with strike price \( K \) is given by:

\[
C(K, t_{n+1}) = \sum_{j=0}^{n-1} \left[ \lambda^n_i p^n_i + \lambda^n_{i+1}(1 - p^n_{i+1}) \right] \max[S^n_{j+1} - K, 0] \tag{3.49}
\]

Substituting \( K = S^n_i \) into this formula, using the equation for forward prices \((i)\), and assuming \( S^n_i \leq S^n_i \leq S^n_{i+1} \) we obtain:

\[
e^r\Delta t C(S^n_i, t_{n+1}) = \lambda^n_i (S^n_{i+1} - S^n_i) + \sum_{j=0}^{n-1} \lambda^n_j (F^n_j - S^n_i) \tag{3.50}
\]

Rearranging this equation, and once again using the forward relation \((3.48)\) gives:

\[
S^n_{i+1} = \frac{S^n_i [e^{r\Delta t} C(S^n_i, t_{n+1}) - \Sigma] - \lambda^n_i S^n_i (F^n_i - S^n_{i+1})}{[e^{r\Delta t} C(S^n_i, t_{n+1}) - \Sigma] - \lambda^n_i (F^n_i - S^n_{i+1})} \tag{3.51}
\]

\[
F^n_i = \frac{S^n_i - S^n_{i+1}}{S^n_{i+1} - S^n_i} \tag{3.52}
\]

where \( \Sigma = \sum_{j=i+1}^{n-1} \lambda^n_j (F^n_j - S^n_i) \). If we take \( S^n_{i+1} \) to be known, then the above equation gives \( S^n_{i+1} \) in terms of known quantities (recall that the interpolated volatility surface, and hence the prices of calls with all strike prices and maturities is are known). A similar analysis for the case of the put option yields:

\[
S^n_i = \frac{S^n_{i+1} [e^{r\Delta t} P(S^n_i, t_{n+1}) - \Sigma] + \lambda^n_i S^n_i (F^n_i - S^n_{i+1})}{(e^{r\Delta t} P(S^n_i, t_{n+1}) - \Sigma) + \lambda^n_i (F^n_i - S^n_{i+1})} \tag{3.53}
\]

where now \( \Sigma = \sum_{j=0}^{i-1} \lambda^n_j (F^n_j - F^n_i) \). Thus, if we know the value of the underlying at any node at time \( t_{n+1} \) the values at all other nodes are determined, as well as all transition probabilities. If \( n + 1 \) is even (so that there are \( n + 2 \) nodes) then take:

\[
S^n_{n+1} = S \tag{3.54}
\]
where $S$ is the current value of the underlying. Otherwise use the additional constraint:

$$S_{n+\frac{1}{2}}^{n+1} = \frac{S^2}{S_{n+\frac{1}{2}+1}}$$  \hspace{1cm} (3.55)

to uniquely define the tree at time $t_{n+1}$.

The algorithm for constructing binomial trees as presented above is due to Derman and Kani [21]. Barle and Cakici [5] have introduced a number of improvements. From the equation:

$$p^n_i = \frac{F^n_i - S_i^{n+1}}{S_i^{n+1} - S_i^n}$$  \hspace{1cm} (3.56)

it is obvious that in order to avoid negative probabilities in the tree we must have:

$$F_i^n \leq S_i^{n+1} \leq F_{i+1}^n$$  \hspace{1cm} (3.57)

Violations of the above inequality result in arbitrage opportunities due to the negative probabilities. In order to maintain this condition in their tree, Derman and Kani occasionally have to override the calculated values for underlying at a given level. Barle and Cakici replace the assumption $S_i^{n+1} \leq S_i^n \leq S_{i+1}^{n+1}$ with the inequality (3.57), hence removing many of the "overrides" necessary in the original algorithm for the tree. In order to achieve this construction, instead of evaluating the call at strike price $S_i^n$ in (3.49), it is evaluated at $F^n_i$, with a similar condition for the put in (3.53).

As a second improvement, Barle and Cakici allow the centre of the tree to grow at the forward rate, rather than remain constant as in the Derman-Kani algorithm, so equations (3.54) and (3.55) become:

$$S_{n+\frac{1}{2}}^{n+1} = F^n_i$$  \hspace{1cm} (3.58)

and

$$S_i^{n+1}S_{i+1}^{n+1} = (F^n_i)^2$$  \hspace{1cm} (3.59)

Barle and Cakici derive equations analogous to (3.51) and (3.52) for the values of the underlying in their tree. They point out that their recursive
procedure can lead to negative probabilities in a similar way to the one used by Derman and Kani, when the inequality (3.57) is violated. Similar to Derman and Kani they override such values, with any value of $S_{i+1}^n$ which obeys the inequality (3.57). (In their original paper, Derman and Kani enforced logarithmic spacing on such exceptional nodes, which can still lead to negative probabilities due to violations of inequality (3.57)). Barle and Cakici point out that their approach is essentially equivalent to building the tree for futures prices, rather than the underlying.

**Generalizations to American Options**

Until now, the methods we have presented have dealt with inferring probabilistic information from the prices of European options on the underlying. The early exercise feature of American options makes their evaluation much more difficult. However, most options that are traded on exchanges are in fact American, and it would therefore be useful to have a method for determining a binomial tree consistent with the prices of American options as well as Europeans. Chriss [12] has generalized the above techniques to allow the construction of a tree consistent with the prices of American options. The method involves using an iterative nonlinear root finder (each iteration calling for the pricing of an American option on a binomial tree). A nonlinear equation must be solved using this root finder at every node where the corresponding input option is American, with a possibility of early exercise. This significantly increases the computational burden involved in constructing the tree.

**Trinomial Trees**

The binomial trees discussed in the previous section have only one free parameter, which is typically used to specify the growth of the central node in the tree through time. Trinomial (in general, multinomial) trees provide a large number of free parameters that can be manipulated by the user. For example, Derman, Kani and Chriss [23] find that the trinomial trees they construct to be consistent with the implied volatility surface produce smoother implied densities for the underlying than the corresponding binomial trees.
The standard methodology is to employ the additional degrees of freedom afforded by a trinomial lattice to make an exogenous specification of the relevant state space (i.e. prescribe in advance all the time steps, and the nodal values for the underlying at each time). Once the state space has been specified, the tree is uniquely determined by the transition probabilities at every level and time. These can be inferred using the volatility smile by evaluating options in a manner similar to the method for binomial trees, discussed above.

Suppose that the values for the underlying at each node in the tree is fixed, and the transitions probabilities (hence also Arrow-Debreu prices) are known up to level \( n \). We seek to infer the probabilities for transition from level \( n \) to \( n + 1 \) implied by options prices. At time \( n \) the underlying has possible values \( S_i^n \) for \( i = 0, \ldots, 2n \). If the underlying assumes the value \( S_i^n \) at time step \( t_n \), then it will assume the values \( S_{i+2}^{n+1}, S_{i+1}^{n+1}, S_i^{n+1} \) at time \( n + 1 \) with probabilities \( p_i^n, 1 - p_i^n - q_i^n, q_i^n \) respectively. \( p_i^n \) denotes the probability of an “up” move, \( q_i^n \) the probability of a “down” move, with the remaining probability mass absorbed into a “middle” move. A calculation parallel to the one performed in the binomial case yields:

\[
p_i^n = \frac{e^{\Delta t C(S_i^{n+1}, t_{n+1})} - \sum_{j=i+1}^{2n} \lambda_j^n (F_j^n - S_i^{n+1})}{\lambda_i^n (S_{i+2}^{n+1} - S_i^{n+1})}
\]

(3.60)

and

\[
q_i^n = \frac{F_i^n - p_i^n (S_{i+2}^{n+1} - S_i^{n+1}) - S_i^{n+1}}{S_i^{n+1} - S_{i+1}^{n+1}}
\]

(3.61)

where \( F_i^n = p_i^n S_{i+2}^{n+1} + (1 - p_i^n - q_i^n) S_{i+1}^{n+1} + q_i^n S_i^{n+1} \). Again, analogous equations for put options can be obtained by methods identical to those employed in the binomial case.

If the smile is not too pronounced, Dupire [29] suggests that it may be adequate to simply prescribe a constant logarithmic spacing in the underlying, with some care being necessary to that the spacing in the tree is wide enough for the above calculations to be feasible. Derman, Kani and Chriss [23] consider a number of alternative prescriptions for the state space. For example,
spacing could be set in order to have \( p_i^n = q_i^n = \frac{1}{3} \) at each time and level in the tree. Another approach is to make the nodes at time \( t_{n+1} \) in the trinomial tree have the same value as the nodes at the \( 2^n \)th time step in a standard binomial tree. Derman, Kani and Chriss [23] also consider a more sophisticated approach, similar to the one introduced by Nelson and Ramaswamy [66], whereby the time and price variables are rescaled in such a way that implied volatilities remain constant throughout the “transformed” tree. (For more details see Nelson and Ramaswamy [66].)

3.8 Optimization Methods

Our discussion above focussed on interpolating the implied volatility surface in order to construct an implied tree from option prices. There is little justification for why any interpolation technique should produce the appropriate implied volatility surface (equivalently, the appropriate European option prices). A fundamental observation regarding the Breeden-Litzenberger result motivates two other paradigms: optimization and stability. Avellaneda et al. observe that methods based on interpolation “tend to be unstable since the solution is very sensitive to the smoothness and convexity of the function used in the interpolation.” (Under extremely general assumptions, call options prices must be a convex function of the underlying, see Merton [62]).

Let us examine this situation more carefully. Recall the Breeden-Litzenberger result from the previous section:

\[
\frac{\partial^2 C}{\partial K^2}\bigg|_{K=S} \quad (3.62)
\]

For a fixed \( T \) with \( K \) varying solving this equation for \( f \) is known to be ill-posed, and can lead to instability in techniques for numerical solution (see below for an approach that considers the stability of the problem directly).

In this section, we consider an alternative to interpolation: optimization of a measure of the “fitness” of the implied distribution. If we restrict our attention to distributions that price all liquid options in agreement with the
observed volatility smile, mathematically we obtain a constrained optimization problem.

From the previous section, we know that if the complete implied volatility surface is given (equivalently, the prices of options with all expiration dates and strike prices), there is a unique arbitrage free diffusion process that correctly prices all options. In practice, we only observe prices for options with a finite number of strike prices and expiration dates. Even for the prices that are quoted, the market maker's bid-ask spread must be taken into account. Difficulties with the data are further compounded by the fact deep in-the-money and deep out-of-the-money options are highly illiquid, and their prices may be subject to a time lag (difference between the time the price is first quoted and when it is observed).

There are infinitely many probability distributions for the underlying that price all observed options within their respective bid-ask spreads. For pricing applications, we only wish to consider one of these "feasible" distributions. Interpolation of the implied volatility surface provides a method for selecting a distribution, but there is little intuitive evidence that this distribution should in any way be "best". Rubinstein [72] introduces an optimization perspective, where the "fitness" of the distribution is maximized among all possible distributions which correctly reproduce the observed prices of liquid options. The mathematical problem, in its most general form, reduces to:

$$\max_{\varphi} \Theta(\varphi)$$

(3.63)

$$\varphi \in \Omega$$

(3.64)

where $\Theta$ is a measure of the "fitness" of a probability distribution, and $\Omega$ is the set of all feasible distributions (i.e. distributions that price all observed options within their respective bid-ask spreads).

### 3.8.1 Discrepancy from a Prior Distribution

One strategy is to minimize the distance from an assumed prior distribution, $q$. For example, $q$ could be the lognormal assumed by the Black-Scholes
model, with volatility equal to the implied volatility of an at the money option. In this case, we have \( \Theta(\varphi) = -\rho(\varphi, q) \), where \( \rho \) is a measure of the distance between the distributions \( p \) and \( q \). The resulting optimization problem:

\[
\min_{\varphi} \rho(\varphi, q)
\]

(3.65)

\[
\varphi \in \Omega
\]

(3.66)

represents the goal of remaining as close as possible to the prior distribution \( q \), while still pricing consistently with the market.

Rubinstein [72] illustrates the above approach in a situation where all the options considered have the same expiration date, and the state-space is discrete. We seek a discrete approximation to the distribution of the underlying at the maturity date. Uncertainty regarding the level of the underlying at the expiration date is represented by a finite number, \( n + 1 \) of scenarios, where the underlying assumes the values \( S_j, j = 0, \ldots, n \). Let the prior probability of the underlying assuming value \( S_j \) at maturity be \( P_j \). For example, we could use the nodes \( S_j \) and probabilities \( P_j \) generated by an \( n \) step Cox-Ross-Rubinstein binomial tree, constructed using the at the money Black-Scholes implied volatility. Consider \( m \) European options with strike prices \( K_i, i = 1, \ldots, m \). Denote the bid and ask prices for these options as \( C^b_i \) and \( C^a_i \) respectively. Let the current bid and ask price for the underlying be \( S^b \) and \( S^a \). For simplicity, we assume that the underlying pays dividends at a constant one period rate \( \delta \), and that the one period risk free interest rate \( r \) is constant throughout the lifetime of the options.
Using a least squares criterion to measure the distance between probability distributions gives the problem in a quadratic programming form:

\[
\min_{P_j} \sum_{j=0}^{n} (P_j - P'_j)^2
\]  

subject to

\[
\sum_{j=0}^{n} P_j = 1 \quad P_j \geq 0 \quad j = 0, \ldots, n
\]  

\[
S^b \leq \left( \frac{\delta^n \sum_{j=0}^{n} P_j S_j}{r^n} \right) \leq S^a
\]  

\[
C_i^b \leq \left( \frac{\sum_{j=0}^{n} \max[0, S_j - K_i]}{r^n} \right) \leq C_i^a \quad i = 1, \ldots, m
\]

Other measures of discrepancy are of course possible. For instance, the sum of absolute differences:

\[
\sum_{j=0}^{n} |P_j - P'_j|
\]

allows the above quadratic programming problem to assume a linear programming form. Other suggestions include the goodness of fit measure:

\[
\sum_{j=0}^{n} \left( \frac{(P_j - P'_j)^2}{P_j} \right)
\]

and the entropy distance

\[
\sum_{j=0}^{n} -P_j \ln \left( \frac{P_j}{P'_j} \right)
\]

which we shall examine in greater detail in the continuous time setting below.
Jackwerth and Rubinstein [45] give empirical results for each of the above measures of discrepancy. They find that the results are relatively insensitive to both the choice of distance measure $\rho$ and the prior distribution $q$. In some circumstances, each of the discrepancies leads to an implied probability density function that is too ‘jagged’ for intuition. This leads to the proposal of an alternative criterion based on maximizing the “smoothness” of the implied density.

The objective function in (3.67) is replaced with:

$$\min_{P_i} \sum_{j=0}^{n} (P_{j-1} - 2P_j + P_{j+1})^2$$

(3.74)

which has an interpretation in terms of the finite difference approximation of the second derivative of the probability density with respect to the level of the underlying. If we assume that the $S_j$'s are evenly spaced, $S_{j+1} - S_j = h$, for $j = 1, \ldots, n$, then:

$$\frac{\partial^2 P}{\partial S^2} \approx \frac{P_{j+1} - P_j}{h^2} - \frac{P_j - P_{j-1}}{h^2} = \kappa_j (P_{j-1} - 2P_j + P_{j+1})$$

(3.75)

In their implementation, Rubinstein and Jackwerth ignore the weights $\kappa_i$, and square each term in the objective function in order to obtain a measure of the magnitude of the second derivative of the implied density function. The corresponding minimization problem for continuous distributions, where the objective function is the $L^1$ norm of the second derivative of $P$ with respect to $S$, yields a solution in terms of cubic spline (see Mayhew [60]). Jackwerth and Rubinstein solve the problem with a penalty function method, minimizing a function that consists of the objective function plus a weighted sum of the violations on each of the constraints defining the feasible set. This approach achieves greater “smoothness” in return for sacrificing the strict feasibility of the implied density (solutions to the problem typically contain small regions with negative probability, and will occasionally price options outside of their bid-ask spreads).

It remains only to build a tree consistent with the implied probability distribution. Rubinstein’s original tree is constructed in terms of the return on
the stock, and path probabilities through the tree. Jackwerth [43] showed that this construction is equivalent to:

\[ P_{n-1,i} = w(n,i)P_{n,i} + (1 - w(i,j))P_{i,j+1} \quad (3.76) \]

\[ S_{i-1,j} = \frac{w(i,j)P_{i,j}S_{i,j} + (1 - w(i,j + 1))P_{i,j+1}S_{i,j+1}}{(\xi)(w(i,j)P_{i,j} + (1 - w(i,j + 1))P_{i,j+1})} \quad (3.77) \]

where \( w(i,j) = \omega(\text{frac}i) = 1 - \left(\frac{i}{j}\right) \), and the final level of the tree is set by an optimization procedure as outlined above. Here \( P_{n,i} \) is the probability of reaching the node \((n,i)\), and is not the same as the transition probability \( P_i^n \) discussed in previous sections. By considering arbitrary “weight” functions \( w(n,i) = \omega\left(\frac{i}{n}\right) \), Jackwerth is able to generalize the above approach to options with more than one expiration date.

### 3.8.2 The Maximum Entropy Criterion

#### Maximum Entropy Distributions from Options Prices

An alternative technique, mentioned briefly above, involves specifying the fitness function to be its entropy measure. The entropy, \( Y(p) \) of a given probability distribution \( p \) is given by:

\[ Y(p) = -\int_{-\infty}^{\infty} \ln (p(x))p(x)dx \quad (3.78) \]

A related notion is the entropy distance between two distributions \( p \) and \( q \) given by:

\[ \rho(p, q) = \int_{-\infty}^{\infty} \ln \left( \frac{p(x)}{q(x)} \right) p(x) \, dx \quad (3.79) \]

The concept of entropy originated in thermodynamics, where it is used as a measure of the information in a given system. The maximum entropy distribution yields an interpretation as the one that is minimally inferred from a given dataset, in a sense making the fewest ad hoc assumptions regarding the nature of the distribution. For example, the maximum entropy probability distribution defined on the interval \([a, b]\) is simply the uniform distribution on \([a, b]\). In light of the focus of attention on alternatives to log-normality, it is interesting to note that the maximum entropy distribution on
[0, \infty] given only the mean and variance of the distribution is the lognormal. Rubinstein [72], and Jackwerth and Runistein [45] consider the entropy distance as a possible discrepancy measure in the optimization problem (3.65). In the discrete case, this amounts to using the objective function:

$$
\sum_{j=0}^{n} P_j \ln \left( \frac{P_j}{P'_j} \right)
$$

in the nonlinear program (3.67).

Similar to the smoothness criterion, maximum entropy can be used in a nonparametric setting, where no assumptions are made on the existence of a prior distribution. Buchen and Kelly [11] consider the problem in a continuous state setting. Suppose we observe the prices of n call options with strike prices $K_i$, all maturing at time $T$. Let risk-neutral discount for assets at time $T$ to the present be denoted by $D(T)$. Further, denote $c_i(S_T) = D(T)(S_T - K_i)_+$. Then using the maximum entropy criterion, the optimization problem (3.65) assumes the form (where we assume that the underlying $S_T$ cannot assume negative values):

$$
\max \int_0^\infty -p(S_T) \ln (p(S_T)) dS_T
$$

subject to

$$
p(S_T) \geq 0 \quad S_T \in [0, \infty]
$$

$$
\int_0^\infty p(S_T) dS_T = 1
$$

$$
E[c_i(S_T)] = \int_0^\infty p(S_T)c_i(S_T) dS_T = d_i \quad i = 1, \ldots, n
$$

The Lagrangian of the above problem is:

$$
L(p) = -\int_0^\infty \ln (p(S_T)) dS_T + (1 + \lambda_0) \int_0^\infty p(S_T) dS_T + \sum_{i=0}^{n} \lambda_i \int_0^\infty c_i(S_T)p(S_T) dS_T
$$
To solve the optimization problem, we need only determine the $p$ for which:

$$
\int_0^\infty \left( -\ln (p(S_T)) + \lambda_0 + \sum_{i=1}^n \lambda_i c_i(S_T) \right) \delta p(S_T) \, dS_T = 0 \quad (3.86)
$$

This leads to the solution:

$$
p(S_T) = \frac{1}{\alpha} \exp \left( \sum_{i=1}^n \lambda_i c_i(S_T) \right) \quad (3.87)
$$

where

$$
\alpha = \int_0^\infty \exp \left( \sum_{i=1}^n \lambda_i c_i(S_T) \right) \, dS_T \quad (3.88)
$$

A similar result obtains when a prior distribution $q$ is assumed for the underlying and the entropy distance $\rho(p, q)$ is minimized. In this case, the implied probability distribution becomes:

$$
p(S_T) = \frac{q(S_T)}{\alpha} \exp \left( \sum_{i=1}^n \lambda_i c_i(S_T) \right) \quad (3.89)
$$

where now:

$$
\alpha = \int_0^\infty q(S_T) \exp \left( \sum_{i=1}^n \lambda_i c_i(S_T) \right) \, dS_T \quad (3.90)
$$

**Extensions to a Local Volatility Surface**

Avellaneda et al. [3] extend the above discussion to the calibration of an entire volatility surface. Entropy distance from a prior distribution is minimized, so that optimal calibration becomes a *constrained stochastic control* problem:

$$
\min E^Q \left[ \int_0^T \eta(s^2(s)) \, ds \right] \quad (3.91)
$$

subject to

$$
E^Q [e^{-T_i r} G_i(S_{T_i})] = d_i \quad i = 1, \ldots, n \quad (3.92)
$$
where the minimization occurs over all probability distributions $Q$ of diffusions:

$$dS = \mu S dt + \sigma_t S dW$$

(3.93)

such that $\sigma_t$ is a progressively measurable process satisfying $0 \leq \sigma_{\text{min}} \leq \sigma_t \leq \sigma_{\text{max}} < \infty$, and the $G_i$ are the payoff functions of a set of derivative securities whose prices are observed in the market. Here $\eta(\sigma^2)$ can be any pseudo-entropy, i.e. any smooth real-valued positive, strictly convex function defined on $[0, \infty)$ with minimum 0 at $\sigma^2 = \sigma_0^2$, where $\sigma_0^2$ is the assumed prior volatility.

The problem is solved by applying techniques from stochastic control theory. Consider the Legendre dual of $\eta$, $\Phi(X)$ given by

$$\Phi(X) = \min_{\sigma_{\text{min}}^2 < \sigma^2 < \sigma_{\text{max}}^2} [\sigma^2 X - \eta(\sigma^2)]$$

(3.94)

and the Hamilton-Jacobi-Bellman equation:

$$\frac{\partial W}{\partial t} + e^{rt} \Phi \left( \frac{e^{-rt}}{2} S^2 \frac{\partial^2 W}{\partial S^2} \right) + \mu S \frac{\partial W}{\partial S} =$$

$$- \sum_{t < T_i \leq T} \lambda_i \delta(t - T_i) G_i(S), \quad S > 0, t \leq T$$

(3.95)

(3.96)

with final condition $W(S, T + 0) = 0$. The optimal solution of the stochastic control problem can be obtained by finding a global minimum for:

$$V(S, t; \lambda_1, \lambda_2, \ldots, \lambda_M) = W(S, t; \lambda_1, \lambda_2, \ldots, \lambda_M) - \sum_{i=0}^M \lambda_i C_i$$

(3.97)

Denote the parameter values that optimize the above equation by $\lambda_1^*, \ldots, \lambda_M^*$, and the solution to the corresponding equation $HJB$ by $W^*$. Then the volatility of the optimal diffusion process is

$$\sigma^*(S, t) = \sqrt{\Phi'(\frac{e^{-rt}}{2} S^2 \frac{\partial^2 W^*}{\partial S^2})}$$

(3.98)

The above procedure provides a maximum entropy estimate for the local volatility surface. In order to solve the problem numerically, one begins
with an initial estimate of the parameter vector \( \Lambda^1 = (\lambda^1_1, \ldots, \lambda^1_M) \). Next, the corresponding HJB equation must be solved numerically for \( W^1 \). An iterative routine for numerical optimization of a function in several variables (e.g. the BFGS or DFP variable metric methods) is then employed to yield successive improvements \( \Lambda^k, W^k \) until a desired tolerance is reached. The method is computationally expensive, requiring the numerical solution of the partial differential equation \( HJB \) at each iteration. Avellaneda et al. implement their method with a quadratic pseudo-entropy:

\[
\eta(\sigma^2) = \frac{1}{2}(\sigma^2 - \sigma_0^2)^2
\]

and a constant prior \( \sigma_p(S, t) = \sigma_0 \). They test the method on Dollar/Deutschmark currency options data. The shape of the smile in currency options markets is highly erratic, and does not at all exhibit the consistent downward slope that has been observed in index options markets.

**Entropy Maximization with a Historical Prior**

Stutzer [78] [79] uses the maximum entropy criterion in an original manner. Here the exogenous specification of a prior distribution is foresaken in favour of maximizing proximity to the historical behaviour of the underlying. Consider a series of observations on the value of the underlying at past times \(-dt, -2dt, \ldots, -Ndt\), and suppose that we wish to estimate the distribution for the underlying at time \( Tdt \), with \( T < N \). First calculate the \( N - T \) historically observed returns:

\[
R^k = \frac{S(-kdt)}{S(-(k-T)dt)} \quad k = 1, \ldots, N - T
\]

where \( S(-kdt) \) is the price of the underlying \( k \) periods ago. Next, the distribution for the underlying at time \( Tdt \) is discretized into \( N - T \) scenarios corresponding to the values

\[
S_T^k = S_0 R^k \quad k = 1, \ldots, N - T
\]

where \( S_0 \) is the current level of the underlying. Each of the scenarios is given the prior probability \( p_k = \frac{1}{N-T} \). This equivalent to assuming a uniform prior
over the set of scenarios generated by a rolling window over the historical series of returns. Similar procedures can be employed for other variables relative to the prices of derivatives (e.g. the risk-free rate of interest or the dividend rate paid out by the underlying). The uniformity of the prior leads to maximizing the entropy:

$$\sum_{k=1}^{N-T} -p_k \ln(p_k)$$

over the generated scenarios. The optimization is constrained so that the optimal density produces the current value of the underlying as its discounted expected future payoff (the martingale restriction for the risk-neutral pricing distribution - see the first chapter for further details). The set of feasible densities can also be restricted to those that correctly price traded options simply by adding constraints on the expected payoffs of derivatives under the risk-neutral measure, as was demonstrated above. It is important to note that the results obtained by using the above approach can vary significantly depending on whether or not data from the 1987 stock market crash is included in the time series.

### 3.9 A Statistical Technique: Nonparametric Regression

Ait-Sahalia and Lo [1] consider a nonparametric technique from nonlinear regression, known as kernel smoothing, to estimate the implied probability distribution of the underlying. The approach shares with maximum entropy methods the advantage of not assuming any functional form for the implied distribution. Additionally, freedom is allowed over a “smoothness” or “bandwidth” parameter (see below) which can be manipulated to produce smoother implied distributions. The method can be further constrained to produce distributions which are guaranteed to have continuous partial derivatives up to a given order (usually taken to be 4). A drawback is that nonparametric techniques are extremely data-intensive, and make some assumptions regarding the process by which the data is generated.

We consider the call option pricing function $C$, and then employ the Breeden-Litzenbeger result in order to obtain the implied probability density func-
tion. The call option price is taken to be a function of five explanatory variables: the current level of the underlying $S_t$, the option’s strike price $K$, the time to maturity $\tau = T - t$, the risk free interest rate over the lifetime of the option $r_{t, \tau}$, and the dividend rate paid by the underlying over the lifetime of the option $\delta_{t, \tau}$:

$$C = C(Z) = C(S_t, K, \tau, r_{t, \tau}, \delta_{t, \tau})$$  \hspace{1cm} (3.103)

We also have observations of various call prices $C_i$, at points $Z_i$

$$Z_i = [S_{t_i}, K_i, \tau_i, r_{t_i, \tau_i}, \delta_{t_i, \tau_i}] \quad i = 1, \ldots, n$$  \hspace{1cm} (3.104)

The data $Z_i$ will include historical time series, as well as current market data, and hence it is not reasonable to expect the function $C$ to price all the calls correctly. Instead, a least squares criterion is employed

$$\min_{C \in \Omega} \sum_{i=1}^{n} [C_i - C(S_{t_i}, K_i, \tau_i, r_{t_i, \tau_i}, \delta_{t_i, \tau_i})]^2$$  \hspace{1cm} (3.105)

The set $\Omega$ restricts the domain of pricing functions. For example, a minimal requirement in order to employ the Breeden-Litzenberger result is that the call option price be twice continuously differentiable in the strike price variable. Ait-Sahalia and Lo consider only those functions that are four times continuously differentiable in each of the variables. The above sum is minimized by the conditional expectation of $C$ given $Z$ (see for example, Rao [68]). The mathematical problem therefore reduces to the determination of an appropriate estimate $\hat{C}$ of the conditional expectation $E[C|Z]$. 

*Kernel smoothing* estimates $C$ as a sum of the observed values dependent variables $C_i$:

$$\hat{C} = \hat{E}[C|Z] = \frac{\sum_{i=1}^{n} K((Z - Z_i)/h)C_i}{\sum_{i=1}^{n} K((Z - Z_i)/h)}$$  \hspace{1cm} (3.106)

where $K$ the called the *regression kernel* and $h > 0$ is the *bandwidth*. The above estimator is referred to as the *Nadaraya-Watson estimator* for the conditional expectation. Typically, the regression kernel is taken to be a probability density function, since these functions are positive and integrate to one. The parameter $h$ has an important interpretation as a smoothing factor. Small values of $h$ place greater emphasis on those observations $Z_i$
that are closer to $Z$, whereas larger values of $h$ produce smoother functions $C$ in terms of the independent variables. Proper selection of the value for $h$ is a delicate matter, involving a tradeoff between valuing proximity and smoothness.

Ait-Sahalia and Lo specify the kernel $K$ to be the product of one-dimensional kernels in each of the explanatory variables. Typical one-dimensional kernels are:

$$k^{(2)}(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$  \hspace{1cm} (3.107)

$$k^{(4)}(z) = \frac{3}{\sqrt{8\pi}} \left(1 - \frac{z^2}{3}\right) e^{-\frac{z^2}{2}}$$  \hspace{1cm} (3.108)

where the superscript indicates the order of the kernel, i.e. the largest even integer $s$ such that $\int u^l k(u) du = 0$ for $l = 1, \ldots, s-1$, and $\int |u|^s k(u) du < \infty$. To estimate the call option pricing function, consider:

$$k_{S_t} = k_K = k_r = k_r = k_\delta = k^{(4)}$$  \hspace{1cm} (3.109)

This results in the following estimate for the price of a European call option:

$$C(S_t, K, \tau, r_t, \tau, \delta_t, \tau) =$$  \hspace{1cm} (3.110)

$$\sum_{i=1}^n k_S(\frac{S_t-S_{t_i}}{h_S}) k_K(\frac{K-K_{t_i}}{h_K}) k_r(\frac{\tau-\tau_{i}}{h_\tau}) k_r(\frac{\tau_t-\tau_{i_t}}{h_{\tau_t}}) k_\delta(\frac{\delta_t-\delta_{i_t}}{h_\delta}) C_i$$

$$\sum_{i=1}^n k_S(\frac{S_t-S_{t_i}}{h_S}) k_K(\frac{K-K_{t_i}}{h_K}) k_r(\frac{\tau-\tau_{i}}{h_\tau}) k_r(\frac{\tau_t-\tau_{i_t}}{h_{\tau_t}}) k_\delta(\frac{\delta_t-\delta_{i_t}}{h_\delta})$$  \hspace{1cm} (3.111)

As mentioned previously, the choice of the bandwidth parameters $h$ is crucial to the success of the technique. Ait-Sahalia and Lo use:

$$h_j = \alpha_j \sigma_j (Z_j)^{\frac{1}{4+2p}}$$  \hspace{1cm} (3.112)

where $d$ is the number of explanatory variables, indexed by $j$ (here $d = 5$), $p$ is the number of continuous derivatives the function $C$ should have (here $p = 4$), and the constants $\alpha_j$ are chosen in order to minimize the mean-square error of the function $C$.  

99
3.10 Use of a Prior Distribution

In this section, we consider the legitimacy of choosing minimization of the distance from an assumed prior distribution as a technique for selecting the "optimal" measure from those that correctly price the European options traded on the market.

By far the most popular prior distribution is the lognormal distribution assumed in the Black-Scholes model. This is mainly due to the immense popularity and historical significance of the Black-Scholes model. Nonetheless, experimentation with other priors is required in order to determine the distribution that works best. (One criterion would be to compare the stability of the implied distributions arising from different priors across time. Another might examine how the distribution implied by a subset of observed options prices the remaining liquid options.) Of course, in the limit as more constraints are imposed the solution becomes less dependent on the specific form of the objective function, and more on the option prices in the market.

Jackwerth and Rubinstein [45] found that the assumption of a prior distribution led to intuitively unappealing distributions that are extremely "jagged" (this observation led to the adoption of the smoothness criterion discussed above). In general, methods that do not assume a prior do not use any information extraneous to the data. This has been used as a justification for some nonparametric techniques, as it is claimed that they introduce the least bias into the model (see e.g. Buchen and Kelly [11] or Ait-Sahalia and Lo [1]). It could also be seen as a disadvantage, since it ignores any additional information (outside the dataset) that the researcher may bring to the model. Market participants will often have intuitions regarding the future distributions of market instruments, and may only seek to "adjust" their intuition in a way to make it consistent with the market prices of liquid options.

Avellaneda et al. [3] give another argument in favour of the use of prior distributions, motivated by a computational view: "Minimizing ... with respect to the prior stabilizes the far-tails of the probability distribution for the underlying index." Rubinstein [72] argues that the problem of finding an
appropriate implied probability distribution is equivalent to interpolation within the observed strike prices and extrapolation on the tails (other treatments of the subject, such as that of Derman and Kani [21] make this more explicit). Recall that Shimko [75] used quadratic fitting within the observed range of implied volatilities, while exogenously specifying lognormal tails for the implied distribution. In a footnote, Avellaneda et al. state “a unique prescription of the volatility surface far away from traded strikes cannot be obtained precisely from option prices. The introduction of a Bayesian prior serves as an ‘extrapolation mechanism’ for characterizing the volatility in regions where the price information is weak ... as well as a mechanism for smoothing the volatility surface.” It is interesting to notice that when using minimization to estimate implied distributions Jackwerth and Rubinstein find that a prior distribution detracts from smoothness, whereas in the estimation of a volatility surface, Avellaneda et al. [3] assert that prior distributions enhance smoothness. Further testing, as well as a better determination of the desirable properties of implied distributions is required before any final conclusion can be made on the validity of assuming a prior distribution for the underlying asset.

3.11 Stability

The options prices in any given dataset are not exact. Measurement errors are unavoidable due the presence of such market factors as asynchronous price quotes and the bid-ask spread. While it may be reasonable to assume that prices are known within a given tolerance level, it is hardly valid to assert that the data are known exactly. In estimating the local volatility surface $\sigma(S,t)$ from observed prices, it is therefore desirable to select a method whereby small changes in the input data (option prices) lead to only small changes in the resulting output (the local volatility surface).

Bodhurtha and Jermakyan [47] [46] present an approach to volatility surface estimation that examines the problem of stability directly. Again, we consider the most general single factor diffusion process for the underlying:

$$dS = \mu(S,t)Sdt + \sigma(S,t)SdW$$  \hspace{1cm} (3.113)
An analysis identical to the Black-Scholes argument given in the first chapter yields the partial differential equation that any derivative security on this underlying must obey:

\[ rf = \frac{\partial f}{\partial t} + rS\frac{\partial F}{\partial S} + \frac{1}{2}\sigma^2(S,t)S^2\frac{\partial^2 f}{\partial S^2} \]  

(3.114)

If \( f \) is a European call option with strike price \( K \), then we also know the final condition \( f(S,T) = \max(S - K, 0) \) and the boundary condition \( f(0,t) = 0 \).

In theory, we know the process followed by the underlying security, and are interest in deducing option prices from that process. In reality, what we observe is the prices of liquid derivative securities. From these prices we wish to determine the diffusion coefficient \( \sigma(S,t) \) that generated the observed prices. As we have noted above, there are many different values for \( \sigma(S,t) \) that will suffice. Estimating the coefficient of a partial differential equation given noisy observations on the solution is an inverse problem that is known to be ill-posed (see Tikhonov and Arsenin [80]).

By a series of transformations, it is possible to reduce the partial differential equation for the call, together with its boundary conditions into the form:

\[ \frac{\partial U(v, Z)}{\partial v} = \frac{\bar{T} \alpha^2(v, S)}{2} \left( \frac{\partial^2 U(v, Z)}{\partial Z^2} - \frac{\partial U(v, Z)}{\partial Z} \right) \]  

(3.115)

with boundary condition:

\[ U(0, Z) = \max(0, e^Z - 1) \]  

(3.116)

There are two main difficulties with the above equation from a computational viewpoint. The first is the instability of the inverse problem, as discussed above. The second is that the diffusion coefficient \( \alpha(S,t) \) is not constant, varying with both time and the level of the underlying.

The algorithm developed by Bodurtha and Jermakyan proceeds in two stages. First, the above equation is converted into a sequence constant coefficient equations using the method of small parameter. Next, each member of this sequence of problems is stabilized using a Tikhonov regularization. The mathematics involved are not overly difficult, but the derivations are
rather long, and hence we refer readers to the original paper for details. Naturally, the entire (infinite) sequence of problems cannot be solved, so it must be truncated after a finite number of steps. Each iteration in the procedure requires the numerical solution of a partial differential equation, and an optimization problem (in order to determine the regularization parameter in the Tikhonov regularization). The desired stability is obtained, but at a significant computational expense.

3.12 Empirical Tests

Despite the large amount of recent research on implied probability distributions, there has been little work comparing the performance of the various models on market data. As previously mentioned, Jackwerth and Rubinstein [45] analyze the performance of their optimization framework with different objective functions. Their results indicate relatively uniform performance across objectives, the optimization being much more sensitive to the nature of the constraints.

Dumas et al. [27] compare the prediction properties of a parametrized family of volatility surfaces. First the volatility surface is estimated in one of the following forms:

\[
\sigma(S, t) = \alpha_0 \tag{3.117}
\]

\[
\sigma(S, t) = \alpha_0 + \alpha_1 S + \alpha_2 S^2 \tag{3.118}
\]

\[
\sigma(S, t) = \alpha_0 + \alpha_1 S + \alpha_2 S^2 + \beta_1 t + \gamma_1 St \tag{3.119}
\]

\[
\sigma(S, t) = \alpha_0 + \alpha_1 S + \alpha_2 S^2 + \beta_1 t + \beta_2 t^2 + \gamma_1 St \tag{3.120}
\]

Data on S&P 500 index options are used to test the performance of each of the above models. The evaluation procedure is quite simple. First estimate the volatility function from the market. Next determine the option
prices predicted by the estimated model for dates one week in the future. Compare these estimates with the observed future prices. Under many circumstances, the first of the above models (Black-Scholes constant volatility surface) performs best. There are a number of problems with this testing methodology. First, the parametrized family of volatility functions hardly captures the full range of possible functions for the local volatility surface as developed above. Furthermore, it is not at all clear that the test is really evaluating the performance of the models with respect to their stated goal of pricing exotic options in agreement with traded instruments.

3.13 Directions for Future Research

While there has been a great deal of work on implied probability distributions over the past few years, a number of important questions remain unanswered.

- Can any of the above models be extended to inferring implied volatility functions, or in general implied parameters, for stochastic interest rate models?
- Is it possible to extend the results presented here for the single factor diffusion to determining a stochastic process for the volatility implied either by current option prices, or by historical data? What about for the parameters for mixed jump-diffusion processes (see Bates [6] for a first attempt).
- How do the hedges constructed by using the above models perform on actual market data?
- How stable is the behaviour local volatility surface thorough time? Which of the above methods produces the most stable implied surface?

As mentioned previously, a large amount of empirical testing and comparative study is required before any assessment regarding the quality of the above models from a practical, rather than theoretical, standpoint can be made.
Part II

Implied Scenario Probabilities
3.14 Introduction

In this part of the chapter, we see how the estimation of implied distributions (the inverse of the pricing problem), can be achieved by solving a portfolio selection problem. In this sense, it provides a unification of the theory of the previous chapters. The algorithm for determining implied probabilities presented in this chapter employs the Scenario Optimization model outlined in chapter 2 (we briefly review the relevant theory below). The chapter is structured as follows. We begin with a discussion of the basic duality properties of linear programming. We also present a heuristic argument for the relationship between duality in optimization and the inverse pricing problem. The next section provides a summary of the properties of the risk-reward model relevant to the inverse pricing problem. This is followed by a series of examples explaining the properties of the implied scenario probabilities generated by the risk-reward model in some hedging applications involving call options on the S&P 500 index.

3.15 Duality, Reflection and Inversion

We assume that the reader is familiar with the basic duality properties of linear programming. Specifically, to any linear programming problem

\[
\max_x c^T x
\]

subject to

\[
Ax \leq b
\]

\[
x \geq 0
\]

there corresponds another linear programming problem

\[
\min_y b^T y
\]

subject to

\[
A^T y \geq c
\]

\[1\text{The second part of chapter 3 contains a preliminary version of a working paper by the author. Please do not quote or disseminate without permission. Comments are solicited.}\]
such that problem (3.121) has a solution if and only if (3.124) has a solution, and in this case the optimal values of the problems are equal. The problem (3.121) is referred to as the primal problem, and (3.124) is referred to as its dual problem. The objective value of any feasible solution to the dual problem provides an upper bound for the optimal solution of the primal problem. The converse also holds, the objective value of any feasible solution of the primal gives a lower bound on the optimal solution of the dual. There is an obvious one-one correspondence between the variables of the dual, and the constraints of the primal. The optimal value of a dual variable gives the sensitivity of the optimal value of the primal problem with respect to perturbations in its corresponding constraint. Further information on duality on linear programming, and in more general optimization problems, is contained in the book by Luenberger (?).

Many algorithms for solving linear programming problems exploit the duality properties outlined in the above paragraph, by approaching an optimal solution both from above (via dual feasible points), and below (by primal feasible points). This leads rather naturally to the view that the dual is in some sense a reflection of the primal with respect to the optimal solution. The main result examined in this chapter can therefore be stated as follows: the optimal solution of the reflection of the portfolio selection problem gives the solution to the inverse pricing problem. The situation is summarized by the following diagram

```
Pricing        → Portfolio Selection(Primal)
Inversion ↓         ↓ Reflection
      Probabilities ← Portfolio Selection(Dual)
```

Scenario Optimization provides a methodology for inferring the scenario probabilities implied by asset prices. The probabilities appear as the dual variables in the risk-reward optimization model. It remains unresolved whether similar results are attainable in other portfolio selection models, however the above intuition gives some hope for generalizations.
3.16 Recapitulation: The Regret-Reward Model

3.16.1 The Regret-Reward Frontier

We begin by briefly summarizing the important relations between the geometric properties of the regret-reward frontier and the structure of the market.

- Perfect replication of the benchmark portfolio is possible if and only if the regret-reward frontier hits the regret axis.

- The market is arbitrage free and complete with respect to the benchmark if and only if the regret-reward frontier passes through the origin and has strictly positive slope.

- If the benchmark portfolio crosses the reward axis at a strictly positive regret, then it dominates all possible combinations of tradeable market securities.

3.16.2 Notation

The following notation is employed:

- $S$ - the number of future scenarios.

- $N$ - the number of instruments available to construct the hedge.

- $\tau$ - the vector containing the value of the benchmark under each of the possible future scenarios, $\tau \in \mathbb{R}^n$.

- $c$ - the current market price of the benchmark.

- $D$ - the market payoff matrix $d_{i,j}$ denotes the amount paid by security $i$ under scenario $j$. $D \in \mathbb{R}^{N \times S}$.

- $q$ - the vector containing the prices of the replicating instruments. $q \in \mathbb{R}^S$, $q_i$ is the current market price of security $i$. 
3.16.3 Implied Probabilities

Consider the linear programming formulation of the risk-return problem:

\[
\min_{x,y^+,y^-} p^T y^- \quad (3.127)
\]

subject to

\[
D^T x - y^+ + y^- = \tau \quad (3.128)
\]

\[
p^T (y^+ - y^-) - q^T x \geq K - c \quad (3.129)
\]

\[
y^+, y^- \geq 0 \quad (3.130)
\]

The dual of the above problem is:

\[
\max_{\lambda,\pi} \tau^T \pi + (K - c)\lambda \quad (3.131)
\]

subject to

\[
D\pi - \lambda q = 0 \quad (3.132)
\]

\[
0 \leq \pi - \lambda p \leq p \quad (3.133)
\]

\[
\lambda \geq 0 \quad (3.134)
\]

The first constraint in the dual problem yields that any pair of dual variables \( \pi, \lambda \) with \( \pi \geq 0 \) will provide a market price vector:

\[
D\left(\frac{\pi}{\lambda}\right) = q \quad (3.135)
\]

If the variables \( \pi \) and \( \lambda \) also satisfy

\[
\tau^T \left(\frac{\pi}{\lambda}\right) = c \quad (3.136)
\]

then \( \frac{\pi}{\lambda} \) is a *state price vector* for the market \((D, \tau)\). If we denote \( \psi = \frac{\pi}{\lambda} \), then the vector \( \hat{\psi} \) where:

\[
\hat{\psi}_i = \frac{\psi_i}{\sum_i \psi} \quad (3.137)
\]
provides a probability measure for the scenarios under consideration. The methodology in this paper selects as state price vector \( \psi^* \) corresponding to the optimal dual variables for the linear program with \( K = K^* \) that maximizes the risk-adjusted return:

\[
RAR_r(K) = \frac{K}{MR_r(K)}
\]  

(3.138)

The mathematical properties of the implied scenario probabilities are outlined in detail in Dembo [20], or see the second chapter of this thesis.

3.16.4 Role of the Model Inputs

The fundamental inputs to any application of the Scenario Optimization algorithm are the scenarios and the benchmark. The scenarios prescribe the null sets, and hence determine the set of eligible pricing measures. The benchmark provides a criterion by which the performance of each market security is measured. Standard optimization routines only examine return, which can be regarded as using a null benchmark. The use of a benchmark transforms the weighting given to certain events (hence altering the geometry of the optimization problem).

3.17 Examples

In this section, we proceed through a number of examples illustrating the use of the risk-reward framework for determining implied scenario probabilities. We begin with an example in an idealized market, and then introduce market imperfections to make the situation more realistic. We examine the sensitivity of the procedure to its key inputs: the scenarios under consideration, the prior distribution assumed for the scenarios, and the instruments used to construct the hedge.

3.17.1 Perfect Markets
Uniform Prior Distribution

To begin, we consider a discrete approximation to the Black-Scholes model. The perfect market assumptions hold (subject to group constraints- see below). There are no transactions costs are taxes. We consider a replicating portfolio consisting of a large number of European call options on a market index. Each option expires in 164 days, and we shall generate scenarios based on the value of the index at the expiration date. The strike prices for the options range from 250 to 500, with a strike price available at each multiple of 5 in between. The current level of the index is 354.75. The risk-free interest rate is 9%, expressed in annual compounding. The underlying index is assumed to pay no dividends during the life of the options. All options are priced using the Black-Scholes formula, with volatility \( \sigma = 0.171 \). Here, according to market convention, volatility is expressed as an annual rate.

We optimize with respect to a set of evenly spaced scenarios on the value of the underlying index at expiration. The scenarios range from an extreme drop (shift of -100) to a large jump (+170), with a scenario at each multiple of 5 in between. The scenarios were generated using a RiskScript macro written by Ben DePrisco. To begin with, we assume a uniform (equally weighted) prior distribution. This is appropriate, since the uniform is the distribution that provides the least “bias” or “prior information” to the model. (This intuitive notion is formalized by the entropy of the distribution - see section A of this chapter). The example will therefore determine whether, given the Black-Scholes prices, and no additional information, the Scenario Optimization model will infer the lognormal density that generated the prices.

The freedom in the choice of benchmark portfolio is an important advantage of Scenario Optimization. Most of the examples in this paper are calculated with a null benchmark (paying off zero under all scenarios), although in some cases we also consider replicating the payoff of one of the options with the others. In order to avoid getting an optimal solution that involves no trades, we impose additional group constraints to ensure that some trading is done. Since these group constraints may be binding in the optimal solution, they may have important implications for the behaviour of the implied scenario
probabilities (see the section on group constraints below). Optimization is performed using the RiskWatch risk management package, distributed by Algorithmics Inc. In order to solve the parametric linear programming problem to generate the risk-return frontier, repeated calls are made to the CPLEX linear programming package.

The risk-return model, using the null portfolio as benchmark, and the full range of options described above produces the following risk-reward frontier:

![Figure 1: Frontier in Perfect Market](image)

Since the above curve passes through the origin, we know that the market is complete with respect to the benchmark. Further, the positive slope at zero indicates that there is no arbitrage in the model. The optimal dual variables at the point where the Risk-Adjusted-Return is maximized yield the following risk-neutral scenario probabilities:
which give a reasonable discrete approximation to the continuous time log-normal density. Thus, the risk-return model has passed the first test, and we proceed to examining the sensitivities of the above density to changes in the model inputs.

**Tail Scenarios with a Uniform Prior**

Often we are interested in the performance of the hedge under extreme movements in the market. To test the performance of Scenario Optimization under such movements, we considered the above optimization problem with additional scenarios in the left tail. Specifically, we added scenarios with downward shifts of -120, -115, -110, and -105 to the data. We maintain the assumption of a uniform prior distribution (see below for a parallel case with a lognormal prior distribution). An important observation in this example is that none of the options in our replicating portfolio will expire in the money under any of the newly generated scenarios. As can be seen from the risk-reward frontier:
markets remain complete, and there is no possibility of arbitrage with respect to the null benchmark. This should not be surprising. Since each of the options has value identically zero in the new scenarios, and the target also has value zero, any portfolio weights for the options will reproduce the benchmark exactly. It follows from both elementary linear algebra and financial intuition that the introduction of new scenarios could not possibly add arbitrage to a market where it was previously absent. The dual benchmark-neutral probabilities yield the following (initially surprising) picture:

Upon closer consideration these probabilities are not so strange. In a sense, the model has no instruments that it can use to obtain “information” regarding these scenarios. It therefore ascribes the uniform prior probabilities
to them (we shall consider the parallel case with a lognormal prior below). We postulate that if the market traded options with lower strike prices, the risk-reward model would have more information, and therefore would better reproduce the Black-Scholes lognormal. Repeating the optimization with a richer set of options, we find that this is in fact the case. We include four more options with strike prices lower than those contained in the original portfolio (maintaining the constant spacing on strike prices).

Benchmark Neutral Probabilities – THEO/Value

It is interesting to note that the above is a property only the left tail of the distribution. Extreme right tail mean that all options expire in the money, with linear payoffs of the same slope, but separated by strike. Different options have different payoffs under each of these scenarios, and therefore it is possible to infer some information regarding implied probabilities from these scenarios.

Observant readers will have noticed a slight discrepancy between the lognormal distribution and the discrete approximations shown above in the right tail of the distribution. This has a simple explanation in view of the model variables, and also has important applications for our next section. The implied probability distribution seems to follow the lognormal exactly until the up move of 130, and then drops off. Probabilities are then zero until the final scenario, an up move of 170). There is a slight “spike” of probability mass at this point. The drop corresponds to the strike price of the last traded option (500). The “spike” is the final scenario under consideration.
The next section presents a more dramatic illustration of this phenomenon.

**Smaller Replicating Portfolio**

The above examples assumed that an unreasonably large range of options are available on the market. In reality, the available strike prices will be much sparser, and even some of the traded options (those with very high or very low strike prices) will be subject to liquidity constraints. In order to address this issue, we considered a reduced replicating portfolio, with only call options with only the following strike prices: 250, 275, 300, 325, 330, 335, 340, 345, 350, 355, 360, 365, 370, 375, 380, 385. The motivation for our choice was that options with these strike prices with 164 days to expiration were traded on the S&P 500 on January 2, 1991 (corresponding to an index level of 354.75). The data was acquired from Rubinstein [72]. From the risk-reward curve:

![Risk-Reward Curve](image)

we see that the market remains complete and that there is no arbitrage. The risk-neutral probabilities exhibit the following behaviour:
Inside the range where the strike prices are dense (325-385, corresponding to shifts in the index of approximately -30 to +30), the lognormal distribution is reproduced nicely. In the left tail, there are “spikes” of implied probability corresponding to the strike prices of the available options. The drop in probability in the right tail is very pronounced. The shape of the implied distribution appears to be determined principally by the prices of traded instruments within the range of available strikes, and by the prior distribution outside this range (see the section below, where a similar calculation is done with a lognormal prior for further justification of this hypothesis).

### 3.17.2 Varying the Prior Distribution: Lognormality

The assumption of a uniform prior distribution in the previous section was used in order to avoid any bias in the prior distribution. We now proceed to examine the affects of just such a bias. Typically, investors will have their own intuitions about the weights that are appropriate for various market scenarios (these need not necessarily be beliefs about the actual probabilities of the future scenarios, and could incorporate risk aversion - see Dembo [20]). Given that we know that the prices were generated using the lognormal distribution assumed by the Black-Scholes model, it would be interesting to see how our model performs with a lognormal assumption for the prior distribution.

In order to facilitate comparison, we consider the identical scenario sets as
above, except with different scenario weights. The following algorithm was used to generate the lognormal distribution. The extreme tail scenarios are assumed to be sufficiently far enough out to have zero probability. For any other scenario, \( S_n \) (where scenarios are ordered by increasing value for the underlying), we calculate the probability mass assigned by the lognormal to the interval:

\[
\left[ \frac{S_n + S_{n-1}}{2}, \frac{S_n + S_{n+1}}{2} \right]
\]  

(3.139)

and assign this probability to the scenario \( S_n \). Of course, in order to use the lognormal in the Black-Scholes model, rates must first be converted to continuous compounding. This results in a risk-free interest rate of 0.086177696 and a volatility of 0.15785808. In the calculations, a function for calculating values of the cumulative normal distribution written by Alex Kreinin was used.

For the first example, we use all the options, and optimize with respect to the null portfolio. This results in the following risk-return frontier:

Risk–Reward Curve

![Risk–Reward Curve](image)

---

118
and the corresponding risk-neutral probabilities:

![Benchmark Neutral Probabilities – THEO/Value](image)

The agreement with lognormality here is not surprising. We have essentially repeated the first elementary test above, except with the true risk-neutral distribution as the prior distribution.

Recall that when we included additional tail scenarios with a uniform prior the effect produced on the tail of the implied distribution was quite dramatic. It was noted then that the implied distribution followed the shape of the prior distribution on the region where the call options and the null portfolio cannot be distinguished. When we replace the prior uniform with the lognormal distribution, we obtain the following graphs:
Again, we find that the implied distribution closely mirrors the prior, however in this case it is the near zero lognormal distribution, rather than uniform weights, that are obtained. Since the implied density so closely resembles the true risk-neutral density that generated the prices, "smoothing" the distribution by introducing options with even lower strike prices (as was done with the uniform prior) is no longer necessary.

We conclude this section by considering the sparser set of strikes corresponding to the options traded on the S&P 500 on January 2, 1991 (see above for details). For the uniform prior, this resulted in a dramatic change in the character outside of the strike prices of traded options of traded options.
The probability “spikes” still occur, giving a departure from strict lognormal, but the external knowledge of the correct prior distribution brought results much closer to the lognormal that generated the prices originally. The departure in the left tail from lognormality (and the resulting correction) bear a similar interpretation to the uniform case, corresponding to the beginning of sparsity in the strike prices, and the final traded strike. The interpretation of the right probability “spike” corresponding neither to the final strike price, nor the highest scenario value. Suggestions regarding the interpretation of this graph are, of course, welcome.
3.18 Performance on Market Data

Now that we have developed an intuition for how the model performs in idealized situations, let us examine its performance on a set of market data. We use the data given in Rubinstein [72]. The dataset consists of the call options on the S&P 500 index with maturity in 164 days on January 2, 1990, and is given for completeness in an appendix. The data illustrate the downward sloping volatility smile that has been characteristic of index options markets in recent years (see part A of this chapter):
The risk-free interest rate is taken to be 9% (expressed with annual compounding). The index is assumed to pay out a dividend yield at 3.6% expressed in annual compounding (3.537% in continuous compounding). The market also incorporates a significant bid-ask spread (see the dataset in the appendix). Optimizing with the same benchmark and constraints as in the previous examples yields the following risk-reward frontier:
The optimal duals variables give the following implied probabilities:

**Benchmark Neutral Probabilities – THEO/Value**

The resulting distribution exhibits the excess weight in the left tail that is prevalent in the probability distribution implied by index options after the market crash in 1987. The distribution is still highly dependent on the strike prices of the options that are included in the replicating portfolio.

### 3.18.1 Regarding the Data

Some of the variables used as input to the above optimization are not stated directly in the paper source (namely Rubinstein [72]). Compounding con-
ventions are not specified, and the dividend payout rate for the S&P is not given. We have tried to “back out” these parameters from the data on implied volatilities that is provided, but make no claims that the data use as input to the Scenario Optimization model is identical to the S&P 500 data for January 2, 1991, as used by Rubinstein. Since we have taken care to declare all input variables, and the dimensions used to express them, the examples in this paper should be completely reproducible.

3.18.2 Group Constraints

The theoretical results presented in this section did not consider the effects of additional constraints in the optimization model. This affects the structure of the optimization problem, and therefore may influence the choice of the vector of risk-neutral probabilities. For further details, see [73].

3.19 Concluding Remarks and Directions for Future Research

We consider the following as directions for important future research on implied scenarios probabilities:

- Determine the mathematical nature of the sensitivity of the implied distributions to changes in the input probabilities.

- Consider the behaviour of the implied distributions under changes in the benchmark portfolio.

- Can the relationship between duality in optimization and the determination of implied probabilities (inverse of the pricing problem) be extended to a more general (continuous time) setting? (for a discrete time approach to the fundamental theorem of asset pricing using optimization, see Rogers [69], continuous time generalizations remain elusive).

We have developed the mathematical properties of the scenario probabilities implied by the risk-return optimization model. A number of examples
have illustrated the properties of these probabilities where prices are given by a simple theoretical model that is well understood. It is important to investigate the behaviour of the model in these familiar situations so that the results obtained in more complicated problems can be properly interpreted. We have also identified avenues for future research and interesting open problems.
Bibliography


[64] Bruce Mizrach, *Did option prices predict the emt crises*, working paper, 17, 1996.


132


