Derivative Pricing and Optimal Execution of Portfolio Transactions in Finitely Liquid Markets



Michael David Mitton

The Queen's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2005

To my parents.

Acknowledgements

First and foremost I would like to thank my supervisor, Sam Howison, for providing the guidance and support so crucial to allowing me to pursue this research. Your tolerance and patience in attempting to convert this ex-physicist into a mathematician has been commendable and I have learned a lot from you. I also owe a massive debt of gratitude to my parents whose support, both financial and emotional, has been incalculable. What I am today could not have happened without you and I will never forget it. Special thanks go to The Queen's College, The University of Oxford, and the government of the United Kingdom for funding this research through the Cyril and Phillis Long Studentship, the Clarendon Fund Award, and the Overseas Research Studentship, respectively. Finally, I would like to say thank-you to the amazing network of friends and family that have surrounded and supported me over the years. What remains of my sanity only does so because of you.

Abstract

In real markets, to some degree, every trade will incur a non-zero cost and will influence the price of the asset traded. In situations where a dynamic trading strategy is implemented these liquidity effects can play a significant role. In this thesis we examine two situations in which such trading strategies are inherent to the problem; that of pricing a derivative contingent on the asset and that of executing a large portfolio transaction in the asset.

The asset's finite liquidity has been incorporated explicitly into its price dynamics using the Bakstein-Howison model [4]. Using this model we have derived the no-arbitrage price of a derivative on the asset and have found a true continuous-time equation when the bid-ask spread in the asset is neglected. Focussing on this pure liquidity case we then employ an asymptotic analysis to examine the price of a European call option near strike and expiry where the liquidity effects are shown to be most significant and closed-form expressions for the price are derived in this region. The asset price model is then extended to incorporate the empirical fact that an asset's liquidity mean reverts stochastically. In this situation the pricing equation is analyzed using the multiscale asymptotic technique developed by Fouque, Papanicolaou, and Sircar [22] and a simplified pricing and calibration framework is developed for an asset possessing liquidity risk. Finally, the derivative pricing framework (both with and without liquidity risk) is applied to a new contract termed the American forward which we present as a possible hedge against an asset's liquidity risk.

In the second part of the thesis we investigate how to optimally execute a large transaction of a finitely liquid asset. Using stochastic dynamic programming and attempting only to minimize the transaction's cost, we first find that the optimal strategy is static and contains the naive strategy found in previous studies, but with an extra term to account for interest rates neglected by those studies. Including time risk into the optimization procedure we find expressions for the optimal strategy in the extreme cases when the trader's aversion to this risk is very small and very large. In the former case the optimal strategy is simply the cost-minimization strategy perturbed by a small correction proportional to the trader's level of risk aversion. In the latter case the problem is shown to be much more difficult; we analyze and derive implicit closed-form solutions to the much-simplified perfect liquidity case and show numerical results to demonstrate the agreement of the solution with our intuition.

Contents

1 Introduction **1**

1.1 Background Material and Terminology 2

1.2 Derivative Pricing in a Perfectly Liquid Market 3

1.2.1 American Contracts in the Black-Scholes Model 5

1.3 Structural Characteristics of a (More) Realistic Market 7

1.4 Literature Review 9

1.5 The Bakstein-Howison Liquidity Asset Price Model 13

1.5.1 Calibration of the Model Parameters 18

1.6 Outline of the Thesis 19

2 Derivation of Derivative Pricing Equations **20**

2.1 Leading-Order Equation for Arbitrary γ, λ, and α 21

2.2 The Case $\gamma \ll \delta t$ 25

2.3 American Contracts in the Bakstein-Howison Model 28

2.4 Conclusions 29

3 Asymptotic Analysis of the European Call Option in the $\gamma = 0$ Model **31**

3.1 Introduction 31

3.2 The Outer Solution 33

3.3 Inner Solution 37

3.4 Matching 44

3.5 Conclusions 46

4 Hedging Liquidity Risks with the American Forward 47
4.1 Liquidity Risks in an Imperfect Market 47
4.2 Hedging Liquidity Risks 48
4.3 European Forward in the Black-Scholes Model 49
4.4 American Forward in the Black-Scholes Model 50
4.4.1 The Case $D = 0$ 51
4.4.2 The Case $D > 0$ 51
4.5 The Forward in the $\gamma = 0$ Model 53
4.5.1 The Case $D = 0$ 54
4.5.2 The Case $D > 0$ 54
4.6 Conclusions 58

5 Derivative Pricing in a Market with Stochastic Liquidity - Part I 60
5.1 Previous Literature 61
5.2 Outline of Work 62
5.3 Asset Price Model 63
5.4 Deriving the Pricing Equation 64
5.5 Asymptotic Analysis of the Pricing Equation 67
5.6 Calibration of the Model 74
5.6.1 Region of Validity 76
5.7 Conclusions 77

6 Derivative Pricing in a Market with Stochastic Liquidity - Part II 78
6.1 The Stochastic Liquidity Bakstein-Howison (BH) Asset Price Model 79
6.2 Derivation of the Stochastic Market Depth Pricing Equation 80
6.3 Rescaling the Stochastic Market Depth Pricing Equation 84
6.4 Analysis of the Rescaled Pricing Equation 85
6.4.1 The Case $(c = d, d)$ 86
6.4.2 The Case $(c = 2d, d)$ 94
6.4.3 The Case $(c > 2d, d)$ 98
6.5 Calibration of the Stochastic Market Depth Model 104
6.5.1 Calibration Procedure 112

6.6 The American Forward with Liquidity Risk 113
6.6.1 Solving for the American forward Price 118
6.7 European Call Option with Liquidity Risk 124
6.8 Conclusions 128

7 Minimizing Transaction Costs in a Finitely Liquid Market 131
7.1 Previous Literature 132
7.2 Outline of work 134
7.3 The Bakstein-Howison Liquidity Asset Price Model 135
7.4 Formulation of the Cost-Minimization Problem 137
7.5 Derivation of the Equivalent Bellman Equation Formulation 138
7.6 Solving for the Optimal Trading Strategy 141
7.7 Discussion 145

8 Optimal Liquidation with Small Risk Aversion 147
8.1 Outline of Chapter 147
8.2 Formulation of the Mean-Variance Problem 148
8.3 Difficulties with Variance Minimization 149
8.4 The Auxiliary Formulation for the Risk-Adjusted (RA) Problem 151
8.5 The Bellman Equation for the Auxiliary Problem 155
8.6 Solving for the Risk-Adjusted Optimal Trading Strategy 156
8.6.1 Solving the Auxiliary Equation 156
8.6.2 Solving the Auxiliary Problem for Small Risk Aversion 160
8.6.3 Finding the Solution of the RA Problem from the Auxiliary Problem 164
8.7 Discussion 166

9 Optimal Liquidation with Large Risk Aversion 170
9.1 A Simplified Risk-Adjusted Problem for Large Λ 172
9.2 Solution of the Auxiliary for the Perfectly Liquid RA Problem 174
9.3 Discussion of Analytical Results 182
9.4 Calculating the Solution of the RA Problem 183
9.5 Numerical Results and Further Discussion 185

10 Conclusions and Areas for Future Study **189**

List of Figures

1.1 An order book for a fictitious asset with a spot price of 100. 8
1.2 Asset price tree representing the price process given by equations (1.20) - (1.22). 17

3.1 The first liquidity correction to the Black-Scholes value of a European call option in the outer region. 38
3.2 Plots of $u_0(y)$ for $\theta = 0.1$, 1, and 10. 40
3.3 Numerical solution of $\hat{V} - \max(x, 0)$ and its leading-order asymptotic approximation for $\tau = 10^{-10}$. 43

4.1 Plots of $\tilde{C}^{a,BS}(t, S)$ and $\tilde{P}^{a,BS}(t, S)$ for the American forward. 52
4.2 Plots of $\tilde{C}^{a,BS}(t, S) - \max(\Phi^C, \tilde{C}^{e,BS}(t, S))$ and $\tilde{P}^{a,BS}(t, S) - \max(\Phi^P, \tilde{P}^{e,BS}(t, S))$ for the American forward. 53
4.3 Long and short positions of the American call and put-forwards in the $\gamma = 0$ model. 55
4.4 The deviation of the plots of Figure 4.3 from their corresponding Black-Scholes values. 56

6.1 The seven regions of the (c, d) space over which the form of the rescaled pricing equation (6.20) varies. 86
6.2 The first correction, ϵh_{1_2}, to the bid and ask prices of both the call and put forward contracts. 120
6.3 The second correction, $\epsilon^{\frac{3}{2}} h_{1_3}$, to the bid and ask prices of both the call and put forward contracts. 121

6.4 Dependence of the second correction, $\epsilon^{\frac{3}{2}} h_{1_3}$, to the bid and ask prices of the call forward on the parameter κ_2. 122
6.5 Dependence of the second correction, $\epsilon^{\frac{3}{2}} h_{1_3}$, to the bid and ask prices of the put forward on the parameter κ_2. 123
6.6 The dominant liquidity correction to the Black-Scholes ask price of a European call option. 126
6.7 The second (liquidity risk) correction to the Black-Scholes ask price of a European call option 127
6.8 The implied market depth as a function of strike, K, for the ask price of a European call option. 128

8.1 The dynamic correction to the leading-order optimal strategy for several asset price paths with small Λ/λ. 168
8.2 The difference between the dynamic correction for the increasing price process and the constant price process and the difference between the dynamic correction for the decreasing price process and the constant price process. . 169

9.1 Optimal holdings for a trader with a large level of risk aversion in a perfectly liquid market. 185

Chapter 1

Introduction

In this thesis we will examine two problems related to the finance of finitely liquid markets. By a finitely liquid market we mean one in which there exist costs associated with trading the asset and where the trading itself has a feedback effect on the asset price; conversely, in a perfectly liquid market no such transaction costs or feedback effects exist. In the first part of our work we will develop a framework for pricing derivatives in a finitely liquid market and then focus on three applications and extensions of the model. In the second part we will then examine the problem of how to optimally execute a large portfolio transaction in such a market.

As our work will centre around the costs associated with trading an asset, a model for the dynamics of the asset price that incorporates trading effects is crucial. The model we will use is the Bakstein-Howison (BH) liquidity model [4] as it possesses a structure portable across a wide range of applications. The main purpose of this first chapter, along with introducing the necessary background material and concepts, will be to motivate and develop the BH model for use in the rest of our work.

While the two parts to this thesis are related the by BH model, they are somewhat independent pieces of work and will be treated as such. As the work on derivative pricing will be presented first, we will motivate the development of the BH model from a derivative pricing perspective and bias the introduction in this direction. When we present our work on the optimal execution of portfolio transactions we will then fully introduce that subject in its own right.

The rest of this chapter will proceed as follows: In Section 1.1 we will briefly introduce some necessary background material and terminology. In Section 1.2 we present a common model for the dynamics of an asset price in a perfectly liquid market and then derive the arbitrage-free price of a European vanilla contract on that asset. In Section 1.3 we discuss the consequences of omitting liquidity effects in the analysis of Section 1.2 and then discuss several features of more realistic markets. In Section 1.4 we discuss the previous literature associated with incorporating liquidity effects into a derivative pricing framework and then use these previous studies to motivate our development of the BH model in Section 1.5. In Section 1.6 we will conclude the chapter by outlining the rest of the thesis.

1.1 Background Material and Terminology

A *derivative* is a general term for any financial contract whose value depends on some more fundamental *underlying* asset. Possible forms of the underlying are *equities* such as public shares of IBM or Microsoft, *fixed-income* instruments such as government bonds or commercial paper, *currencies*, or even other derivatives.

Our work on derivative pricing will focus on two specific types of contracts. First, a *forward* is an agreement between two parties to exchange an amount of the underlying for cash at some specific date (the *expiry date*) in the future. The amount of cash to be paid (called the *forward price*) at expiry is determined when the contract is formed and is set so that there is no cost to either party to enter into the agreement. On the other hand, a *call option* is a contract that gives the holder the right, but not the obligation, to purchase the underlying for a pre-specified price (the *strike price*) at the expiry date. Once the contract has been initialized the holder has greater rights and therefore must pay a premium to the writer to compensate for this asymmetric risk. It is the value of the premium (or the value of the forward price in the case of a forward contract) and how it is affected by liquidity effects that will concern us in this first part of the thesis.

One way of classifying derivatives is based upon when the contract can be exercised. As they have been described, both the forward and call option mentioned above are *European* contracts since they can only be exercised at the expiry date. *American* contracts, on the other hand, give the holder the extra freedom of being able to exercise at *any* time up to, and including, expiry.

Crucial to the theory of derivative pricing is the concept of *arbitrage*. An arbitrage opportunity is one in which a market participant can make an instantaneous, risk-free profit above that of the *risk-free rate*, r. The risk-free rate is the rate of return provided by an asset that possesses essentially zero risk, such as a short-term US or UK government bond.

1.2 Derivative Pricing in a Perfectly Liquid Market

The original theory for pricing derivatives was developed by Black and Scholes [8]; to illustrate the deficiencies of this 'ideal' model we will first present their analysis here.

Central to any derivative pricing model is a description of the underlying's dynamics. In the Black-Scholes world we assume the asset follows a *Geometric Brownian Motion*. Let S be the asset price at time t and dS be the change in S over an infinitesimal time dt; if μ, D, and σ, respectively, are the growth rate, dividend yield, and volatility (all assumed to be constant) of the asset, then the dynamics of S are given as

$$\frac{dS}{S} = (\mu - D)dt + \sigma dX, \tag{1.1}$$

where dX is a Wiener process with Gaussian increments of mean 0 and variance dt.

Let the value of a derivative be $V(t, S)$. We will assume V is a continuous function of the current asset price, S, and the time t. The change in the derivative's value over dt is given by its Taylor series expansion

$$dV(t, S) = V_t dt + V_S dS + \frac{1}{2}V_{tt}dt^2 + \frac{1}{2}V_{SS}dS^2 + V_{tS}dtdS + \ldots, \tag{1.2}$$

where $V_t = \partial V/\partial t$, $V_{tt} = \partial^2 V/\partial t^2$, *etc* .... Substituting (1.1) into equation (1.2) gives

$$dV(t, S) = \left[V_t + (\mu - D)SV_S + \frac{1}{2}\sigma^2 S^2 V_{SS}\right] dt + \sigma S V_S dX + O(dt^{3/2}), \tag{1.3}$$

where we have used the result that, for any Wiener process, $\{X_t\}$, $dX^2 = dt$.

The remainder of the analysis requires several simplifying assumptions; they will be listed as they are required.

Assumption 1: Selling assets 'short' (ie. selling assets we do not own) is permitted.

Assumption 2: Assets may be bought and sold in limitless and fractional amounts.

We now construct a portfolio consisting of long one derivative and short Δ units of the underlying (assumption 1 is necessary to allow the short selling of the Δ assets); the value, $\Pi(t, S)$, of this portfolio is

$$\Pi(t, S) = V(t, S) - \Delta S. \tag{1.4}$$

Assumption 3: There are no transaction costs associated with trading in the underlying and all traders are considered price takers; that is, the traders' trades do not affect the price of the asset.

Assuming there is no cost involved in trading the underlying then the change in the value of the portfolio over dt is

$$d\Pi(t, S) = dV(t, S) - \Delta dS - \Delta DSdt, \tag{1.5}$$

where the $-\Delta Ddt$ term is a result of the fact that we receive a dividend payment of $DSdt$ for each asset and we hold $-\Delta$ of them, thus removing $\Delta DSdt$ from the value of the portfolio. Substituting equations (1.1) and (1.3) into (1.5) yields

$$d\Pi(t, S) = \left[V_t + (\mu - D)SV_S + \frac{1}{2}\sigma^2 S^2 V_{SS} - \Delta DS - \Delta(\mu - D)S\right] dt + \sigma S \left[V_S - \Delta\right] dX. \tag{1.6}$$

Choosing $\Delta = V_S$ eliminates the leading-order stochastic component of $d\Pi$ by matching the risk in the derivative with the risk in the underlying; with this choice of Δ equation (1.6) reduces to

$$d\Pi(t, S) = \left[V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} - DSV_S\right] dt, \tag{1.7}$$

leaving the change in value of the portfolio over dt completely deterministic.

Assumption 4: The market for the underlying is arbitrage-free.

With our specific choice of $\Delta = V_S$, the dynamics of the portfolio are deterministic and it is therefore risk-free. Since the market is arbitrage-free and the portfolio is risk-free it must earn a rate of return equal to the risk-free rate, r. Mathematically this can be written as

$$d\Pi(t, S) = \left[V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} - DSV_S\right] dt = r\Pi dt. \tag{1.8}$$

Substituting (1.4) into (1.8) above and dividing by dt gives

$$V_t + (r - D)SV_S + \frac{1}{2}\sigma^2 S^2 V_{SS} - rV = 0. \tag{1.9}$$

Equation (1.9) is the Black-Scholes equation and gives the price of any European derivative that depends only on the current values of S and t. The equation is second-order backwards parabolic and so we must specify two boundary conditions and one final condition. The valuation of specific contracts is incorporated through these boundary conditions. For instance, if T and K denote the expiry date and strike price, respectively, then the boundary conditions for a European call option are $V(T, S) = \max(S - K, 0)$, $V(t, 0) = 0$, and $V(t, S) \sim S$ as $S \to \infty$.

The deterministic nature of equation (1.7) was obtained through the choice $\Delta = V_S$; since V_S is dependant on both t and S it follows that equation (1.9) is only instantaneously valid at time t. In order to maintain a risk-free portfolio the hedge quantity, Δ, will have to be adjusted at every $t \in [0, T)$ by trading in the underlying. This trading strategy, known as a *delta-hedging* strategy, is therefore fundamental to the concept of a Black-Scholes value of the derivative. Assumption 3 is necessary so that we are able to carry out the delta-hedging strategy continuously and without cost.

1.2.1 American Contracts in the Black-Scholes Model

In Chapters 4 and 6 we will examine an American version of a forward contract; *i.e.* a forward where the holder has the right of early exercise. The Black-Scholes pricing framework developed above is valid only for European contracts. When early exercise is permissible however, not only must we find V at a given point, (t, S), but also whether the contract should be exercised at that point and so we must modify the analysis to take this factor into account.

To formulate the problem, consider a generic derivative contract with value $V(t, S)$ and payoff function, $\Phi(S)$. Now imagine there exists some point (t, S) during the contract's life such that $V(t, S) < \Phi(S)$. If the contract is American then this situation causes an arbitrage opportunity: We could buy the contract for $V(t, S)$, immediately exercise it for $\Phi(S)$ and therefore make an instantaneous risk-free profit of $\Phi(S) - V(t, S) > 0$. For an American derivative we must therefore impose the constraint $V(t, S) \geq \Phi(S)$ for any (t, S)

when the contract is exercisable.

We now split S into two regions, $Z = \{S : V(t,S) > \Phi(S)\}$ and $Z^* = \{S : V(t,S) = \Phi(S)\}$, and define the Black-Scholes operator $L_{BS}[V] \equiv V_t + (r-D)SV_S + \frac{1}{2}\sigma^2 S^2 V_{SS}^2 - rV$; $L_{BS}[V]$ has the physical interpretation of the difference between the return provided by the hedged portfolio and that of the equivalent amount invested in a risk-free asset. In general $\Phi(S)$ is not a solution of the Black-Scholes equation and so $L_{BS}[\Phi(S)] \neq 0$ for all $S \in Z^*$; since $L_{BS}[V(t,S)] > 0$ would allow for an arbitrage opportunity we can conclude that $L_{BS}[V(t,S)] < 0$ for $S \in Z^*$. But now imagine holding V for some $S \in Z^*$; since $L_{BS}[V(t,S)] < 0$, the return on our portfolio is less than it would be if the equivalent amount were invested in a risk-free asset. For all $S \in Z^*$ the optimal strategy is to exercise the contract and the region Z^* is therefore referred to as the *exercise* region.

In the region Z we have $V(t,S) > \Phi(S)$ for all $S \in Z$. In this region the contract is free to satisfy the Black-Scholes equation and so $L_{BS}[V(t,S)] \equiv 0$ for all $S \in Z$. Since $L_{BS}[V(t,S)] = 0$ the return on our hedged portfolio is the same as it would be if the equivalent amount were invested in a risk-free asset; there is no advantage to exercising the contract and so the optimal strategy for $S \in Z$ is to hold.

The valuation problem thus far has $L_{BS}[V(t,S)] = 0$ where $V(t,S) > \Phi(S)$ and $L_{BS}[V(t,S)] < 0$ where $V(t,S) = \Phi(S)$ along with the contract specific boundary conditions. The problem is that the boundary points between Z and Z^* are not known. Let us denote these *free boundaries* by $S_f^i(t)$ where $i \in 1, 2, \ldots m$ for a contract with m such points. To fix the position of the free boundaries we must impose two further constraints; from arbitrage considerations both V and V_S must be continuous across $S_f^i(t)$. The fully specified valuation problem for an American derivative with m free boundaries is therefore

$$\begin{aligned} L_{BS}[V(t,S)] &= 0 && S \in Z, \\ L_{BS}[V(t,S)] &< 0 && S \in Z^*, \\ V(t,S_f^i(t)) &= \Phi(S_f^i(t)), \\ V_S(t,S_f^i(t)) &= \Phi_S(S_f^i(t)), \end{aligned} \quad \Bigg\} \quad i = 1, \ldots m \tag{1.10}$$

with contract specific boundary conditions.

1.3 Structural Characteristics of a (More) Realistic Market

Generally, there are three types of participants in a market: *sellers* who submit sell orders (or *asks*) which provide a supply of the asset; *buyers* who submit buy orders (or *bids*) and provide demand for the asset; and a centralised *market maker* who accepts and displays these orders and then forms a mean spot price based on the aggregate supply and demand.[1]

Because of differing personal opinions about the present worth and the future prospects of the company that has issued the underlying there will usually be a range of buy and sell orders in the market. A representation of a fictitious asset's *order book* is shown in Figure 1.1. The assets available to market participants are arranged in price *layers* at regularly spaced intervals. The bid and ask layers closest to the spot price will be referred to as the *optimal* buy and sell levels, respectively, and all other layers as the sub-optimal layers. The distance between the optimal buy and sell prices is known as the bid-ask spread. It exists to provide insurance for the market maker and can therefore be thought of the cost of immediate access to the asset [42].

We introduce the concept of the asset's *market depth* as the total number of units of the asset available to be bought or sold at a given price. One possible measure for this quantity is the inverse of the slope of the best-fit line through the asset's order book; for the fictitious asset in Figure 1.1 this is shown as a dashed line. A 'deep' market will have a very small slope so that even a very large trade can be completed entirely within the optimal layer. If the market for the asset is shallow, however, then a trader may need to tap sub-optimal layers forcing up (down) the average price paid (obtained) per asset in order to complete the trade. Because it permits them to get in and out of their positions more cheaply and easily, rational traders will always prefer a deep market to a shallow one.

A trade occurs in the market when a buyer (seller) agrees to the price of one of the sell (buy) orders. At this point the market maker removes the appropriate orders from the board and readjusts the spot price to reflect the new supply-demand equilibrium.

In the Black-Scholes analysis above, it was found that the concept of a delta-hedging

[1] We note that this is a somewhat simplified view of the market maker's role within a centralised market. Often, instead of simply displaying prices, the market maker will hold stock of the underlying to provide traders with immediate supply and demand for the asset. The simplified view, however, *is* a good assumption in a market such as an electricity market where the market maker does not does not usually have the capacity to store the asset.

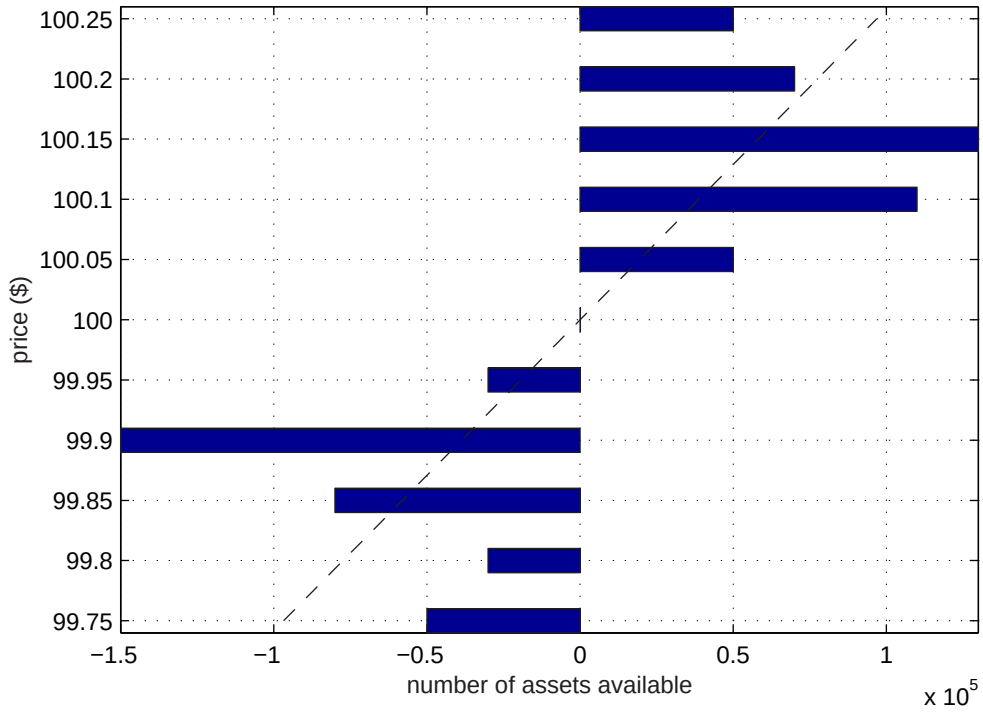


Figure 1.1: An order book for a fictitious asset with a spot price of 100. Bars represent individual market layers; negative numbers indicate bid orders. The dashed line is the best-fit line to the data and its slope is a possible measure for the inverse market depth of the asset.

strategy was central to the value of a derivative. But from the above discussion it is clear that in a real market every trade will incur some positive transaction cost. Not only will a trader not be able to buy or sell at the spot price, but his average transaction price may also be further above (below) the price of the optimal layer if the trade is large and more than one layer needs to be tapped. Since there are an infinite number of rehedging trades required to maintain a delta-neutral portfolio and since not all traders have the same level of transaction costs it is not immediately clear whether the derivative's value in this type of market is finite or even unique.

1.4 Literature Review

There have been many studies into incorporating the transaction cost effect into an asset price model for the purpose of derivative pricing; a good review by Whalley is presented in [43]. In [33] Leland develops a pricing model in continuous time with discrete rehedging for single options in which transaction costs are proportional to the value of the traded quantity. He derives an equation for the option's value similar to Black-Scholes, but with a volatility modified by the rehedging interval and the level of transaction costs. He also derives a trading strategy analogous to the delta-hedging strategy and is able to show that as the trading interval tends to zero this strategy replicates the option's payoff inclusive of its total transaction costs. Boyle and Vorst [9] develop a similar model to Leland, but in discrete time and find similar results as the number of timesteps tend to infinity and/or the transaction cost proportionality constant tends to zero. In both of these studies, only single options are analyzed and the results are valid only for contracts with strictly concave or convex payoff functions. Hoggard, Whalley, and Wilmott [27] extend Leland's model to contracts with arbitrary payoffs and derive a non-linear pricing equation for a derivative's value with discrete rehedging.

As it will prove to be a useful comparison with our results in Chapter 2 we will briefly outline the proportional transaction costs analysis of Hoggard, Whalley, and Wilmott [27]. We begin in a discrete-time setting where the asset price dynamics are given by

$$\frac{\delta S}{S} = \mu \delta t + \sigma \phi \sqrt{\delta t},$$

where ϕ is an $\mathcal{N}(0,1)$ random variable and δt is a non-infinitesimal time. As in Section 1.2 we begin with a portfolio consisting of long one derivative and short Δ units of the underlying. After a small, but not infinitesimal, amount of time, δt, the change in value of the portfolio will be

$$\delta \Pi(t,S) = \left[V_t + (\mu - D) S V_S + \frac{1}{2}\sigma^2 S^2 V_{SS} \phi^2 - \Delta D S - \Delta(\mu - D) S \right] \delta t + \sigma S \left[V_S - \Delta \right] \phi \sqrt{\delta t} - kS|\nu|,$$

where k is the level of the trader's proportional transaction costs and ν is the number of

assets traded at time t. Comparing this last equation with equation (1.6) we see that the change in value of the hedging portfolio in the presence of transaction costs is exactly that in value in the Black-Scholes world less the amount, $kS|\nu|$, required to trade the ν assets; this cost impacts negatively as transaction costs will always detract from the value of the portfolio. If we now delta hedge (that is, choose $\Delta = V_S$) we find that, unlike in the Black-Scholes world, the risk in the portfolio is not completely eliminated; that is, the market has now been made incomplete by the addition of the transaction cost effect. This inability to hedge away all asset price risk is a result of the fact that transaction costs force us to hedge discretely and so there is uncertainty in the number of assets required to hedge the derivative over δt. The number of assets traded, ν, however, is the change in Δ between two trades and so, assuming δt is small, is approximately

$$\nu \approx \sigma S V_{SS} \phi \sqrt{\delta t},$$

where, for future reference, we note that V_{SS} (also known as the contract's Γ) is the amount of the underlying required to maintain a hedged portfolio. We now use this expression for ν to calculate the expected transaction costs during the trading interval; this is

$$E[kS|\nu|] = \sqrt{\frac{2}{\pi}} k\sigma S^2 \left|V_{SS}\right| \sqrt{\delta t}.$$

Using this expression for the expected transaction costs we now require the expected rate of return of the hedging portfolio to equal the risk-free rate; that is

$$E[\delta \Pi] = r \Pi \delta t.$$

The equation satisfied by a European contract in the presence of proportional transaction costs is then

$$V_t + (r - D) S V_S + \frac{1}{2}\sigma^2 S^2 V_{SS} - \sqrt{\frac{2}{\pi \delta t}} k\sigma S^2 \left|V_{SS}\right| - rV = 0, \tag{1.11}$$

which is precisely that derived by Hoggard, Whalley, and Wilmott in [27] and mentioned above. The key points of this brief analysis are: 1. the presence of transaction costs causes the market to become incomplete and we can no longer hedge perfectly; as a result 2. only

the *expected* return on the portfolio will grow at the risk-free rate. These points will be addressed again in Chapter 2.

Taleb [42] defines liquidity as the ability to trade in an asset not only without cost, but also without significantly affecting its price as a result. There have been a number of studies that have focused on relaxing the price-taker assumption and examined the resulting *market impact* feedback effect both on the price of the asset itself and on the price of a derivative contingent on that asset.

Jarrow [29] uses a discrete-time economy consisting of a large trader and a group of small *price takers* to examine whether market manipulation strategies can exist in a market in which trading impact effects are present. A market manipulation strategy is a strategy that generates a rate of return greater than this risk-free rate; their preclusion is therefore an alternative way of stating that a model is arbitrage-free. In this study the impact effect is modelled explicitly as a function of the large trader's trade size and it is shown, in general, that manipulation strategies are possible for the large trader in this type of market. More specifically, these manipulation strategies will exist if the asset price process is dependent on the history of the large trader's holdings.

In [30] Jarrow extends his model from [29] and analyzes the pricing and hedging of derivative contracts in this framework. He again finds that arbitrage opportunities are, in general, possible in this type of model. In addition to the no-arbitrage condition found in [29], he finds that another condition is required when the large trader can invest in a derivative as well. This additional *synchrony* condition requires the asset price to immediately react to and reflect the holdings of the large trader, both in the underlying asset and in the derivative. When these conditions are satisfied Jarrow finds that manipulation strategies are forbidden and a unique derivative price can be obtained; this price is simply the Black-Scholes price with a random volatility modified by the impact parameter of the model.

In [24], [25], [41], and [39], the feedback effect of a dynamic trading strategy is examined through its effect on the volatility of the asset price process. The method used in all of these papers is similar: The market is split into a large group of *reference* (or *small* in [39]) traders who trade based on the fundamental value of the asset and a smaller group of *programme* (or *large* in [39]) traders who trade based on some other strategy. An aggregate demand

function is then specified for each of the two groups and enforcing market equilibrium then results in a modified asset price process.

In [25], Frey and Stremme focus specifically on the modification to the volatility due to a dynamic trading strategy. When the programme traders follow a delta-hedging strategy for a simple option, the authors are able to derive the modification caused by the programme traders to the volatility assumed by the reference traders. In [24], Frey extends the work in [25] by considering the valuation of options on the volatility-modified price dynamics and finds a trading strategy for a programme trader the super-replicates and therefore gives an upper bound to the price of an option.

In [41], Sircar and Papanicolaou again derive a feedback-modified asset price process, but use this result to focus on the pricing of derivatives in this market. They manage to derive a pricing equation similar to Black-Scholes, but with a nonlinear volatility parameter modified by both the fraction of the total assets traded by the programme traders and their utility function. This pricing equation is then analyzed asymptotically for a linear utility function and when the programme traders are a small fraction of the market and have shown, amongst other things, that feedback effects alone are not enough to account for the observed volatility smile.

In [39], Schonbucher and Wilmott derive a pricing equation for the paper value of a replicating portfolio that has the same form as found in [41] and analyze this equation numerically. They also focus on the possibility of market manipulation strategies and arbitrage opportunities within the model and show in general that these situations do exist; their presence demonstrated by the emergence of jumps in the price process where the contract's Γ is large.

In the next section we will present the BH model developed by Bakstein and Howison [4]. This model has several advantages to the models described above. First, the market impact function used in the model contains both temporary and permanent impact effects and is modelled explicitly as a nonlinear function of the trade size. Second, unlike any of the previous models, the BH model incorporates both a market impact *and* transaction cost effect. Finally, as we will see in Chapter 2, when the assett's bid-ask spread can be considered negligible its market is complete and arbitrage free and therefore cannot admit the market manipulation strategies that existed in some of the models mentioned above.

As it was manipulation strategies that prevented a unique derivative price in these models we should be able to derive unique prices for a generic derivative contract using the BH model unlike in the previous studies.

1.5 The Bakstein-Howison Liquidity Asset Price Model

The central problem with using Black-Scholes in a finitely liquid market is that, as we showed in Section 1.2, the price of a derivative is inherently tied to a dynamic trading strategy in the underlying. If transaction costs are incorporated into the analysis the derivative's price may become unbounded as the trading interval tends to zero; we must therefore assume rehedging takes place at non-infinitesimal intervals. We begin with a discrete-time, finite horizon economy. Over the set of permissible trading times, $\{t_i : i = 0, 1, 2 \dots N\}$ (where $t_0 = 0$ and $t_N = T$), we define two price processes; let S_{t_i} and B_{t_i} be the values of the underlying asset and a risk-free bond, respectively, at time t_i. Let Ω be the finite set of all possible states, ω_j, of the system; in our case a state can be thought of as one realization of the stochastic asset price path.

The bond price dynamics are deterministic and can be expressed as

$$B_{t_0} = 1,$$
$$B_{t_{i+1}} = (1 + r_{(t_i,t_{i+1})})B_{t_i},$$

where $r_{(t_i,t_{i+1})}$ is the spot rate at time t_i used until time t_{i+1}.

To describe the dynamics of the asset price we now (fictitiously) divide the market into two groups. The first of these is a large and influential trader whose trades will impact on the price. This large trader can be thought of, more generally, as an aggregate of influential traders who all have knowledge of each others trading strategies and is exactly the programme trader used in the previous literature. The second group is simply the rest of the traders in the market which we will refer to as the *background* group. Crucially we assume all traders within the background to be price takers and that they have no knowledge of each others trading activity. The net effect of the actions of the background is to cause a stochastic change in the asset price; to model this we use the binomial model of Cox, Ross, and Rubenstein (CRR) [15]. At each time, t_i, the asset price can move to one of two values;

with probability p_u, $S_{t_i} \to u \cdot S_{t_i}$ and with probability $p_d = 1 - p_u$, $S_{t_i} \to d \cdot S_{t_i}$. That is,

$$S_{t_0} = S_0,$$
$$S_{t_{i+1}}(\omega_j) = \begin{cases} u \cdot S_{t_i} & \text{with Prob}(\omega_j = \omega_u) = p_u, \\ d \cdot S_{t_i} & \text{with Prob}(\omega_j = \omega_u) = p_d, \end{cases} \tag{1.12}$$

for some initial asset price, S_0. We note the standard result (see, for example [7]) that the CRR model is complete (and therefore free of arbitrage) when the condition $u \geq 1 + r \geq d$ is satisfied.

For future reference we note that the parameters, u, d, p_u, and p_d of the binomial model can be chosen in such a way that the discrete process converges to a Geometric Brownian Motion (GBM) with drift μ and volatility σ,

$$dS = \mu S dt + \sigma S dX, \tag{1.13}$$

as $\delta t \to dt$ for an infinitesimal interval, dt. Specifically, we require the conditional mean, $\hat{\text{E}}[S_{t_{i+1}}|S_{t_i}]$, and variance, $\text{Var}[S_{t_{i+1}}|S_{t_i}]$,of the binomial process over one timestep to be equal to that of the continuous-time process for the same interval. That is,

$$\hat{\text{E}}[S_{t_{i+1}}|S_{t_i}] = e^{\mu \delta t} S_{t_i}, \qquad \text{(GBM)} \tag{1.14}$$

and

$$\hat{\text{E}}[S_{t_{i+1}}|S_{t_i}] = (p_u u + p_d d) S_{t_i}. \qquad \text{(binomial)} \tag{1.15}$$

Similarly, the variances for each process are

$$\text{Var}[S_{t_{i+1}}|S_{t_i}] = e^{2\mu \delta t}(e^{\sigma^2 \delta t} - 1) S_{t_i}^2, \qquad \text{(GBM)} \tag{1.16}$$

and

$$\text{Var}[S_{t_{i+1}}|S_{t_i}] = (p_u u^2 + p_d d^2) S_{t_i}^2. \qquad \text{(binomial)} \tag{1.17}$$

Equating (1.14) with (1.15) and (1.16) with (1.17) gives two equations for the three unknowns u, d, and p_u (where there are only three unknowns since $p_d = 1 - p_u$). To fully specify the solution we note that $0 \leq p_u \leq 1$ and $0 < d < u$; we use the Jarrow-Rudd [28] condition by choosing $p = 1/2$ which gives

$$\begin{aligned} u &= e^{\mu\delta t}\left(1 + \sqrt{e^{\sigma^2\delta t} - 1}\right), \\ d &= e^{\mu\delta t}\left(1 - \sqrt{e^{\sigma^2\delta t} - 1}\right), \quad \text{and} \\ p &= \frac{1}{2}, \end{aligned} \tag{1.18}$$

and causes the binomial process of (1.12) to converge to the GBM of (1.13) in the limit of continuous time.

To incorporate liquidity effects into the model we now focus on the actions of the large trader. Let H_{t_i} and $\hat{H}_{t_i}$ be the large trader's holdings at time t_i in the asset and bond, respectively. We will assume that both H_{t_i} and $\hat{H}_{t_i}$ are fully determined given S_{t_i}; that is, the amount traded by the large trader during any period is fully predictable given the asset price at the beginning of that period.

The liquidity effects described in Section 1.3 all increase the costs associated with trading in the asset. Let $\overline{S}_{t_i}$ be the average price paid by the large trader when trading an amount, $\delta H_{t_i} = H_{t_{i+1}} - H_{t_i}$, and γ and λ be measures for the bid-ask spread width and the market depth, respectively.

Liquidity effects will first impact the large trader through the average price paid per asset when executing a trade. This price, $\overline{S}_{t_i}$, should be dependent on both the bid-ask spread and the market depth; with a non-zero spread the average transaction price for a buy (sell) will automatically be above (below) the spot price and with a shallow market more than one layer may have to be tapped thus increasing the cost of the trade. Furthermore, we expect:

- $\overline{S}_{t_i}$ should be an increasing function of H_{t_i}.
- For a sell order, $\overline{S}_{t_i} < S_{t_i}$. In the limit of a *very* large order we should have $\lim_{\delta H_{t_i} \to -\infty} \overline{S}_{t_i} = 0$.
- For a buy order, $\overline{S}_{t_i} > S_{t_i}$. In the limit of a *very* large order we should have

$\lim_{\delta H_{t_i} \to \infty} \overline{S}_{t_i} = \infty$.

One possible form for this impact function (and the one that will be used in this work) has been suggested in [4]; it is

$$\overline{S}_{t_i} = S_{t_i} \left[1 + \text{sign}(\delta H_{t_i})\gamma\right] e^{\lambda \delta H_{t_i}}. \tag{1.19}$$

Once the large trader's transaction is complete the market will be briefly out of equilibrium. The re-equilibration of the asset price will occur as the other traders in the market adjust their orders in response to the actions of the large trader. It is reasonable to expect that the new, permanent spot price will depend on both the previous spot, S_{t_i}, and the large trader's average price paid per asset. We will model this asset price slippage by taking the new spot price to be the geometric average of these two quantities. This average will then be weighted by a slippage parameter, α, to incorporate the extent to which the background traders believe the large trader's trade bears any new information. Specifically, this permanent market impact is modelled by

$$S_{t_i} \to S_{t_{i+1}} = S_{t_i}^{\alpha} \overline{S}_{t_i}^{1-\alpha}$$

Two notable cases for the model are: 1. $\alpha = 0$. In this case $S_{t_{i+1}} = \overline{S}_{t_i}$ so the background traders believe the large trader's average transaction price fully reflects the true state of the market; and 2. $\alpha = 1$. In this case they believe the large trader's actions contain no new information so the new spot price remains unchanged at $S_{t_{i+1}} = S_{t_i}$.

Over one timestep the asset price dynamics can be expressed as

$$\overline{S}_{t_i} = S_{t_i} \left[1 + \text{sign}(\delta H_{t_i})\gamma\right] e^{\lambda \delta H_{t_i}} \qquad \text{(transaction cost)} \tag{1.20}$$

$$S_{t_i} \to S_{t_i}^{\alpha} \overline{S}_{t_i}^{1-\alpha} = S_{t_i} \left[1 + \text{sign}(\delta H_{t_i})\gamma\right]^{1-\alpha} e^{\lambda(1-\alpha)\delta H_{t_i}} \qquad \text{(price slippage)} \tag{1.21}$$

$$S_{t_i} \overline{S}_{t_i} \to \begin{cases} u \cdot S_{t_i} \left[1 + \text{sign}(\delta H_{t_i})\gamma\right]^{1-\alpha} e^{\lambda(1-\alpha)\delta H_{t_i}} \\ d \cdot S_{t_i} \left[1 + \text{sign}(\delta H_{t_i})\gamma\right]^{1-\alpha} e^{\lambda(1-\alpha)\delta H_{t_i}} \end{cases} \qquad \text{(stochastic change)} \tag{1.22}$$

To aid in our development of the model, Figure 1.2 shows the first two timesteps of the asset price dynamics. In summary, a timestep begins when a stochastic change finishes from the previous timestep (at t_0 we imagine that the trader simply begins with a mishedged

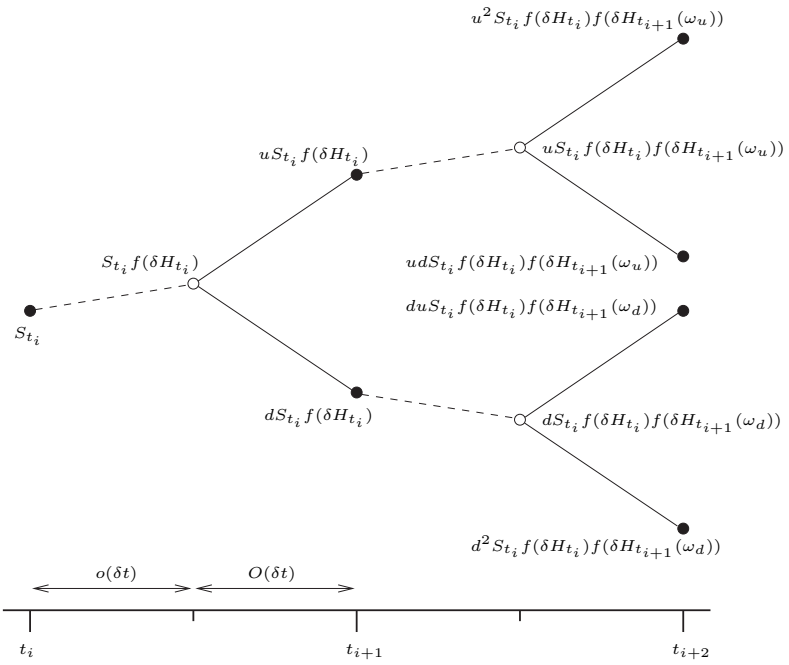


Figure 1.2: Asset price tree representing the price process given by equations (1.20) - (1.22). Explicit values of S have been shown for nodes with $t \leq t_{i+2}$. Dashed lines represent price slippage and solid lines asset price diffusion. The factor $f(\delta H_{t_{i+j}}(\omega)) = [1 + \text{sign}(\delta H_{t_{i+j}}(\omega))\gamma]^{1-\alpha} e^{\lambda(1-\alpha)\delta H_{t_{i+j}}}$ for $j \in \{0, 1, 2\}$.

portfolio). The large trader then adjusts his holdings in the underlying in order to rehedge his portfolio. The rehedge quantity, δH_{t_i}, is calculated to take into account the subsequent price slippage and the average price paid per asset for this trade is given by equation (1.20). After this transaction the asset price will briefly be out of equilibrium due to the fact that the large trader's trade will have removed some orders from the board. In response the background traders may adjust their orders resulting in a permanent price shift given by (1.21). While it is reasonable to believe that a readjustment of all the background traders' holdings may take some time, we will assume that this permanent slippage is an instantaneous effect (or, at least occurs on a much shorter timescale than the stochastic (diffusive) change in the price). Finally, to complete the timestep the asset price then undergoes a stochastic change to $u \cdot S_{t_i}^{\alpha} \overline{S}_{t_i}^{1-\alpha}$ with probability p_u, or to $d \cdot S_{t_i}^{\alpha} \overline{S}_{t_i}^{1-\alpha}$ with probability p_d.

Note Because of the shift in the asset price due to each of the large trader's trades it is not clear whether the asset price tree generated by the Bakstein-Howison model recombines. For derivative pricing purposes when using the full model in a discrete-time setting, recombination is an important issue: After N timesteps and without recombination there would be 2^N possible states to the system; in this situation valuing even simple derivatives on the asset would become a path-dependent and very difficult task. While it is relatively obvious that there is no reason the asset price process should recombine for an arbitrary trading strategy, whether or not it recombines for a delta-hedging strategy for a specific derivative contract has proven a difficult question to answer and remains an open area for research. In Chapter 2 we use the Bakstein-Howison asset price model to derive the equation satisfied by the price of a derivative in a finitely liquid market and find that when the bid-ask spread of the asset can be considered negligible a true continuous-time pricing equation emerges. This $\gamma = 0$ equation will be our sole derivative pricing focus; since recombination is immaterial in the continuous-time limit, this issue will not prove a problem for our work.

1.5.1 Calibration of the Model Parameters

So far γ, λ, and α have only loosely been defined as parameters for, respectively, the bid-ask spread, market depth, and price slippage. In order to calibrate the model with real data we must give them more precise definitions. For this work γ is a direct measure of the bid-ask spread and defined as the fractional increase of $\overline{S}_{t_i}$ over S_{t_i} as $\delta H \to 0$. The market depth parameter, λ, has dimension (number of assets traded)$^{-1}$ and is defined as the relative change in the average transaction price per asset traded. This definition does not take into account the fact that the market depth *felt* by the holder of a very large position in the asset will be less than for someone holding a smaller position. To incorporate this fact we will need to normalize λ with some measure of the market size; this point will be addressed in Chapter 2. Finally, α is the degree to which the average transaction price slips back to this price after a trade (*i.e.* if $\alpha = 0$ the new spot price will exactly equal the average transaction price). These parameters have been explicitly calculated for several large cap equities on the German stock exchange in [4]. For these stocks γ takes values in the range $8 \times 10^{-4} \to 1.2 \times 10^{-3}$, λ in the range $(8 \times 10^{-8} \to 3 \times 10^{-7})$(no. of assets traded)$^{-1}$, and α in the range $0.42 \to 0.51$.

1.6 Outline of the Thesis

As was mentioned at the beginning of the chapter, the first part of our work will revolve around the pricing of derivatives on finitely liquid assets. In Chapter 2 we will use the BH asset price model to derive the Black-Scholes-equivalent derivative pricing equations for a contract on a finitely liquid asset. In Chapter 3 we investigate the asymptotic form of the solution to this pricing equation for a European call option near strike and expiry where liquidity effects are most prominent. In Chapter 4 we then apply the pricing model to a new contract termed the *American forward*. This contract has a very simple structure and we believe may be a very effective, yet inexpensive, hedge against liquidity risks. Finally, in Chapters 5 and 6 we then extend the BH model to allow for a stochastic market depth. We derive the derivative pricing equation for this stochastic liquidity model and then find asymptotic approximations to this equation in several regions of the model's parameter space.

In the second part of the thesis we investigate how to optimally execute a transaction of a large number of assets in a finitely liquid market. Using the BH model to generate the asset price we examine in Chapter 7 the simplest situation in which the optimal trading strategy is that which minimizes the strategies' transaction costs. In Chapters 8 and 9 we then generalise the problem and solve for the trading strategy that simultaneously minimizes the cost of the transaction and its associated risk; in Chapter 8 we do this for the situation where the level of the trader's risk-aversion is very small and in Chapter 9 when this risk-aversion is very large.

Finally, in Chapter 10 we make concluding remarks and suggest ideas for future research.

Chapter 2

Derivation of Derivative Pricing Equations

The goal of this chapter is to derive the equation(s) for the price of a derivative contingent on an asset in a finitely liquid market whose price dynamics are given by the BH model developed in Chapter 1. We will first derive the pricing equation for general γ, λ, and α in Section 2.1. We will see that the equation in this general situation is quite complex and extracting its qualitative features is difficult; in Section 2.2 we will therefore examine the negligible bid-ask spread case which is what the work in Chapters 3, 4, 5, and 6 will expand upon. In Section 2.3 we extend the $\gamma = 0$ pricing framework to account for the early exercise feature of American contracts and then finally conclude the chapter in Section 2.4 by making brief concluding remarks.

We assume a world in which trading occurs instantaneously at a discrete set of times $\{t_0, \ldots, t_N = T\}$ and where the period length, $\delta t = t_i - t_{i-1}$ is a constant for all i. Even though market activity occurs discretely, we assume that the value of the derivative is continuous in both t and S; this can be justified by assuming t and S to be continuous and that the market participants only 'awaken' every δt to trade.

Because of the nonlinear form of the BH model we should expect the equation(s) satisfied by a derivative contract to be likewise nonlinear. With forethought we therefore define n to be the number of assets that the derivative is written on, $V(t, S)$ to be the total value of the contract on these n assets, and $\tilde{V}(t, S)$ to be the value of the contract *per unit of*

underlying. With any linear pricing equation it is clear that the relation $V(t,S) = n\tilde{V}(t,S)$ holds, but this is not necessarily so in the nonlinear case. We also define a dimensionless market depth parameter, $\overline{\lambda} = n\lambda$, in terms of this number. This normalized $\overline{\lambda}$ automatically takes into account the fact that the effective depth of the market will be less when more assets need to be traded.

2.1 Leading-Order Equation for Arbitrary γ, λ, and α

To begin, we construct a portfolio with value $\Pi(t,S;nH)$ consisting of long one derivative written on n units of the underlying and with value $V(t,S)$, and short nH units of the underlying. The value of this risky portfolio is given by

$$\begin{aligned}\Pi(t,S;nH) &= V(t,S;nH) - nHS, \\ &= n\tilde{V}(t,S;nH) - nHS. \end{aligned} \tag{2.1}$$

At the same time we will consider a number, $n\hat{H}_t$, of risk-free bonds each with value B_t at time t such that the value of this cash holding is equivalent in value to the risky portfolio; that is

$$\Pi(t,S;nH) = n\tilde{V}(t,S;nH) - nHS = n\hat{H}_t B_t. \tag{2.2}$$

We begin the derivation in the middle of period t (labelled by any $\circ$ in Figure 1.2) when all trading and slippage has occurred so that nH is the correct hedge quantity. At this point the background agents trade over a time, δt, causing the asset price to move to $S + \delta S$; the value of the risky portfolio at time $t + \delta t$ after this movement is

$$\Pi(t+\delta t, S+\delta S; nH) = n\tilde{V}(t+\delta t, S+\delta S; nH) - nH\cdot(S+\delta S) - nH(S+\delta S)D\delta t$$

where we have explicitly shown the dependence of $\tilde{V}$ on the amount of the underlying held against it and included the effect of the asset paying a constant dividend yield, D. At the end of period t the portfolio is mishedged; after rehedging by an amount $n\delta H$ its value will

be

$$\begin{aligned}\Pi(t,S;nH)+\delta\Pi \equiv&\Pi(t+\delta t,S+\delta S;n(H+\delta H)\\ =&nV(t+\delta t,S+\delta S;n(H+\delta H))-n(1+D\delta t)(H+\delta H)(S+\delta S).\end{aligned}\quad(2.3)$$

But this act will cause a reaction in the market resulting in a price slippage

$$S+\delta S\to(S+\delta S)[1+\gamma\text{sign}(n\delta H)]^{1-\alpha}e^{n\lambda(1-\alpha)\delta H}.\quad(2.4)$$

Now consider just the derivative with value $\tilde{V}(t,S;nH)$ at time t. After diffusion and rehedging its new value will be $\tilde{V}(t+\delta t,S+\delta S;n(H+\delta H))$, but the act of trading the $n\delta H$ assets will modify the asset price so that it slips according to (2.4). The price of the derivative while holding $n(H+\delta H)$ assets against it each worth $S+\delta S$ is equivalent to holding nH assets each worth the slipped value of the asset price; mathematically this is expressed as

$$\tilde{V}(t+\delta t,S+\delta S;n(H+\delta H))=\tilde{V}(t+\delta t,(S+\delta S)[1+\gamma\text{sign}(n\delta H)]^{1-\alpha}e^{n\lambda(1-\alpha)\delta H};nH),\quad(2.5)$$

and equation (2.3) can then be written

$$\begin{aligned}(\Pi(t,S;nH)+\delta\Pi)/n=&\tilde{V}(t+\delta t,(S+\delta S)[1+\gamma\text{sign}(n\delta H)]^{1-\alpha}e^{\bar{\lambda}(1-\alpha)\delta H};nH)\\ &-(1+D\delta t)(H+\delta H)(S+\delta S)[1+\gamma\text{sign}(n\delta H)]^{1-\alpha}e^{\bar{\lambda}(1-\alpha)\delta H},\end{aligned}\quad(2.6)$$

where we have used the definition $n\lambda=\bar{\lambda}$. If we expand (2.6) in a Taylor series around the point (t,S) we get

$$\begin{aligned}(\Pi(t,S;nH)+\delta\Pi)/n=&\tilde{V}+(\tilde{V}_t-HSD)\delta t+\tilde{V}_S\left((S+\delta S)[1+\gamma\text{sign}(n\delta H)]^{1-\alpha}e^{\bar{\lambda}(1-\alpha)\delta H}-S\right)\\ &+\frac{1}{2}\tilde{V}_{SS}\left((S+\delta S)[1+\gamma\text{sign}(n\delta H)]^{1-\alpha}e^{\bar{\lambda}(1-\alpha)\delta H}-S\right)^2\\ &-(H+\delta H)(S+\delta S)[1+\gamma\text{sign}(n\delta H)]^{1-\alpha}e^{\bar{\lambda}(1-\alpha)\delta H}+O(\delta S\cdot\delta t).\end{aligned}\quad(2.7)$$

Choosing $H=\tilde{V}_S$, as in the Black-Scholes analysis, eliminates the leading-order component of the portfolio's risk due to the stochastic change in the asset price. With this choice for

H we have

$$\begin{aligned}\delta H = \delta \tilde{V}_S =& \tilde{V}_{SS}\delta S + \tilde{V}_{tS}\delta t \\ \approx& \tilde{V}_{SS}\delta S,\end{aligned} \tag{2.8}$$

for small δt. Since $\delta H = O(\delta S)$ it is small and we can expand the exponential

$$e^{\bar{\lambda}(1-\alpha)\delta H} = 1 + \bar{\lambda}(1-\alpha)\delta H + \frac{1}{2}\bar{\lambda}^2(1-\alpha)^2\delta H^2 + \cdots . \tag{2.9}$$

For ease of notation let

$$\begin{aligned}\theta &\equiv 1 + \gamma \text{sign}(n\delta H), \\ &= 1 + \gamma \text{sign}(\delta H),\end{aligned} \tag{2.10}$$

where the second equality follows from the fact that $n \geq 0$ and therefore does not affect the sign of δH. Once equations (2.8) and (2.9) are substituted into (2.7) we are left with

$$\begin{aligned}(\Pi + \delta\Pi)/n = \tilde{V} + (\tilde{V}_t - nDS\tilde{V}_S) - S\tilde{V}_S\delta t \\ + \left[\left(\theta^{2(1-\alpha)} - \theta^{1-\alpha}\right)\left(\bar{\lambda}(1-\alpha)S\tilde{V}_{SS} + 1\right)\tilde{V}_{SS} - \theta^{1-\alpha}S\tilde{V}_{SS}\right]\delta S \\ + \left[\frac{1}{2}\left(2\theta^{2(1-\alpha)} - \theta^{1-\alpha}\right)\left(\bar{\lambda}(1-\alpha)S\tilde{V}_{SS} + 2\right)\bar{\lambda}(1-\alpha)S\tilde{V}_{SS}^2\right. \\ \left. -\theta^{1-\alpha}\left(\bar{\lambda}(1-\alpha)S\tilde{V}_{SS} + 1\right)\tilde{V}_{SS} + \frac{1}{2}\theta^{2(1-\alpha)}\tilde{V}_{SS}\right]\delta S^2 + O(\delta S^3).\end{aligned} \tag{2.11}$$

We now focus on the dynamics of the risk-free portfolio defined in equation (2.1). Beginning again in the middle of period t, the value of the portfolio is $\Pi(t, S; H) = n\hat{H}_t B_t$. After diffusion this value will grow deterministically so that

$$\Pi(t, S; nH) \to n\hat{H}_t B_t(1 + r\delta t).$$

At the end of period t we then rehedge the portfolio so that $n\hat{H}_t \to n\hat{H}_t + n\delta\hat{H}_t$; the final value of this equivalent rehedged risk-free portfolio will therefore be

$$\Pi(t, S; H) + \delta\Pi(t, S; H) = n(\hat{H}_t + \delta\hat{H}_t)B_t(1 + r\delta t). \tag{2.12}$$

We require that the risky and risk-free portfolios together be self-financing; that is, any buying (selling) of assets is completely funded through the selling (buying) of risk-free bonds. Since our model includes liquidity cost effects the appropriate self-financing condition is

$$\delta H \cdot S[1 + \text{sign}(\delta H)\gamma]e^{\lambda \delta H} + \delta \hat{H}_t \cdot B_t(1 + r\delta t) = 0, \tag{2.13}$$

where the first term is the total cost to purchase δH assets when the spot price is S. Substituting equations (2.13) and (2.1) into equation (2.12) gives

$$(\Pi + \delta\Pi)/n = (\tilde{V} - HS)(1 + r\delta t) - \delta H(S + \delta S)[1 + \text{sign}(n\delta H)\gamma]e^{\bar{\lambda}\delta H} \tag{2.14}$$

which says that, under the self-financing constraint, the cash-equivalent portfolio grows deterministically at the risk-free rate less the amount required to rehedge the risky portfolio. Again choosing the hedge quantity $H = \tilde{V}_S$, expanding the exponential term, and then substituting equations (2.8) and (2.9) into (2.14) gives

$$(\Pi(t, S; H) + \delta\Pi)/n = \tilde{V} - S\tilde{V}_S - rS\tilde{V}_S\delta t + r\tilde{V}\delta t - \theta S\tilde{V}_{SS}\delta S + \theta(\bar{\lambda}S\tilde{V}_{SS} + 1)\tilde{V}_{SS}\delta S^2 + O(\delta S^3). \tag{2.15}$$

Equations (2.11) and (2.15) both describe the value of the portfolio after diffusion, rehedging, and slippage; since we have required that the system be self-financing they should therefore be equivalent to $O(\delta S^3)$. Equating (2.11) and (2.15) and dropping the time indeces for notational simplicity gives

$$\begin{aligned}
&\left[\tilde{V}_t + (r - D)S\tilde{V}_S - r\tilde{V}\right]\delta t + \frac{1}{2}\left(\theta^{1-\alpha} - 1\right)^2 S^2\tilde{V}_{SS} \\
&\quad + \left[\left(\theta^{2(1-\alpha)} - \theta^{1-\alpha}\right)(\bar{\lambda}(1-\alpha)S\tilde{V}_{SS} + 1)\tilde{V}_{SS} + \left[\theta - \theta^{1-\alpha}\right]S\tilde{V}_{SS}\right]\delta S \\
&\quad + \left[\frac{1}{2}\left(2\theta^{2(1-\alpha)} - \theta^{1-\alpha}\right)(\bar{\lambda}(1-\alpha)S\tilde{V}_{SS} + 2)\bar{\lambda}(1-\alpha)S\tilde{V}_{SS}^2\right. \\
&\quad \left. \theta^{1-\alpha}(\bar{\lambda}(1-\alpha)S\tilde{V}_{SS} + 1)\tilde{V}_{SS} + \theta(\bar{\lambda}S\tilde{V}_{SS} + 1)\tilde{V}_{SS} + \frac{1}{2}\theta^{2(1-\alpha)}\tilde{V}_{SS}\right]\delta S^2 = 0.
\end{aligned} \tag{2.16}$$

Equation (2.16) is the equation describing the leading-order dynamics of the derivative's value and is valid for all $t \in \{t_0, \ldots, T\}$. For an arbitrary bid-ask spread and market depth the leading-order component of a derivative's price in a finitely liquid market is clearly

stochastic. The presence of the (stochastic) δS and θ terms in (2.16) is a result of the fact that the present market model is incomplete and results in an analogous situation to that encountered in Section 1.4 with the pure transaction cost model. The residual risk remaining in the hedging portfolio results from the fact that we have had to hedge discretely and thus from the uncertainty in the rehedging costs between trading intervals. As we showed in Section 1.4 with the pure transaction cost model, to generate a deterministic price for derivative we must take the expectation of (2.16) thereby requiring that the expected rate of return on the hedging portfolio be equal to the risk-free rate.

The current liquidity model, however, is more complex than the pure transaction cost model. Once the expectation of (2.16) is taken we will have terms of $O(\sqrt{\delta t})$ and $O(\delta t)$, but multiplying these terms will be several different powers of γ. Depending on the relative magnitudes of γ and δt the form of the leading-order pricing equation will change. In addition to the $\gamma = O(\sqrt{\delta t})$ pure transaction cost case already examined, we believe a distinct pricing equation exists for the case $\gamma = O(\delta t)$ and possibly for other cases as well, but leave this as an open area for future research.

2.2 The Case $\gamma \ll \delta t$

A very interesting situation occurs when $\gamma = o(\delta t)$. In this case we are left with a pure liquidity model in the sense that the bid-ask spread can be neglected and any trading effects are caused entirely by price slippage. When $\gamma = o(\delta t)$, the leading-order form of the pricing equation (2.16) reduces to

$$\left[\tilde{V}_t + (r-D)S\tilde{V}_S - r\tilde{V}\right]\delta t + \left[\frac{1}{2}\tilde{V}_{SS} + \bar{\lambda}S\tilde{V}_{SS}^2 + \frac{1}{2}\bar{\lambda}^2(1-\alpha)^2S^2\tilde{V}_{SS}^3\right]\delta S^2 = 0. \qquad (2.17)$$

There are no δS terms in equation (2.17); in the continuous-time limit $\delta S^2 \sim \sigma^2 S^2 \delta t$ as $\delta t \to 0$ and (2.17) reduces to

$$\tilde{V}_t + (r-D)S\tilde{V}_S + \frac{1}{2}\sigma^2S^2\tilde{V}_{SS} - r\tilde{V} + \bar{\lambda}\sigma^2S^3\tilde{V}_{SS}^2 + \frac{1}{2}\bar{\lambda}^2(1-\alpha)^2\sigma^2S^4\tilde{V}_{SS}^3 = 0, \qquad (2.18)$$

which, henceforth, we will refer to as the $\gamma = 0$ pricing equation and is valid for all $\delta t \ll 1$ in the region $\bar{\lambda} \ll 1$ and $\gamma = o(\delta t)$.

Comments

- In the $\gamma = 0$ model a delta hedging strategy perfectly hedges the derivative contract and $\tilde{V}$ is deterministic as a result. It is clear, therefore, that model incompleteness in a finite liquidity setting is entirely the result of the bid-ask spread, or transaction cost, effect. A direct consequence of this model completeness is the fact that the $\gamma = 0$ pricing equation is valid in the limit of continuous time; if the bid-ask spread for the underlying is negligible (as it is in the case for some very large traders) the derivative can be rehedged at arbitrarily small intervals without fear of its value becoming unbounded.

- Like the general pricing equation (2.16), the $\gamma = 0$ equation is nonlinear in $\tilde{V}$. A contract written on a finitely liquid underlying with a negligible bid-ask spread displays non-trivial dependence the number, n, of that underlying it is written upon.

- Defining
$$\hat{\sigma}^2 = \sigma^2\left(1 + 2\bar{\lambda}S\tilde{V}_{SS} + \bar{\lambda}^2(1-\alpha)^2S^2\tilde{V}_{SS}^2\right),$$
we can rewrite the $\gamma = 0$ equation as
$$\tilde{V}_t + rS\tilde{V}_S + \frac{1}{2}\hat{\sigma}^2S^2\tilde{V}_{SS} - r\tilde{V} = 0.$$
In the $\gamma = 0$ model the price of a derivative is simply the Black-Scholes value with the liquidity-modified volatility, $\hat{\sigma}$.

  The important fact about $\hat{\sigma}$, though, is that it is the volatility felt specifically by the large trader who trades according to the delta-hedging strategy. Any price taker in the market could not use this volatility to price a derivative. To see why this is so, we note that within the expansion for $\hat{\sigma}$ there is a term that contains $\bar{\lambda}(1-\alpha)$ and another that contains only $\bar{\lambda}$. Recalling that α incorporates the permanent market impact effect of the large trader's trading activity, it is therefore an *observable* (by all market participants) impact effect. On its own, however, $\bar{\lambda}$ is only felt by the large trader through his transaction costs and cannot be observed by the rest of the market.

  What we have, therefore, is the situation where there are two effective volatilities for the underlying. One of these is the liquidity-modified volatility, $\hat{\sigma}$, which is that

felt by the large trader while carrying out the delta-hedging strategy and the other is simply σ which is 'felt' by all market participants with no knowledge of the large trader's activities.

There is a well-established bank of literature on pure transaction cost derivative pricing models, but very little, if any, on explicit, pure liquidity models such as the $\gamma = 0$ model. For this reason, the remainder of our work on derivative pricing in a finitely liquid market in Chapters 3-6 will focus on the $\gamma = 0$ model alone.

Before we can proceed, however, one final issue with (2.18) needs to be resolved. Regardless of whether we are long or short the derivative (and are therefore long or short Γ, respectively, for a European vanilla contract), the $\tilde{V}_{SS}^2$ term will be strictly positive. This transaction cost-like effect will increase hedging costs to both positions which will result in a bid-ask spread in the derivative's price. This is an interesting result: Even though the bid-ask spread in the underlying has been neglected, one has been generated in the price of the derivative by the $\gamma = 0$ model.

In a finitely liquid market we expect the Black-Scholes price of a derivative, $\tilde{V}_{BS}$, to lie between the ask price (*i.e.* the price demanded by the writer), $\tilde{V}_a$, and the bid price (the price the buyer expects to pay), $\tilde{V}_b$, so that $\tilde{V}_a > \tilde{V}_{BS} > \tilde{V}_b$.[1] Equation (2.18) gives the ask price of a derivative; to see why we must analyze the effects of its nonlinear terms. Since $\bar{\lambda} \geq 0$, as was stated above, the $\tilde{V}_{SS}^2$ term must be strictly positive. The $\tilde{V}_{SS}^3$ term, however, can be positive or negative depending on the form of the contract. In the simple case of a call option, $\tilde{V}_{SS} > 0$; since $\tilde{V}_{SS}^2 > 0$ and both terms in (2.18) have positive sign then $\tilde{V}$ as given by this equation must be larger than $\tilde{V}_{BS}$ and so it must be the ask price. To find the bid price we note that that the holder of the contract receives the negative of the writer's payoff, $\tilde{V}(T, S)$; with the transformation $\tilde{V} \to -\tilde{V}$ we are then left with the equation for the equivalent bid price. The valuation equations are then

$$\tilde{V}_{a_t} + (r - D)S\tilde{V}_{a_S} + \frac{1}{2}\sigma^2 S^2 \tilde{V}_{a_{SS}} - r\tilde{V}_a + \bar{\lambda}\sigma^2 S^3 \tilde{V}_{a_{SS}}^2 + \frac{1}{2}\bar{\lambda}^2(1-\alpha)^2\sigma^2 S^4 \tilde{V}_{a_{SS}}^3 = 0, \quad (2.19)$$

[1] In this work we will always refer to the *price* of a derivative contract; the bid price being the price the buyer feels he should pay for the contract, while the ask price is the price the seller feels the buyer should pay. We specifically distinguish the derivative's price from its *value* to each side of the contract once entered into; for the long position these two quantities will be the same, but for the short position the price and value will be the negative of one another.

for the ask price, and

$$\tilde{V}_{b_t} + (r-D)S\tilde{V}_{b_S} + \frac{1}{2}\sigma^2 S^2 \tilde{V}_{b_{SS}} - r\tilde{V}_l - \bar{\lambda}\sigma^2 S^3 \tilde{V}^2_{b_{SS}} + \frac{1}{2}\bar{\lambda}^2(1-\alpha)^2\sigma^2 S^4 \tilde{V}^3_{b_{SS}} = 0 \quad (2.20)$$

for the bid price given the same final condition, $\tilde{V}(T,S)$.

2.3 American Contracts in the Bakstein-Howison Model

For use in Chapters 4 and 6 we need to extend the $\gamma = 0$ pricing model to account for the early exercise feature of American contracts. Let

$$\begin{aligned} L_{a_{BH}}[\tilde{V}_a(t,S)] = \tilde{V}_{a_t} + (r-D)S\tilde{V}_{a_S} + \frac{1}{2}\sigma^2 S^2 \tilde{V}_{a_{SS}} - r\tilde{V}_a \\ + \bar{\lambda}\sigma^2 S^3 \tilde{V}^2_{a_{SS}} + \frac{1}{2}\bar{\lambda}^2(1-\alpha)^2\sigma^2 S^4 \tilde{V}^3_{a_{SS}}, \quad (2.21) \\ L_{b_{BH}}[\tilde{V}_b(t,S)] = \tilde{V}_{b_t} + (r-D)S\tilde{V}_{b_S} + \frac{1}{2}\sigma^2 S^2 \tilde{V}_{b_{SS}} - r\tilde{V}_b \\ - \bar{\lambda}\sigma^2 S^3 \tilde{V}^2_{b_{SS}} + \frac{1}{2}\bar{\lambda}^2(1-\alpha)^2\sigma^2 S^4 \tilde{V}^3_{b_{SS}}, \quad (2.22) \end{aligned}$$

and, as was defined in subsection 1.2.1, let the hold region, Z, be defined by $Z \in \{S : \tilde{V}(t,S) > \Phi(S)\}$ and the exercise region, Z^*, by $Z^* \in \{S : \tilde{V}(t,S) = \Phi(S)\}$, where $\Phi(S)$ is the contract's payoff function. Note that the physical interpretation of $L_{a_{BH}}$ and $L_{b_{BH}}$ is completely analogous to L_{BS}; they are the rate of return above the risk-free rate that the seller and buyer, respectively, think the hedged portfolio should earn inclusive of the hedging costs arising from the liquidity effects. As a result we must have $L_{a_{BH}}[\tilde{V}_a(t,S)] = 0$ and $L_{b_{BH}}[V_b(t,S)] = 0$ when the contract should be held and $L_{a_{BH}}[\tilde{V}_a(t,S)] < 0$ and $L_{b_{BH}}[V_b(t,S)] < 0$ when it should be exercised (remember V_a and V_b represent the derivative's *price* giving $L_{a_{BH}}[\tilde{V}_a(t,S)] < 0$ in the exercise region, not $L_{a_{BH}}[\tilde{V}_a(t,S)] > 0$). To avoid arbitrage at a free boundary we must impose the usual free boundary conditions that the derivative's price and its Δ must be continuous across the boundary. Allowing for the fact that the free boundary may be different for the bid and ask positions, we define $S^i_{bf}(t)$ and $S^i_{af}(t)$ to be the i^{th} $(i = 1, \ldots, m)$ free boundary for the bid and ask prices, respectively.

The full free boundary problem for a generic American contract in the $\gamma = 0$ model is

$$\begin{aligned} L_{a_{BH}}[\tilde{V}_a(t,S)] &= 0 && S \in Z, \\ L_{a_{BH}}[\tilde{V}_a(t,S)] &< 0 && S \in Z^*, \\ \left.\begin{aligned} \tilde{V}_a(t,S^i_{af}(t)) &= \Phi(S^i_{af}(t)), \\ \tilde{V}_{a_S}(t,S^i_{af}(t)) &= \Phi_S(S^i_{af}(t)), \end{aligned}\right\} && i = 1,\dots m, \end{aligned} \tag{2.23}$$

for the ask price, and

$$\begin{aligned} L_{b_{BH}}[\tilde{V}_b(t,S)] &= 0 && S \in Z, \\ L_{b_{BH}}[\tilde{V}_b(t,S)] &< 0 && S \in Z^*, \\ \left.\begin{aligned} \tilde{V}_b(t,S^i_{bf}(t)) &= \Phi(S^i_{bf}(t)), \\ \tilde{V}_{b_S}(t,S^i_{bf}(t)) &= \Phi_S(S^i_{bf}(t)), \end{aligned}\right\} && i = 1,\dots m, \end{aligned} \tag{2.24}$$

for the bid price.

2.4 Conclusions

In this chapter we have derived the price of a derivative contract contingent on a finitely liquid asset whose dynamics are given by the BH model. For arbitrary magnitudes of the bid-ask spread and market depth we have found the equation for the leading-order component of the price to be stochastic. When the bid-ask spread can be considered negligible, however, it has been shown that delta hedging eliminates all risk from the hedging portfolio and a true continuous-time pricing model results. We can therefore conclude that the bid-ask spread in the full BH model and, more precisely, the fact that this spread forces us to hedge discretely, is what specifically causes the incompleteness of the full model. Interestingly, even though the spread in the underlying is neglected in the $\gamma = 0$ model, we have also found that a spread is generated in the price of the derivative due to the presence of a transaction cost-like Γ^2 term in the pricing equation.

The nonlinear terms in the $\gamma = 0$ pricing equation are driven by Γ^2 and Γ^3 and so we expect the liquidity effects to be most significant where the contract's Γ is largest; near any discontinuities in the payoff or its slope and close to expiry. In the next chapter we will

analyze the $\gamma = 0$ pricing equation for a simple European call option in exactly this region to determine precisely how these nonlinear effects enter the pricing framework.

Chapter 3

Asymptotic Analysis of the European Call Option in the $\gamma = 0$ Model

3.1 Introduction

Except for very simple contract structures, an exact analytical solution to the $\gamma = 0$ pricing equation (2.18) is virtually impossible due to its strong nonlinearities. While numerical solutions can be calculated, these calculations can be computationally expensive and sometimes too time-consuming to be of practical use. More of a potential problem, though, is that calculating a numerical solution to (2.18) can simply be very difficult to carry out at any points near expiry where the payoff function or its slope is discontinuous as V_{SS}^2 and V_{SS}^3 are not well defined at these points. Instead, it would be useful to have a simple analytical approximation to the solution that could be calculated quickly while also providing a reasonable level of accuracy.

If a small parameter exists in a pricing equation then this fact can be exploited using asymptotic analysis to achieve the desired analytic approximation. The use of asymptotic analysis in finite liquidity models is sparse and mainly oriented towards transaction cost models. Research in this area centers around the work by Whalley and Wilmott [44] (and in more depth in [43]) and Barles and Soner [5] in which the authors examine a utility-

based pricing equation (for instance, that derived by Davis, Panas and Zariphopoulou [16] or Hodges and Neuberger [26]) in the limit of a small parameter. In the case of Whalley and Wilmott [44] a small level of transaction costs is assumed, while Barles and Soner [5] investigate the system's behaviour when the size of the option portfolio and the trader's risk-aversion are large. In these limits the authors are able to derive pricing equations that greatly simplify the general three-dimensional nonlinear free-boundary problem for the derivative's price as well as correspondingly simple optimal trading strategies that super-replicate the contract.

Asymptotic methods of approximating pricing models involving market impact are far less prevalent. In [41] Sircar and Papanicolaou analyze a nonlinear pricing equation where the nonlinearities arise from feedback effects due to trading in the underlying. In the limit of a small fraction of programme traders in the market they find the leading-order correction to the Black-Scholes price is driven by a term proportional to Γ^2 and thus analyze a transaction cost-like model very similar to that of Whalley and Wilmott [44]. To our knowledge the only work on a problem involving pure liquidity effects (*i.e.* permanent market impact effects) is a brief mention in [4] about using a regular expansion to approximate a derivative's value in the $\gamma = 0$ BH model. In addition to any explicit calculations, what is neglected in [4] is the presence of a boundary layer that exists for times very close to expiry and cannot be determined using a regular expansion only.

In this chapter we will use the fact that the market depth, $\bar{\lambda}$, is small and find approximate solutions to the $\gamma = 0$ BH pricing equation for a European call option on a non-dividend paying asset. In addition to solving for the solution consistent with the regular expansion suggested in [4], we will also demonstrate the presence of a boundary layer for times very near to expiry and close to strike and then formally match the two regions thereby calculating a solution globally valid for small $\bar{\lambda}$. We will carry out the analysis specifically for the ask price of the contract, but the method used is easily extended to the bid price with the usual transformation $V \to -V$ and $\Phi \to -\Phi$ where Φ is the contract's payoff for the short position.

3.2 The Outer Solution

The valuation problem for the European call option is

$$\begin{gathered}\tilde{V}_t + rS\tilde{V}_S + \frac{1}{2}\sigma^2 S^2 \tilde{V}_{SS} - rV + \bar{\lambda}\sigma^2 S^3 \tilde{V}_{SS}^2 + \frac{1}{2}\bar{\lambda}^2(1-\alpha)^2\sigma^2 S^4 \tilde{V}_{SS}^3 = 0, \\ \tilde{V}(T,S) = \max(S-K,0), \\ \tilde{V}(t,0) = 0, \qquad \text{and} \qquad \tilde{V}(t,S) \sim S \quad \text{as} \quad S \to +\infty.\end{gathered} \tag{3.1}$$

As usual T, K, r, and σ are the expiry date, strike price, risk-free rate of interest, and the underlying's volatility.

We begin by non-dimensionalizing (3.1) using $x = (S-K)/K$, $\tau = \frac{1}{2}\sigma^2(T-t)$, and $\hat{V} = \tilde{V}/K$. With the redefinition $k = 2r/\sigma^2$, the problem becomes

$$\begin{gathered}\hat{V}_\tau = k(1+x)\hat{V}_x + (1+x)^2\hat{V}_{xx} - k\hat{V} + 2\bar{\lambda}(1+x)^3\hat{V}_{xx}^2 + \bar{\lambda}^2(1-\alpha)^2(1+x)^4\hat{V}_{xx}^3, \\ \hat{V}(0,x) = \max(x,0), \\ \hat{V}(\tau,-1) = 0, \qquad \text{and} \qquad \hat{V}(\tau,x) \sim x \quad \text{as} \quad x \to +\infty.\end{gathered} \tag{3.2}$$

To find an approximation to $\hat{V}$ in the outer region we pose the regular expansion

$$\hat{V} \sim V_0^o + \bar{\lambda}V_1^o + \cdots \qquad \text{as} \quad \bar{\lambda} \to 0, \tag{3.3}$$

which results in the $O(1)$ problem

$$\begin{gathered}V_{0_\tau}^o = k(1+x)V_{0_x}^o + (1+x)^2 V_{0_{xx}}^o - kV_0^o, \\ V_0^o(\tau,-1) = 0, \qquad \text{and} \qquad V_0^o(\tau,x) \sim x \quad \text{as} \quad x \to +\infty.\end{gathered} \tag{3.4}$$

Problem (3.4) tells us that the leading-order behaviour of the solution in the outer region is just the Black-Scholes value of the option. While (3.4) can be solved exactly, the resulting form of V_0^o will make the $O(\bar{\lambda})$ problem for V_1^o intractable. Instead, we will rescale the $O(1)$ problem and find an approximation to V_0^o that is valid in some restricted domain of the outer region. Let δ be an artificial small parameter and X and $\bar{\tau}$ be $O(1)$ variables. If we rescale (3.4) according to

$$\tau = \delta\bar{\tau} \qquad \text{and} \qquad x = \sqrt{\delta}X, \tag{3.5}$$

and pose the inner expansion

$$V_0^o \sim \sqrt{\delta} V_{00}^o + \delta V_{01}^o + \cdots , \tag{3.6}$$

then V_{00}^o obeys

$$\begin{gathered} V_{00_{\bar{\tau}}}^o = V_{00_{XX}}^o, \\ V_{00}^o(\bar{\tau}, X) \to 0 \quad \text{as} \quad X \to -\infty \qquad \text{and} \qquad V_{00}^o(\bar{\tau}, X) \sim X \quad \text{as} \quad X \to +\infty, \end{gathered} \tag{3.7}$$

and is a valid approximation to $\hat{V}$ in the outer region for $\tau \ll 1$ and $|x| \ll \sqrt{\tau}$.

To solve (3.7) we seek a similarity solution of the form $V_{00}^o(\bar{\tau}, X) = \sqrt{\bar{\tau}}\Psi_0(\zeta)$ where $\zeta = \frac{X}{\sqrt{\bar{\tau}}}$; with this choice Ψ_0 satisfies

$$\frac{d^2\Psi_0}{d\zeta^2} + \frac{1}{2}\zeta\frac{d\Psi_0}{d\zeta} - \frac{1}{2}\Psi_0 = 0, \tag{3.8}$$

$$\Psi_0 \to 0 \quad \text{as} \quad \zeta \to -\infty, \qquad \text{and} \qquad \Psi_0 \sim \zeta \quad \text{as} \quad \zeta \to +\infty. \tag{3.9}$$

Equation (3.8) along with the boundary conditions (3.9) has the solution

$$\Psi_0(\zeta) = \frac{\zeta}{2}\left[\operatorname{erf}\left(\frac{\zeta}{2}\right) + 1\right] + \frac{1}{\sqrt{\pi}} e^{-\frac{\zeta^2}{4}}, \tag{3.10}$$

where $\operatorname{erf}(\cdot)$ is the error function and is defined as

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-y^2} dy.$$

In the (τ, x) space the leading-order behaviour of the outer solution is thus

$$V_0^o \sim \frac{x}{2}\left[\operatorname{erf}\left(\frac{x}{2\sqrt{\tau}}\right) + 1\right] + \sqrt{\frac{\tau}{\pi}} e^{-\frac{x^2}{4\tau}} + O(\delta), \tag{3.11}$$

which, again, is valid in the region $\tau \ll 1$ and $|x| \ll \sqrt{\tau}$.

We will neglect the higher-order terms, V_{01}^o, V_{02}^o, ..., and focus on the $O(\bar{\lambda})$ correction

to V_0^o. The problem satisfied by V_1^o is

$$\begin{gathered} V_{1_\tau}^o = k(1+x)V_{1_x}^o + (1+x)^2 V_{1_{xx}}^o - kV_1^o + \frac{1}{2\pi\tau}(1+x)^3 e^{-\frac{x^2}{2\tau}}, \\ V_1^o(\tau,-1) = 0, \qquad \text{and} \qquad V_1^o(\tau,x) \to 0 \quad \text{as} \quad x \to +\infty. \end{gathered} \tag{3.12}$$

The problem 3.12 is very similar in form to that analyzed in [41]; whereas in their work, Sircar and Papanicolaou derive an exact solution to this equation in integral form, we will instead focus on a restricted region of the (τ, x) space and derive a simple, closed-form solution in this region. If we rescale (3.12) with (3.5) and pose the expansion

$$V_1^o \sim V_{10}^o + \sqrt{\delta} V_{11}^o + \cdots, \tag{3.13}$$

then the leading-order behaviour of the $O(\bar{\lambda})$ correction is given by

$$\begin{gathered} V_{10_{\bar\tau}}^o = V_{10_{XX}}^o + \frac{1}{2\pi\bar\tau} e^{-\frac{X^2}{2\bar\tau}}, \\ V_{10}^o(\bar\tau, X) \to 0 \quad \text{as} \quad X \to \pm\infty. \end{gathered} \tag{3.14}$$

In deriving (3.14) we have neglected all terms of $O(\delta)$ and higher in our approximation of V_0^o. Since $V_{10}^o = O(\bar{\lambda})$ then this approximation will only be valid for $\delta \ll \bar{\lambda}$ (*i.e.* only where the $O(\bar{\lambda})$ correction term is much larger than the largest of the neglected terms in the expansion for V_0^o) and therefore, since $\tau = O(\delta)$, where $\tau \ll \bar{\lambda}$.

Problem (3.14) has an exact solution. Under the similarity transformation

$$V_{10}^o(\bar\tau, X) = \Psi_1(\zeta), \qquad \text{and} \qquad \zeta = \frac{X}{\sqrt{\bar\tau}}, \tag{3.15}$$

Ψ_1 obeys the equation

$$\frac{d^2\Psi_1}{d\zeta^2} + \frac{1}{2}\zeta\frac{d\Psi_1}{d\zeta} + \frac{1}{2\pi} e^{-\frac{\zeta^2}{2}} = 0, \tag{3.16}$$

with boundary conditions

$$\Psi_1 \to 0 \quad \text{as} \quad \zeta \to \pm\infty. \tag{3.17}$$

Integrating equation (3.16) twice, imposing the boundary conditions (3.17), and changing

back to the (τ, x) variables gives the result

$$V_1^o(\tau, x) \sim \frac{1}{4}\left[1 - \text{erf}\left(\frac{x}{2\sqrt{\tau}}\right)^2\right] + O(\sqrt{\delta}). \tag{3.18}$$

From arguments above we know the upper bound on the region of validity of approximation (3.18) is $\tau \ll \bar{\lambda}$, $x \ll \sqrt{\tau}$. There will also be a lower bound to this region which will occur when the outer expansion, (3.3), breaks down; that is when

$$V_1^o \sim \frac{1}{\bar{\lambda}} V_0^o. \tag{3.19}$$

To determine the lower bound we first note, from equation (3.11), that

$$V_0^o \sim \sqrt{\tau}, \tag{3.20}$$

and, after differentiating (3.11) twice *w.r.t.* x, that

$$V_{0_{xx}}^o \sim \frac{1}{\sqrt{\tau}} \tag{3.21}$$

as $\tau \to 0$ for $x \ll \sqrt{\tau}$. Secondly, if we rewrite equation (3.12) as

$$L_{BS}\left[V_1^o\right] = 2(1+x)^3 V_{0_{xx}}^{o^2},$$

and invoke the small-τ behaviour of $V_{0_{xx}}^o$ from equation (3.21), noting that the Black-Scholes operator, $L_{BS}[\cdot]$, has units of 1/time, then

$$V_1^o \sim 1 \tag{3.22}$$

as $\tau \to 0$ for $x \ll \sqrt{\tau}$. Substituting equations (3.20) and (3.22) into (3.19) tells us that the outer approximation will break down for

$$\tau \sim \bar{\lambda}^2. \tag{3.23}$$

The full bounds on the region of validity of our outer approximation are therefore $\bar{\lambda}^2 \ll \tau \ll \bar{\lambda}$ and $x \ll \sqrt{\tau}$.

Plots of the approximation (3.18) are compared in Figure 3.2 to the numerically calculated first correction, V_1^o, from (3.12). For a value of $\bar{\lambda} = 10^{-4}$ equation (3.18) should be valid for $10^{-8} \ll \tau \ll 10^{-4}$. At $\tau = 5 \times 10^{-6}$ we see the approximation is very good with both the magnitude and the symmetry about $S = K$ of the liquidity effects being reproduced well. Outside the region of validity, at $\tau = 5 \times 10^{-2}$, the symmetry and magnitude of the numerically computed correction break around $S = K$ as τ increases; neither of these features are reproduced in the approximation. This divergence is a result of the non-homogeneous term, V_{0xx}^2, being bounded under the similarity transformation (3.15) and symmetric which would not have been the case if higher-order terms had been retained in the expansion for V_0^o.

3.3 Inner Solution

We have shown that the terms of the outer expansion become of similar size when $\tau \sim \bar{\lambda}^2$; from (3.5) we know $x = O(\sqrt{\tau})$ which suggests (3.2) has a boundary layer of size $\bar{\lambda}^2$ and $\bar{\lambda}$ in the τ and x dimensions, respectively. As suggested by the size of the boundary layer, let θ' and y' be variables (whose sizes will be determined below) defined by $\tau = \bar{\lambda}^2\theta'$ and $x = \bar{\lambda}y'$; if we transform problem (3.2) into the (θ', y') variables and pose the inner expansion

$$\hat{V} \sim \bar{\lambda}V_0^i + \bar{\lambda}^2 V_1^i + \cdots \qquad \text{as} \quad \bar{\lambda} \to 0, \tag{3.24}$$

then the leading-order behaviour of the solution in the boundary layer satisfies

$$\begin{gathered} V_{0_{\theta'}}^i = V_{0y'y'}^i + 2V_{0y'y'}^{i^2} + (1-\alpha)^2 V_{0y'y'}^{i^3}, \\ V_0^i(0, y') = \max(y', 0), \\ V_0^i(\theta', y') \to 0 \quad \text{as} \quad y' \to -\infty, \qquad \text{and} \qquad V_0^i(\theta', y') \sim y' \quad \text{as} \quad y' \to +\infty. \end{gathered} \tag{3.25}$$

In its present form (3.25) is intractable, but we can obtain approximations to the solution in both the small- and large-θ' limits (which will incorporate the initial condition and the conditions to match with the outer solution, respectively). We begin by seeking an approximation to the solution in the small-θ' limit. Let ϵ be an artificial small parameter and let θ and y be $O(1)$ variables. If we rescale equation (3.25) by $\theta' = \epsilon^2\theta$, $y' = \sqrt{\epsilon}y$, and

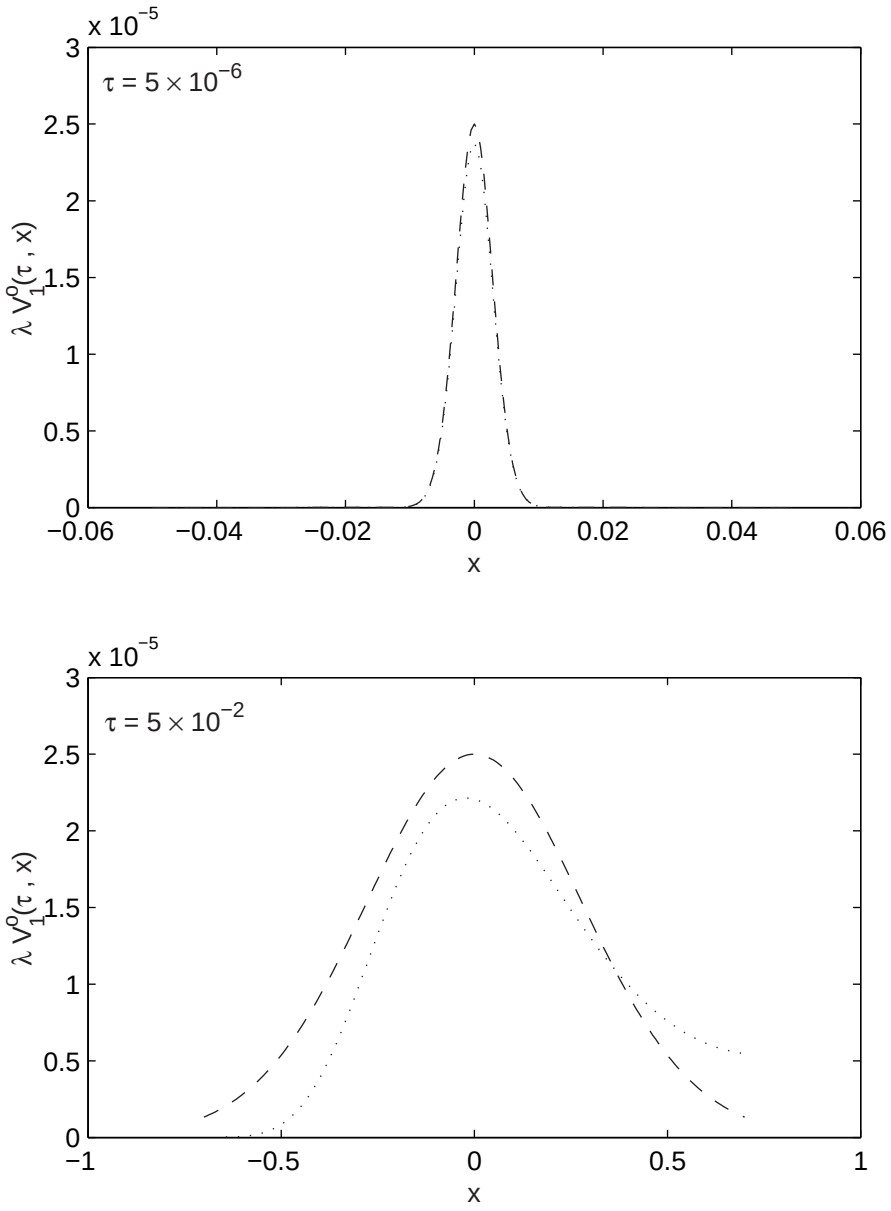


Figure 3.1: The first liquidity correction to the Black-Scholes value of a European call option in the outer region. The dotted line is the numerically calculated value from(3.12) and the dashed line is the leading-order approximation (3.18). Parameter values are $\bar{\lambda} = 10^{-4}$, $\alpha = 0.5$, $r = 0.10$, $\sigma = 0.3$, and $K = 100$.

$V_0^i = \sqrt{\epsilon} v^i$ (so that the intermediate variables θ', y', and V_0^i are taken to be $O(\epsilon)$, $O(\sqrt{\epsilon})$, and $O(\sqrt{\epsilon})$, respectively) we are left with

$$v_\theta^i = (1-\alpha)^2 v_{yy}^{i^3} + 2\sqrt{\epsilon} v_{yy}^{i^2} + \epsilon v_{yy}^i. \tag{3.26}$$

We will find it useful to define $u(\theta, y) = v_{yy}^i(\theta, y)$; differentiating (3.26) twice *w.r.t.* y we then obtain

$$u_\theta = \left((1-\alpha)^2 u^3 + 2\sqrt{\epsilon} u^2 + \epsilon u\right)_{yy}, \tag{3.27}$$

$$u(0, y) = \delta(y), \quad \text{and} \tag{3.28}$$

$$u(\theta, y) \to 0 \quad \text{as} \quad y \to \pm\infty. \tag{3.29}$$

Now posing a further inner expansion

$$u \sim u_0 + \sqrt{\epsilon} u_1 + \cdots, \tag{3.30}$$

we arrive at the problem

$$u_{0_\theta} = (1-\alpha)^2 \left(u_0^3\right)_{yy}, \tag{3.31}$$

$$u_0(0, y) = \delta(y), \qquad \text{and} \qquad u_0(\theta, y) \to 0 \quad \text{as} \quad y \to \pm\infty, \tag{3.32}$$

which gives the leading-order behaviour of the solution in the region $\tau \ll \epsilon^2 \bar{\lambda}^2$, $|x| \ll \sqrt{\epsilon}\bar{\lambda}$.

Equation (3.31) is a porous medium equation of order 3; with the point-source boundary conditions (3.32) it has the solution

$$u_0(\theta, y) = \begin{cases} \frac{1}{\sqrt{12}(1-\alpha)\theta^{\frac{1}{4}}} \left[C_1^2 - \frac{y^2}{\sqrt{\theta}}\right]^{\frac{1}{2}} & |y| < C_1 \theta^{\frac{1}{4}}, \\ 0 & |y| \geq C_1 \theta^{\frac{1}{4}}, \end{cases} \tag{3.33}$$

which is the order-3 Barenblatt-Pattle function [36]. The constant $C_1^2 = \frac{2\sqrt{12}(1-\alpha)}{\pi}$ has been determined from the condition

$$\int_{-C_1\theta^{\frac{1}{4}}}^{C_1\theta^{\frac{1}{4}}} u_0(\theta, y) dy = 1,$$

which ensures the proper behaviour at the boundaries for the call option (*i.e.* that $V \sim y$ as $y \to +\infty$). To demonstrate the form of the the Barenblatt-Pattle function we show plots of $u_0(y)$ for several values of θ in Figure 3.2.

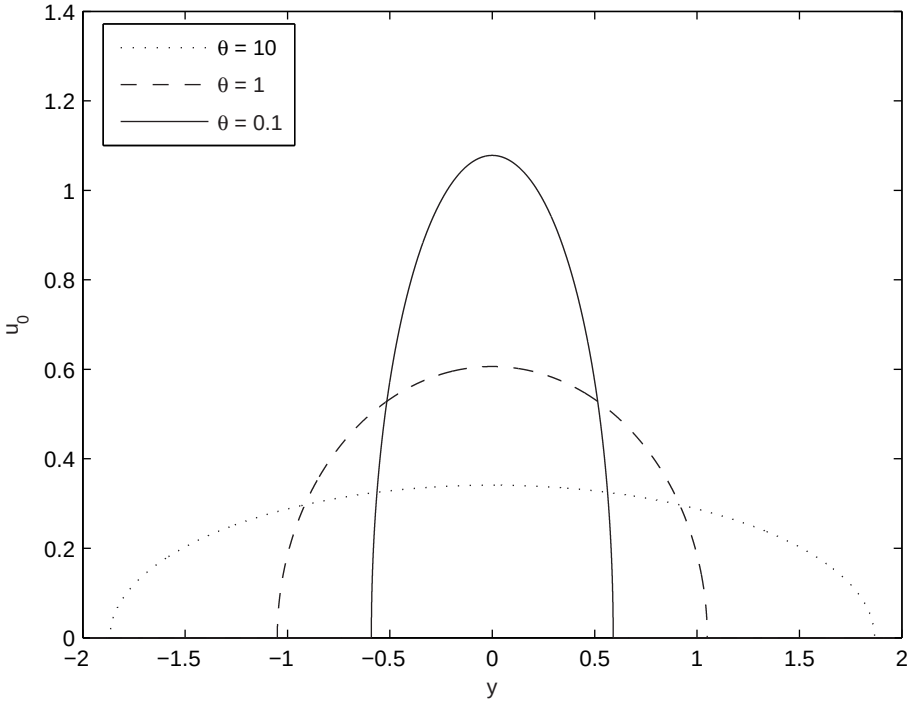


Figure 3.2: Plots of $u_0(y)$ for $\theta = 0.1,\ 1,$ and 10.

From (3.33) it is clear that $u_0 \to 0$ with infinite slope at $y = \pm C_1 \theta^{\frac{1}{4}}$, but because of the (linear) ϵu term in (3.27) we know that the exact solution u must be strictly positive for all y even if the initial condition has compact support. To incorporate this infinite support into the boundary layer approximation without calculating the $O(\epsilon)$ term in the expansion (3.24) (where the compact support first enters) we will seek a small tail correction to u_0 in the form of a travelling wave solution (*i.e.* one that is stationary *w.r.t.* the moving boundaries $y = \pm C_1 \theta^{\frac{1}{4}}$).

We will focus on the tail solution near the boundary moving in the positive y direction; the analysis for the boundary moving to the left in the region $y < 0$ is completely analogous. Let $s(\theta; \epsilon)$ be the centre of the tail region and z an $O(1)$ variable. If we transform into the

moving boundary reference frame by the rescaling

$$y = s(\theta;\epsilon) + \epsilon z, \qquad \text{and} \qquad u = \sqrt{\epsilon} f,$$

we obtain

$$\sqrt{\epsilon} f_\theta - \dot{s} f_z = \left((1-\alpha)^2 f^3 + 2f^2 + f\right)_{zz}, \tag{3.34}$$

$$f(\theta, z) \to 0 \quad \text{as} \quad z \to +\infty, \tag{3.35}$$

where $\dot{s} = ds/d\theta$. Note that since we have focused on the positive boundary we have only the boundary condition for $z \to +\infty$. The free parameter left by this choice will be $s(\theta, \epsilon)$ which will be determined from matching with the Barenblatt-Pattle outer solution. If we now pose the expansion

$$f \sim f_0 + o(1), \tag{3.36}$$

then we are left with the ODE

$$\left((1-\alpha)^2 f_0^3 + 2f_0^2 + f_0\right)_{zz} + \dot{s} f_{0_z} = 0, \tag{3.37}$$

$$f_0 \to 0 \quad \text{as} \quad z \to +\infty. \tag{3.38}$$

Integrating equation (3.37) once gives

$$\left(3(1-\alpha)^2 f_0^2 + 4f_0 + 1\right) f_{0_z} + \dot{s} f_0 = k_1(\theta),$$

and imposing the boundary condition fixes $k_1(\theta) \equiv 0$. Integrating once more yields the final result

$$\frac{3}{2}(1-\alpha)^2 f_0^2 + 4f_0 + \ln(f_0) = k_2(\theta) - \dot{s} z. \tag{3.39}$$

The second integration constant, $k_2(\theta)$, simply acts to shift f_0 horizontally, but in constructing the problem for the tail, its lateral position, $s(\theta, \epsilon)$, has been left free and will be fixed by matching to the solution (3.33); $k_2(\theta)$ can therefore be considered arbitrary to $O(1)$. For the remainder of the analysis $k_2(\theta)$ will simply be defined by $f_0(\theta, 0) = 1$ which gives $k_2(\theta) = \frac{3}{2}(1-\alpha)^2 + 4$.

We now need to determine the form of $s(\theta;\epsilon)$. The tail solution will overlap the inner

region of the boundary layer for $z < 0$; in the limit $z \to -\infty$ we have

$$f_0 \sim \left[\frac{-2\dot{s}z}{3(1-\alpha)^2}\right]^{\frac{1}{2}} + \cdots .$$

If we then make the substitution $y = C_1\theta^{\frac{1}{4}} + \xi$ into (3.33) and expand we find

$$u_0(\theta,\xi) = \left[\frac{C_1\theta^{-\frac{3}{4}}\xi}{6(1-\alpha)^2}\right]^{\frac{1}{2}} + \cdots \qquad \text{as} \quad \xi \to 0^+,$$

and so for the two solutions to match we must have $\dot{s} = \frac{1}{4}C_1\theta^{-\frac{3}{4}}$. Integrating *w.r.t.* θ and requiring $s(0;\epsilon) = 0$ then gives the (somewhat expected) result

$$s(\theta;\epsilon) \sim C_1\theta^{\frac{1}{4}} + o(1). \tag{3.40}$$

Equation (3.40) shows that, to leading order, the position of the tail solution is fixed by the moving boundary of the Barenblatt-Pattle function. In the (τ, x) space the composite solution can therefore be written

$$\hat{V}(\tau,x) \sim \begin{cases} \bar{\lambda}\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{x'} f_0(\tau,x'')dx''dx' & x \le -C_1\sqrt{\bar{\lambda}}\tau^{\frac{1}{4}}, \\ \bar{\lambda}\int\limits_{-\infty}^{-|x|}\int\limits_{-\infty}^{-|x'|} f_0(\tau,x'')dx''dx' + \dfrac{\sqrt{\bar{\lambda}}\tau^{\frac{1}{4}}}{\sqrt{12}(1-\alpha)}\left[-\dfrac{1}{6}\left(C_1^2 - \dfrac{x^2}{\bar{\lambda}\sqrt{\tau}}\right)^{\frac{3}{2}}\right. & \\ \left. + \dfrac{C_1^2}{2}\left[\dfrac{x}{\sqrt{\bar{\lambda}}\tau^{\frac{1}{4}}}\sin^{-1}\left(\dfrac{x}{C_1\sqrt{\bar{\lambda}}\tau^{\frac{1}{4}}}\right) + \left(C_1^2 - \dfrac{x^2}{\bar{\lambda}\sqrt{\tau}}\right)^{\frac{1}{2}}\right]\right] + \dfrac{x}{2} & |x| < C_1\sqrt{\bar{\lambda}}\tau^{\frac{1}{4}}, \\ x + \bar{\lambda}\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{x'} f_0(\tau,x'')dx''dx' & x \ge C_1\sqrt{\bar{\lambda}}\tau^{\frac{1}{4}}, \end{cases} \tag{3.41}$$

where u_0 has been integrated twice *w.r.t.* x and its boundary conditions have been implemented.

In Figure 3.3 is a plot of a numerically calculated $\hat{V}$ at $\tau = 10^{-10}$ as well as the equivalent asymptotic approximation given by (3.41). The numerical method for approximating $\hat{V}$ from (3.24) and (3.25) requires extra attention for very small τ due to the combination of a slope discontinuity in the payoff function at $x = 0$ and the presence of $\hat{V}_{xx}^2$ and $\hat{V}_{xx}^3$

terms in the PDE. If the solution were advanced forward using a standard finite difference routine there might immediately be a problem of overloading the machine when calculating the discrete approximations to the nonlinear terms of the equation near $\tau = 0$ and $x = 0$. To avoid this problem the routine has been split into two steps. For the first 10% of the timesteps in the calculation, the solution calculated is simply the Black-Scholes solution; this is done to smooth out the initial data.[1] After this initial smoothing-out period the nonlinear terms of the Bakstein-Howison equation are re-incorporated into the calculations and the solution is advanced forward to the desired value of τ.

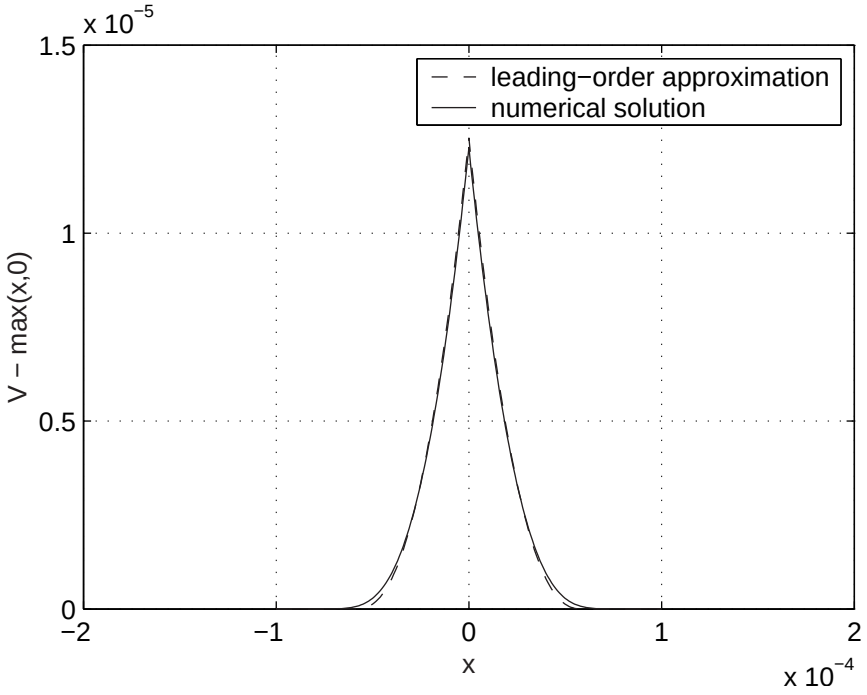


Figure 3.3: Numerical solution of $\hat{V} - \max(x, 0)$ and its leading-order asymptotic approximation for $\tau = 10^{-10}$. Parameter values used were $\lambda = 10^{-4}$, $\alpha = 0.5$, $r = 0.10$, $\sigma = 0.3$, and $K = 100.0$.

While the approximation (3.41) shown in Figure 3.3 is, in general, good, it underestimates the exact solution for intermediate values of x in the boundary layer. The accuracy of the approximation (3.41) could be improved by calculating higher-order terms of the expansion (3.30). In a paper on the diffusion of dopant ions in silicon, King and Please [32]

[1] Initial periods of 5% and 15% of the total number of timesteps in the calculation were also tried and the final solution was found to be relatively insensitive to the choice.

examine an equation very similar to (3.27) except without a u^3 term. In their analysis they calculate not only the leading-order Barenblatt-Pattle solution, but also its first-order correction and then match with the tail solution through higher-order terms in the expansion for $s(\theta;\epsilon)$. This correction term is found as the solution of an ODE in terms of a similarity variable, $\eta = y/\theta^{\beta}$. In [32] where u^2 is the highest-order spatial term in the governing equation this solution can be expressed simply in terms of Legendre functions whose asymptotic forms near $y = \pm C_1\theta^{\frac{1}{4}}$ are well known. This same calculation has been attempted for the valuation problem of this chapter, but because of the u^3 term u_1 cannot be expressed simply in terms of any special functions and so matching its asymptotic form with the tail solution is far more difficult.

3.4 Matching

To demonstrate the matching of the inner and outer approximations we first recall that, to leading order in the outer region

$$\hat{V} \sim \frac{x}{2}\left(\operatorname{erf}\left(\frac{x}{2\sqrt{\tau}}\right)+1\right)+\sqrt{\frac{\tau}{\pi}}e^{-\frac{x^2}{4\tau}}$$

and

$$\hat{V}_{xx} \sim \frac{1}{2\sqrt{\pi\tau}}e^{-\frac{x^2}{4\tau}} \tag{3.42}$$

as $\bar{\lambda} \to 0$.

In order to match with this leading-order component of the outer solution we need the behaviour of the leading-order inner solution, $V_0^i(\theta', y')$, in the limit $\theta' \to +\infty$. Since we lack an exact V_0^i we will need to take a more indirect approach to find its large-θ' behaviour. We begin by restating the problem for the leading-order behaviour of the call option in the boundary layer; it is

$$V^i_{0_{\theta'}} = V^i_{0y'y'} + 2V^{i^2}_{0y'y'} + (1-\alpha)^2 V^{i^3}_{0y'y'}, \tag{3.43}$$

$$V_0^i(0, y') = \max(y', 0), \tag{3.44}$$

$$V_0^i(\theta', y') \to 0 \quad \text{as} \quad y' \to -\infty, \qquad \text{and} \qquad V_0^i(\theta', y') \sim y' \quad \text{as} \quad y' \to +\infty. \tag{3.45}$$

If we make the substitution $w = V^i_{0_{y'y'}}$ then the problem becomes

$$w_{\theta'} = \left(w + 2w^2 + (1-\alpha)^2 w^3\right)_{y'y'}, \tag{3.46}$$

$$w(0, y') = \delta(y'), \qquad \text{and} \tag{3.47}$$

$$w(\theta', y') \to 0 \qquad \text{as} \quad y' \to \pm\infty. \tag{3.48}$$

Now integrating equation (3.46) *w.r.t.* y' gives

$$\begin{aligned}\frac{d}{d\theta'}\int_{-\infty}^{+\infty} w dy' &= \int_{-\infty}^{+\infty} \left(w + 2w^2 + (1-\alpha)^2 w^3\right)_{y'y'} dy' \\ &= \left(w_{y'} + 4ww_{y'} + 3(1-\alpha)^2 w^2 w_{y'}\right)\Big|_{-\infty}^{+\infty}.\end{aligned}$$

Since $w \to 0$ (and therefore $w_{y'} \to 0$) as $y' \to \pm\infty$ we are left with

$$\frac{d}{d\theta'}\int_{-\infty}^{+\infty} w(\theta', y') dy' = 0,$$

showing that the mass of w is a conserved quantity.

The next step in determining the large-θ' behaviour of the boundary-layer solution involves noting that, as was shown in Section 3.3, the leading-order behaviour of the boundary-layer solution for small θ' is the Barenblatt-Pattle function (3.33). While w is very large near $y' = 0$ for small θ' (resulting in the dominance of w^3 over w^2 and w in this region), as θ' increases the localized mass of the initial condition spreads out and w will decrease in this region. Since the mass of w is conserved the nonlinear terms of (3.46) do not act as source terms; for large enough θ' we therefore expect $w \ll 1$ so that $w \gg w^2 \gg w^3$ and the linear diffusion equation will determine the leading-order behaviour of the solution in this large-θ' region. Given that $w(0, y') = \delta(y')$ we should therefore expect w to behave as the fundamental solution of the diffusion equation; that is

$$w(\theta', y') \sim \frac{1}{2\sqrt{\pi\theta'}} e^{-\frac{y'^2}{4\theta'}} \qquad \text{as} \quad \theta' \to +\infty. \tag{3.49}$$

By comparing (3.42) to (3.49) we see that the $\theta' \to \infty$ limit of the leading-order boundary-layer solution is the same as the $\tau \to 0$ limit of the outer solution thus confirming the

validity of the forms of our original inner and outer expansions.

3.5 Conclusions

In this chapter we have performed an asymptotic analysis on the ask price of a European call option in the $\gamma = 0$ model and have found that there exists a boundary layer of size $\tau = O(\overline{\lambda}^2)$ and $x = O(\overline{\lambda})$. In the outer region the solution is approximately the Black-Scholes value of the option with a first correction due to the transaction cost effect. Within the boundary layer the form of the pricing equation changes significantly so that the nonlinear liquidity terms dominate the option's value. To analyze the solution's behaviour within the boundary layer we have also focussed on the region near expiry and the strike price. Within this restricted region it has been found that, to leading order, the option's Γ falls sharply to zero at finite x for finite τ; this compact support within the solution implies that no rehedging of the contract is necessary for asset prices outside this range.

To summarize, for $\tau \gg \overline{\lambda}^2$ the Black-Scholes value of the contract dominates and the liquidity effects (themselves dominated by the transaction cost effect) enter only as higher-order corrections. As τ decreases the magnitude of the liquidity terms increase near the strike price until $\tau = O(\overline{\lambda}^2)$ and $x = O(\overline{\lambda})$ when they become of equal order to the linear terms in the pricing equation. As τ further decreases the liquidity effects begin to dominate the option's value and its Γ becomes much more localised near the strike price. Finally, for very small τ within the boundary layer the price slippage effect begins to completely dominate the option's value; since this effect enters through a power of $\hat{V}_{xx}$ the option's Γ displays the compact support (neglecting terms that are exponentially small) of its payoff function and it effectively does not need to be rehedged for asset prices outside of a narrow range of the strike price.

Chapter 4

Hedging Liquidity Risks with the American Forward

This chapter will focus on a valuation an American-type forward contract with the intention of it being used as a hedge against liquidity risks.

4.1 Liquidity Risks in an Imperfect Market

In Chapter 2 we discussed the pricing of derivative contracts in a finitely liquid market. As well as affecting the dynamics of the underlying, liquidity effects, such as the bid-ask spread and market impact, can pose a more serious threat when hedging a derivative. When deriving equation (2.18) it was necessary to choose the hedge quantity, $H = \tilde{V}_S$, to eliminate the leading-order component of the portfolio's randomness caused by the exogenous diffusion of the asset. But the liquidity effects can themselves prevent this hedging strategy from being carried out.

Imagine the case when one or more layers of a market are completely empty; this can occur if, for instance, the asset price crashes and investor confidence is very low. Even though the asset's value may still be non-zero it may be impossible to sell the required number of assets for *any* price; this phenomenon is termed a *liquidity hole* and it can have serious consequences for anyone holding a derivative on that asset. Maintaining a risk-free derivative portfolio is inherently tied to delta-hedging in the underlying; if trading cannot

be carried out the portfolio will become unhedged and the holder could end up with a large amount of unwanted risk. If a hole occurs near a discontinuity in the derivative's payoff or its slope and near expiry (where its Γ is largest) the results of this mishedging could be disastrous.

4.2 Hedging Liquidity Risks

To prevent the mishedging caused by liquidity holes it would be useful to have a guaranteed supply and demand of the underlying during the life of the derivative so that its holder could trade regardless of the state of its market. In [40] Scholes first mentions liquidity options (in the form of a simple put option on the underlying) as a form of pre-packaged liquidity that could be sold to derivative holders. Because of the risk involved, it is unclear whether anyone would be willing to provide this liquidity at a reasonable price; Scholes suggests that institutional investors with long horizons as the natural source. As an attempt to construct a form of this pre-packaged liquidity, Bakstein and Howison [3] suggest options on the Greeks of a derivative as a method of providing a guaranteed source of liquidity for hedging that derivative. For example, the magnitude of a rehedge quantity to maintain a delta-neutral position for a derivative is equal to the Γ of the contract; a put option on Γ would therefore provide the necessary ability to sell these assets if its market falls out.

The idea in this chapter is to use call and put versions of the forward contract (from now on referred to as a *call-forward* and *put-forward*) as inexpensive forms of supply and demand in the underlying. With the standard European forward the holder receives the asset for the forward price only at the expiry date. For the purpose of guaranteeing liquidity at *any* time during the primary derivative's life, this aspect of the contract is undesirable since it is *a priori* unknown when, if at all, the liquidity in the market will drop. The contract must therefore deliver the protection whenever the holder needs it. If the forward is given an American structure then the supply/demand would be at the holder's discretion and it could be utilized when it is needed.

The idea with using an American forward as a liquidity hedge is as follows:

- If we hold a derivative on an asset in an illiquid market then we would enter into an American foward agreement on the same underlying with the same expiry with some

third party.

- If, while hedging the original derivative, the liquidity in the market drops then the forward could be exercised giving the holder the ability to trade in the underlying and thus maintain the delta-neutrality of the original portfolio.

The attractiveness of the forward as a liquidity hedge is that it is usually constructed to have zero initial value to both parties and therefore does not require any premium payment by the holder.

The remainder of the chapter will present results for pricing the forward contract. To differentiate between specific forms of the contract we will use $\tilde{C}$ and $\tilde{P}$, respectively, as the values of the call-forward and put-forward *per unit of underlying* and Φ^C and Φ^P as their respective payoff functions. Furthermore, the superscripts e and a will designate the European and American versions of the contract while BS and BH will be used for the value of the contract in the Black-Scholes and Bakstein-Howison models, respectively; for instance, $\tilde{P}^{a,BH}(t,S)$ is the value of the American put-forward in the Bakstein-Howison model. Finally, when considerring the Bakstein-Howison model we will assume a market depth of $\lambda = 10^{-7}$ and so different values of $\overline{\lambda}$ will correspond to contracts on the appropriate number of assets; for example $\overline{\lambda} = 10^{-2}$ corresponds to a market with a depth of 10^{-7}/(asset) and a contract on 10^5 assets. Before focussing on the forward in a finitely liquid market we will first value and discuss the properties of the contract in a perfectly liquid Black-Scholes model as a comparison.

4.3 European Forward in the Black-Scholes Model

For the European call and put-forwards, the buyer and seller agree to exchange a number of assets each for the pre-determined forward price per asset, $\tilde{F}$, at the expiry date T. Simple arbitrage arguments lead us to the fair value of the forward price in a perfectly liquid market. As with all derivatives, the value of the contract is equal to the cost of hedging it; from the perspective of the short position, to hedge the contract requires holding the underlying. The cost of this hedge is simply the cost of purchasing the underlying at $t = 0$ and funding at the risk-free rate, r, less the total dividend payment received during this

time; that is

$$\tilde{F}^{e,BS} = S_0 e^{T(r-D)}, \tag{4.1}$$

(where $S_0 = S(t = 0)$) for both the call and put-forwards. With $T = 0.75$, $r = 0.10$, $D = 0.09$ and $S_0 = 99.2528055$ we find $\tilde{F}^{e,BS} = 100$.

4.4 American Forward in the Black-Scholes Model

The pricing problem for both the American call and put-forwards in the Black-Scholes model is given in equation (1.10). The boundary conditions for the call-forward are

$$\begin{aligned} \tilde{C}^{a,BS}(t,S) &\to Se^{-D(T-t)} \qquad \text{as} \quad S \to +\infty, \\ \tilde{C}^{a,BS}(t,0) &= -\tilde{F}e^{-r(T-t)}, \\ \Phi^C &= S - \tilde{F}, \end{aligned} \tag{4.2}$$

and those for the put-forward are

$$\begin{aligned} \tilde{P}^{a,BS}(t,S) &\to -Se^{-D(T-t)} \qquad \text{as} \quad S \to +\infty, \\ \tilde{P}^{a,BS}(t,0) &= \tilde{F}e^{-r(T-t)}, \\ \Phi^P &= \tilde{F} - S. \end{aligned} \tag{4.3}$$

Except for special cases (for instance, when $D = 0$, which is discussed below), the problem (1.10) with boundary conditions (4.2) or (4.3) has no exact analytical solution. All results for the remainder of the chapter have been generated numerically using explicit finite difference routines. Specifically, the solution is calculated at each time using central differences for all spatial approximations and a forward difference approximation in time (where the transformation $t \to -t$ has been made). At each time the solution is advanced ahead one timestep; if any solution value is less than its corresponding payoff value then the solution takes on this value. This step imposes the early exercise constraint by ensuring that the conditions $L_{BS}[V] = 0$ for $V > \Phi$ and $L_{BS}[V] < 0$ for $V = \Phi$ are obeyed .

4.4.1 The Case $D = 0$

When the forward is contingent on an underlying that pays no dividend the analysis is particularly simple. For the European call-forward we have found $\tilde{C}^{e,BS}(t,S) = Se^{-D(T-t)} - \tilde{F}e^{-r(T-t)}$. When $D = 0$ we see that $\partial\tilde{C}^{e,BS}/\partial S = \partial\Phi^C/\partial S$ for all $t \leq T$. As t moves backwards from expiry $\tilde{C}^{e,BS} > \Phi^C$ for all S; there can therefore be no free boundary with the American call-forward and its value must be identical to that of its European equivalent. That is,

$$\tilde{C}^{a,BS} = S - \tilde{F}e^{-r(T-t)}. \qquad (D = 0) \tag{4.4}$$

The value of the European put-forward in the Black-Scholes model is $\tilde{P}^{e,BS}(t,S) = \tilde{F}e^{-r(T-t)} - Se^{-D(T-t)}$. When $D = 0$ we again see that $\partial\tilde{P}^{e,BS}/\partial S = \partial\Phi^P/\partial S$ for all $t \leq T$, but now $\tilde{P}^{e,BS} < \Phi^P$ for all S. From our arguments in Chapter 1 every point (t,S) must therefore be a point where early exercise is optimal. We see that the put-forward on a non-dividend paying asset should be exercised at the earliest possible time. Generalizing, if the contract is given a Bermudan structure so that exercise can only occur at a set of times $\{t_0, t_1, \ldots t_m \leq T\}$ (where $t_0 < t_1 < \cdots < t_m$) then its value is

$$\tilde{P}^{a,BS}(t,S) = \begin{cases} \tilde{F}e^{-r(T-t)} - S & 0 \leq t < t_0 \\ \tilde{F} - S & t_0 \leq t \leq T \end{cases} \qquad (D = 0) \tag{4.5}$$

which gives the result $\tilde{F}^{a,BS} = S_{t_0}e^{rt_0}$ for the forward price.

4.4.2 The Case $D > 0$

For $D > 0$ the boundary condition at $S = \infty$ forces $\tilde{C}^{a,BS}(t,S) < \Phi^C$ (or, equivalently, in the case of a put-forward the boundary condition at $S = 0$ forces $\tilde{P}^{a,BS}(t,S) < \Phi^P$) and therefore a free boundary to form; plots of $\tilde{C}^{a,BS}(t,S)$ and $\tilde{P}^{a,BS}(t,S)$ are shown in Figure 4.1. Figure 4.2 shows $\tilde{C}^{a,BS}(t,S) - \max(\Phi^C, \tilde{C}^{e,BS}(t,S))$ and $\tilde{P}^{a,BS}(t,S) - \max(\Phi^P, \tilde{P}^{e,BS}(t,S))$ for the corresponding plots of Figure 4.1. With the call-forward there appear to be three distinct regions to the solution. For $S > S_f \quad (\approx 130$ in our example), $\tilde{C}^{a,BS}(t,S) = \Phi^C$; this is the early exercise region for the contract. In the hold region and as $S \to 0$ the solution rapidly approaches the linear form of its European equivalent,

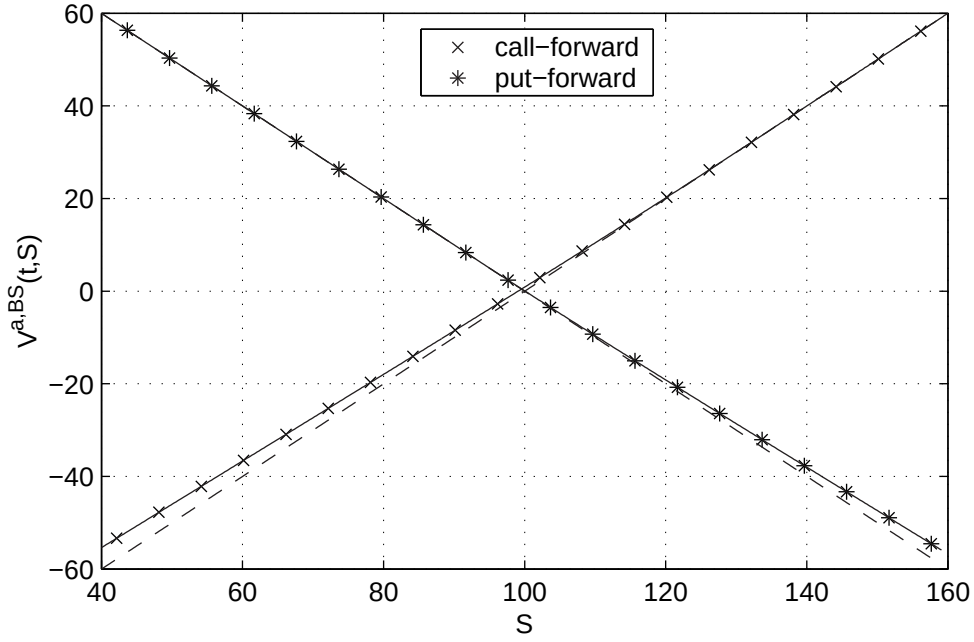


Figure 4.1: Plots of $\tilde{C}^{a,BS}(t,S)$ and $\tilde{P}^{a,BS}(t,S)$ for $r = 0.10$, $\sigma = 0.3$, $D = 0.09$, $\tilde{F} = 100$, and $T - t = 0.75$. Dashed lines show the payoffs for the respective contracts.

$\tilde{C}^{e,BS}(t,S) = Se^{-D(T-t)} - \tilde{F}e^{-r(T-t)}$. Again, with the parameters used for the results of the above figures, $\tilde{C}^{a,BS}$ is not significantly different from $\tilde{C}^{e,BS}$ for $S < 60$. Finally, in the hold region for $S \approx \tilde{F}$ the arbitrage requirement that $\tilde{C}^{a,BS}(t,S) \geq \Phi^C$ introduces curvature into the solution so that the two exterior regions will match and results in the only significantly non-zero data of Figure 4.2. Since this is the region where the solution is most non-linear it will also be the region where the liquidity effects are most significant; this will be discussed more below. The results for $\tilde{P}^{a,BS}(t,S)$ are similar to those of $\tilde{C}^{a,BS}(t,S)$ with one exception: deep in the hold region of the contract, $\tilde{P}^{a,BS}(t,S) - \tilde{P}^{e,BS}(t,S)$ is much larger than the equivalent value for $\tilde{C}^{a,BS}(t,S)$ and we therefore expect the liquidity effects to be significant over a wider range with $\tilde{P}^{a,BS}(t,S)$ than for $\tilde{C}^{a,BS}(t,S)$.

Using the value $S_0 = 99.2528055$ (which was chosen so $\tilde{F}^{e,BS} = 100$), we find $\tilde{F}^{a,BS} = 100.1736584$ for the call-forward and $\tilde{F}^{a,BS} = 99.1929411$ for the put-forward.

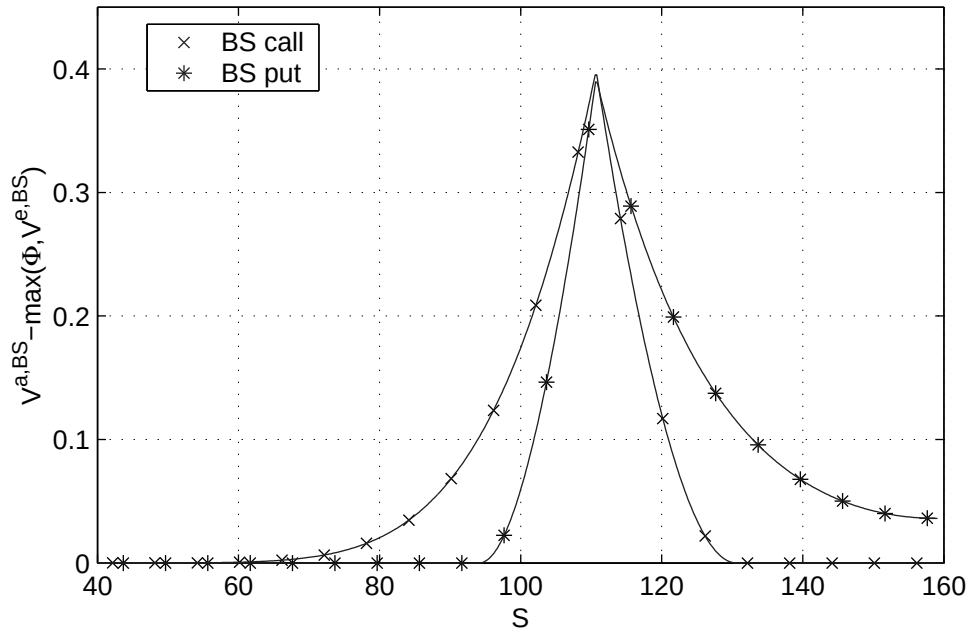


Figure 4.2: Plots of $\tilde{C}^{a,BS}(t,S) - \max(\Phi^C, \tilde{C}^{e,BS}(t,S))$ and $\tilde{P}^{a,BS}(t,S) - \max(\Phi^P, \tilde{P}^{e,BS}(t,S))$ for $r = 0.10$, $\sigma = 0.3$, $D = 0.09$, $\tilde{F} = 100$, and $T - t = 0.75$.

4.5 The Forward in the $\gamma = 0$ Model

The free boundary pricing problem for a general American contract in the $\gamma = 0$ model was defined in Section 2.3. The ask price of the American call-forward is specifically determined by

$$\begin{aligned} L_{a_{BH}}[\tilde{C}_a(t,S)] &= 0 && S \in Z, \\ L_{a_{BH}}[\tilde{C}_a(t,S)] &< 0 && S \in Z^*, \\ \tilde{C}_a(t, S_{af}(t)) &= \Phi^C(S_{af}(t)), \\ \tilde{C}_{a_S}(t, S_{af}(t)) &= \Phi^C_S(S_{af}(t)), \end{aligned} \tag{4.6}$$

and that for the American put-forward by

$$\begin{aligned} L_{a_{BH}}[\tilde{P}_a(t,S)] &= 0 && S \in Z, \\ La_{BH}[\tilde{P}_a(t,S)] &< 0 && S \in Z^*, \\ \tilde{P}_a(t,S_{af}(t)) &= \Phi^P(S_{af}(t)), \\ \tilde{P}_{a_S}(t,S_{af}(t)) &= \Phi^P_S(S_{af}(t)), \end{aligned} \tag{4.7}$$

where the boundary conditions for each contract are the same as those in the Black-Scholes model and are given in (4.2) and (4.3). As usual, the bid prices are found simply by solving (4.6) and (4.7) for $V \to -V$ and $\Phi \to -\Phi$.

4.5.1 The Case $D = 0$

The non-linear effects of the $\gamma = 0$ model enter through powers of $\tilde{V}_{SS}$; in Section 4.4.1 we found that $\tilde{C}^{a,BS}_{SS} = 0$ and $\tilde{P}^{a,BS}_{SS} = 0$ for all S and $t \leq T$ and so we have

$$\tilde{C}^{a,BH}(t,S) = \tilde{C}^{a,BS}(t,S), \qquad (D=0) \tag{4.8}$$

and

$$\tilde{P}^{a,BH}(t,S) = \tilde{P}^{a,BS}(t,S). \qquad (D=0) \tag{4.9}$$

4.5.2 The Case $D > 0$

In Figure 4.3 are the values of both long and short positions for the American forwards in the $\gamma = 0$ model. The deviation of these values from the Black-Scholes equivalents is shown in Figure 4.4. It is clear that the correction to the Black-Scholes price due to the liquidity effects is very small in this model; even with a value $\overline{\lambda} = 10^{-2}$ (corresponding to a contract on 10^5 units of the underlying) the relative increase in the Black-Scholes value is $< 10^{-3}$/contract. In the $\gamma = 0$ model we neglect the bid-ask spread in the underlying. From equation (2.16) we see that the effect of the bid-ask spread enters into the valuation at a lower order than that of the market depth; by neglecting the spread we are therefore focussing on higher-order effects and so the small deviation is not surprising.

Figure 4.4 shows the deviation of the American forward's value in the $\gamma = 0$ model from

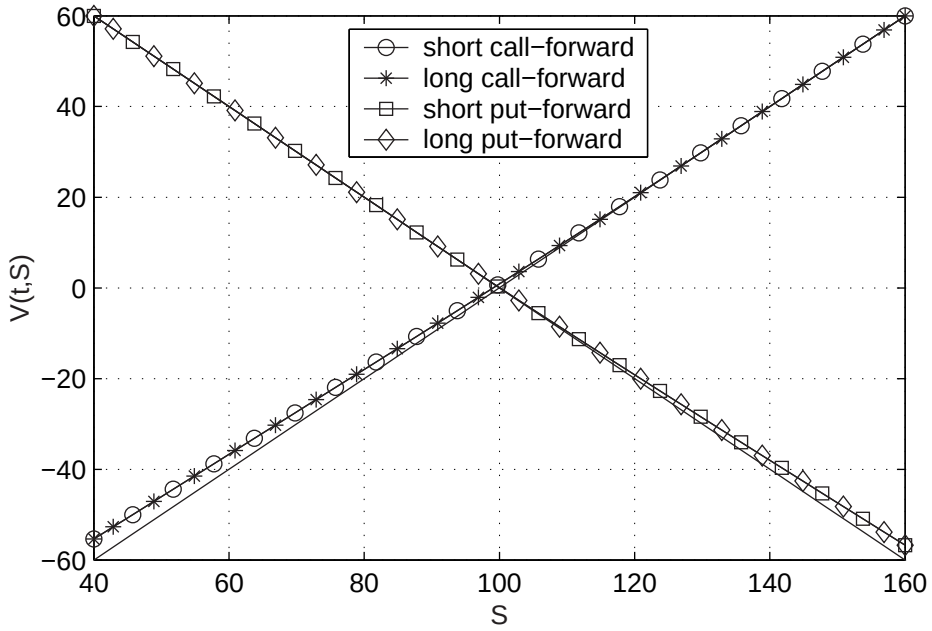


Figure 4.3: Long and short positions of the American call and put-forwards in the $\gamma = 0$ model. Parameter values are $r = 0.10$, $\sigma = 0.3$, $D = 0.09$, $\tilde{F} = 100$, $T = 0.75$, and $\overline{\lambda} = 10^{-2}$.

that in the Black-Scholes model. The liquidity effects are localized in the hold region near the contract's free boundary over precisely the range that the equivalent American contract in the Black-Scholes model is significantly non-linear as was shown in Figure 4.2. Outside this region the contract's Γ is exactly zero (in the early exercise region) or very nearly so (in the hold region) and liquidity effects will be insignificant due to a lack of need to rehedge the contract there.

In Table 4.1 we show forward prices for several values of $\overline{\lambda}$ and α. To reiterate, the forward price is the price paid at time T for an initial asset price, S_0, such that the contract has zero initial value. For the call-forward we see $\tilde{F}^{a,BH}_{\text{long}} < \tilde{F}^{a,BS}$ and $\tilde{F}^{a,BH}_{\text{short}} > \tilde{F}^{a,BS}$; the holder of the contract incurs an additional expense due to the liquidity effects and therefore calculates the forward price to be less than if there were no such costs. For the put-forward $\tilde{F}^{a,BH}_{\text{long}} > \tilde{F}^{a,BS}$ and $\tilde{F}^{a,BH}_{\text{short}} < \tilde{F}^{a,BS}$; due to the cost of hedging the contract the holder feels he should be able to sell the underlying for a higher price than the Black-Scholes value.

We see that a bid-ask spread for $\tilde{F}$ has been generated by the $\gamma = 0$ model; while this

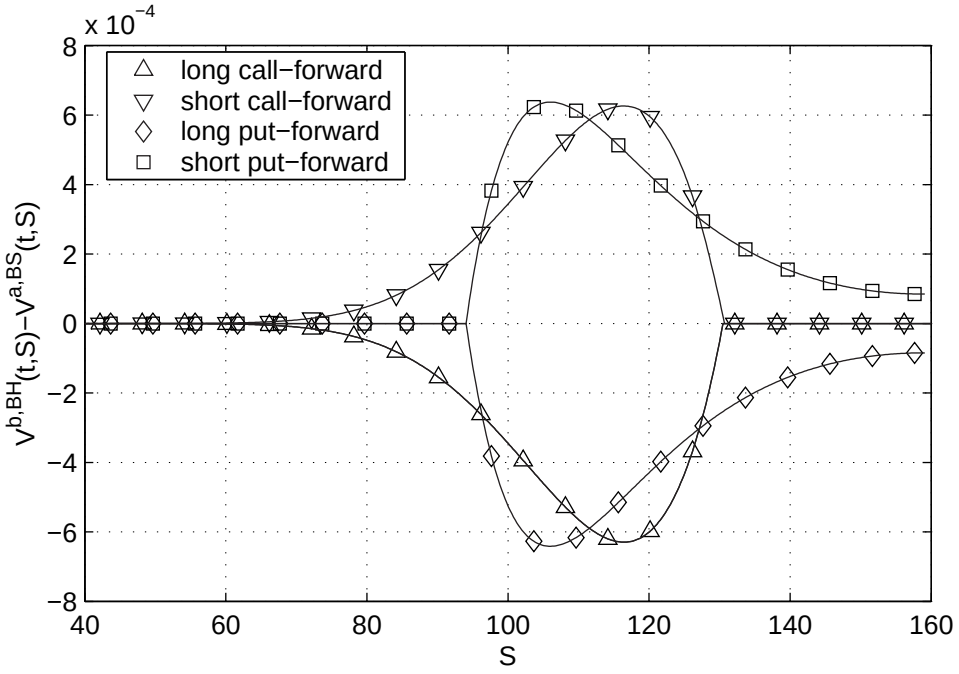


Figure 4.4: The deviation of the plots of Figure 4.3 from their corresponding Black-Scholes values.

spread is not entirely surprising given the presence of trading frictions, it *is* surprising since we have assumed no spread in the underlying. In this case the spread in the forward is a result of the trasaction cost effect enterring into the model through a finite market depth. Also of interest is the fact that the liquidity premium for the short position is greater than that for the long position. This is a result of the fact that the payoff functions for both the call and put-forwards are convex. When the asset price rises the writer of the contract must buy assets to maintain a delta-neutral portfolio; because of the price impact function this buy order will push the asset price even higher resulting in larger rehedging quantities. For the long position an increase is the asset price will require the holder to sell assets. This trade will then push the asset price down which will cancel part of the original change and result in a lower total cost of rehedging the contract over its life.

In terms of dependencies on the liquidity parameters, the bid-ask spread in $\tilde{F}$ is seen to increase with increasing $\overline{\lambda}$. This result is fairly obvious; a higher value of $\overline{\lambda}$ leads to higher transaction costs which, in turn, causes a greater difference in the liquidity adjusted

$(\overline{\lambda}, \alpha)$	$\tilde{F}$(short call)	$\tilde{F}$(long call)	$\tilde{F}$(short put)	$\tilde{F}$(long put)
$(0, -)$	100.173657336	100.173657336	99.1929413115	99.1929413115
$(10^{-6}, 0.5)$	100.173658411 $(1.075\ 10^{-6})$	100.173657302 $(-3.4\ 10^{-8})$	99.1929411057 $(-2.058\ 10^{-7})$	99.1929413647 $(5.32\ 10^{-8})$
$(10^{-4}, 0.5)$	100.173661825 $(4.489\ 10^{-6})$	100.173653888 $(-3.448\ 10^{-6})$	99.1929358395 $(-5.472\ 10^{-6})$	99.192946631 $(5.3195\ 10^{-6})$
$(10^{-2}, 0.5)$	100.174002826 $(3.4549\ 10^{-4})$	100.173312089 $(-3.45247\ 10^{-4})$	99.192408106 $(5.332055\ 10^{-4})$	99.1934740396 $(5.327281\ 10^{-4})$
$(10^{-6}, 0.0)$	100.173658411 $(1.075\ 10^{-6})$	100.173657302 $(-3.4\ 10^{-8})$	99.1929411057 $(-2.058\ 10^{-7})$	99.1929413647 $(5.32\ 10^{-8})$
$(10^{-4}, 0.0)$	100.173661825 $(4.489\ 10^{-6})$	100.173653888 $(-3.448\ 10^{-6})$	99.1929358394 $(-5.473\ 10^{-6})$	99.192946631 $(5.3195\ 10^{-6})$
$(10^{-2}, 0.0)$	100.174002995 $(3.45659\ 10^{-4})$	100.173312259 $(-3.45077\ 10^{-4})$	99.1924076411 $(-5.336704\ 10^{-4})$	99.1934735677 $(5.322562\ 10^{-4})$

Table 4.1: Short and long forward prices per share for the American call and put-forward in the $\gamma = 0$ model with $r = 0.10$, $D = 0.09$, $\sigma = 0.3$, $T - t = 0.75$, and $S_0 = 99.2528055$. The corresponding liquidity premium (*i.e.* $\tilde{F}^{a,BH} - \tilde{F}^{a,BS}$) is shown in brackets.

value perceived by the writer and the holder. Bid-ask spreads can also be seen to increase with decreasing α. A decrease in α corresponds to a greater permanent price slippage; this slippage will necessitate larger rehedgings which will result in larger total hedging costs over the length of the contract and thus a greater spread between the short and long positions.

It was stated at the beginning of the chapter that the American forward was an attractive liquidity hedging tool because of its low cost. Although the typical construction of a forward has $\tilde{F}$ chosen to give the contract zero initial value, this is not necessary; with the same contract structure we could choose any $\tilde{F}$ for a given S_0 and a premium could be exchanged between the two parties. To compare the cost of an American forward liquidity hedge with another possible method, we show the value of an at-the-money American call-forward and American call option in the $\gamma = 0$ model in Table 4.2 below. Both of these contracts allow the holder to purchase a unit of the underlying at the same price at any time during the life of the contract, but the forward is obviously a much less expensive hedging instrument due to the presence of a possible negative payoff for the holder. Also, it is much less affected by the presence of liquidity effects as a result of its smaller Γ near strike.

	$\check{V}_{short}$	$\check{V}_{long}$
American call-forward	0.87292	0.87223
American call option	10.29986	9.87321

Table 4.2: Value of at-the-money American call-forward and call option. $r = 0.10$, $D = 0.09$, $\sigma = 0.3$, $T - t = 0.75$, $K = \tilde{F} = 100$, $\alpha = 0.5$, and $\overline{\lambda} = 10^{-2}$.

4.6 Conclusions

We have constructed and priced an American-type forward contract in the $\gamma = 0$ BH model. Even though the spread in the underlying was neglected, we found that a spread between the bid and ask values of the forward price developed. Furthermore, it has been found that this spread increased with decreasing market depth and increasing price slippage and that it was asymmetric; the liquidity premium being greater for the short position than for the long. Because the call and put forms of the contract provide guaranteed supply and demand of the underlying at any time during its life it is thought that the American forwards could be effective hedging instruments against liquidity risks in an illiquid market. We have found that the price of these contracts is much smaller than the equivalent American option contract due to the possible downside risk to the holder in the forward. Furthermore, because of the very small Γ of the American forwards the liquidity premia of the contracts is also very small; thus not only are the contracts an effective liquidity hedging tool, but they are also very inexpensive to the holder.

Even in a very shallow market with significant price slippage we have seen that the values of the short positions of the American forwards are quite small; while this result may not seem surprising, it is considering the very large loss that the writer could possibly incur. The holder of a liquidity hedge possesses the contract in order to provide a supply or demand of the asset when the liquidity of the market drops significantly which would most likely result from a large boom or crash in the asset price. In deriving the model for the asset price dynamics we assumed that (in the limit of continuous-time) the exogenous diffusion of the asset is generated by a Geometric Brownian Motion, but this construction does not allow for large jumps in the asset price with any significant probability. In valuing the contract we have therefore calculated the liquidity premium for a hedge against the likelihood of the asset making a large change given an asset price model in which these

changes are very unlikely; it is therefore not surprising that the premium is small. As a next step it would be useful to investigate the valuation of the contract in a finitely liquid market in which the exogenous asset price dynamics are generated by a jump-diffusion, or Levy-stable process.

Finally, as has already been stated, our intention for the American forward is that it be used as a hedge against liquidity risks in a finitely liquid market. By definition, liquidity risk is the risk associated with a change in an asset's liquidity, but in this chapter we have assumed that the liquidity of the asset is a constant (through the fact that λ and α are constant). A necessary next step in developing the American forward for use as a liquidity hedge would therefore be to price it in a model in which there exists uncertainty in the level of the market's liquidity. In Chapters 5 and 6 we extend the BH model to allow for a stochastic market depth and then re-derive the pricing equation for a derivative contract under these conditions; we will return the American forward in this more realistic situation at that time.

Chapter 5

Derivative Pricing in a Market with Stochastic Liquidity - Part I

In Chapter 1 we developed an asset price model that incorporated a trading-induced feedback mechanism into the price due to the market impact of a large trader's trade. In addition to this permanent market impact effect, the model also accounted for the effect of the bid-ask spread and the market depth of the asset directly on the cost of the trade to the large trader.

To represent these three effects the model possesses three parameters in addition to those in the standard Black-Scholes model. In Chapters 2, 3, and 4 we have so far treated each of these parameters as constants, but it is well known (see, for instance [4], [12], [13], and [14]) that in reality, market liquidity is not only time-dependent, but also stochastic. In addition to this general behaviour, liquidity is also known to exhibit strong temporal patterns. For instance, on a weekly basis, liquidity varies noticeably from day to day, typically being highest on Tuesdays and lowest on Fridays. But an even more prominent feature is the intra-day variation; liquidity is typically very high to begin the day, falls steadily until approximately midday, and then rises again throughout the afternoon.

To make the BH model more realistic we would like to incorporate the time-dependent nature of the market's liquidity. Since liquidity is greatest at the beginning and end of the day, traders should weight their programs to these times to take advantage of the favourable conditions. Furthermore, since a derivative's price is inherently linked to a hedging strategy

in the underlying, variable liquidity conditions that affect trading strategies should therefore have an effect on the derivative's price.

Probably more important, though, than the overall (deterministic) time-dependent nature of the market's liquidity is the stochasticity that it demonstrates. Our work in Chapter 4 focussed on pricing the American forward contract for it to be used as a hedge against liquidity risk. While it is important to characterise the effect of a known level of market impact and bid-ask spread for derivative pricing and portfolio trading purposes, of far greater importance is the effect of an unforeseen change in their values: As was discussed in Chapter 4, if a liquidity hole appears while hedging a large derivative position it may be impossible to carry-out the hedge leaving open the possibility for large losses.

5.1 Previous Literature

The majority of work in the field of liquidity risk modelling has focussed on incorporating liquidity effects into a capital asset pricing model (CAPM) framework, the so-called liquidity-adjusted CAPM; two examples of this approach are [1] and [37]. The general idea behind this approach is to specify a stochastic model for the supply and/or demand curve for an asset. Along with this finitely elastic curve, traders in the market are modelled as price takers; these two factor combine to affect the traders' wealth, but not the permanent value of the asset price and these studies therefore only examine a stochastic *temporary* market impact effect. With this set-up the traders then maximize their expected utility for a consumption/allocation problem (*i.e.* the Merton problem) over a finite horizon which results in an equilibrium asset price adjusted by the included liquidity effects.

Using a very similar model, the effect of liquidity risk on derivative prices has only been investigated very recently by Cetin, Jarrow, and Protter [10], by Cetin, Jarrow, Protter, and Warachka [11] and subsequently reviewed by Jarrow and Protter [31]. Again, traders are modelled as price takers and the supply curve as finitely elastic and the work therefore again focusses on temporary impact effects only. In [10] it is found that if the conditions are satisfied for an arbitrage-free market and continuous trading strategies are allowed then the market impact function (supply curve) must be horizontal and the standard Black-Scholes price holds. The situation of allowing an upward sloping impact function is specifically examined in [11]; it is found in this case that continuous trading strategies must be forbidden

to admit a unique derivative price. Under this discrete trading condition and using dynamic programming the price is determined numerically from the optimal super-replication strategy for the contract. One interesting conclusion from this work is that the standard delta-hedging strategy is often not the optimal strategy in the presence of liquidity risk.

5.2 Outline of Work

Over this and the next chapter we aim to extend our derivative pricing framework to an underlying asset that possesses liquidity risk. To model this liquidity risk we will take a more straightforward approach than used in [10] and [11]; we will simply exogenously specify a stochastic model for the underlying's market depth that possesses the important mean-reverting characteristic found empirically. With this model it is then a relatively straightforward task to derive a two-dimensional Black-Scholes type pricing equation.

But because of the mean-reverting nature of the asset's market depth and, specifically, that its period of one day is much less than the typical life of a derivative contract, we have an analogous situation to pricing a contract in a market in which volatility is stochastic. In the stochastic volatility example, the fast mean-reversion of the asset's volatility is ingeniously utilized by Fouque, Papinicolaou, and Sircar (FPS) in [18], [19], [20], [21], and [22] to analyze the pricing equation with asymptotic techniques with which they are able to derive a much more simple and easily calibratible pricing framework. As the pricing equation including liquidity risk will be shown to be very complex and because of the structural similarity of our stochastic liquidity model (to be developed in Chapter 6) to the stochastic volatility model used by FPS, we intend to apply this asymptotic technique to the liquidity risk problem with the intention of producing a more applicable pricing framework.

The asymptotic analysis developed by FPS, however, is not trivial. As the BH model for a finitely liquid asset is nonlinear, the pricing equation for a derivative on such an asset will also be nonlinear and the analysis of the stochastic liquidity model will prove even more complicated than for the linear stochastic volatility model. At this point, therefore, we will digress and spend the remainder of this chapter reproducing the original work of FPS for a stochastic volatility model in order to develop the method in a more simple setting. In Chapter 6 we will then return to the problem of pricing a derivative on an asset possessing liquidity risk and extend the FPS analysis to this nonlinear problem.

The rest of the chapter will proceed as follows. In Section 5.3 we will present a model for an asset price where its volatility is driven by a fast, mean-reverting stochastic process. In Section 5.4 we derive the equation for the price of a derivative contingent on this asset and then present the FPS asymptotic analysis of this equation in the limit of fast mean-reversion in Section 5.5. In Section 5.6 we finally discuss how the approximate pricing framework can be calibrated to market data.

5.3 Asset Price Model

We assume a frictionless market for the asset; we therefore model the asset price dynamics with the simple Geometric Brownian Motion

$$dS_t = \mu S_t dt + \sigma_t S_t dW_t, \tag{5.1}$$

where, as usual, dW_t is the increment of a Wiener process, μ is the deterministic growth rate of the asset, and σ_t is its volatility which has been indexed by t to emphasize the fact that it is variable (and specifically stochastic) in this model.

To model the volatility as a mean-reverting process while maintaining some generality, we introduce the stochastic driving variable, y, which is generated by an Ornstein-Uhlenbeck process and let σ_t be some arbitrary function, f, of y. If a is the rate of mean-reversion of the driving process, m is its long-run mean, and β is its (constant) volatility, then the complete asset price model is

$$\begin{aligned} dS_t &= \mu S_t dt + \sigma_t S_t dW_t, \\ \sigma_t &= f(y_t), \\ dy_t &= a(m - y_t)dt + \beta d\hat{W}_t, \end{aligned} \tag{5.2}$$

where $d\hat{W}_t$ is the increment of a Wiener process partially correlated with dW_t; that is

$$dW_t \cdot d\hat{W}_t \sim \rho dt. \tag{5.3}$$

for a correlation coefficient $\rho \in [0, 1]$.

Before proceeding we need to calculate the long-run invariant distribution, $\phi(y)$, of the volatility process as it will be used in a later section. We begin with the probability, $p(y,t;y_0,t_0)$ that the volatility process hits y at time t given that it had a value of y_0 at $t_0 < t$; p then satisfies the Kolmogorov forward equation

$$\frac{\partial p}{\partial t} = \frac{1}{2}\beta^2 \frac{\partial^2 p}{\partial y^2} - \frac{\partial}{\partial y}\left(a(m-y)p\right), \tag{5.4}$$

with appropriate boundary conditions to ensure that p has the necessary properties of a probability distribution. By definition, the time-invariant distribution, $\phi(y)$, will satisfy $\frac{\partial \phi}{\partial t} = 0$ in which case (5.4) reduces to

$$\frac{1}{2}\beta^2 \frac{\partial^2 \phi}{\partial y^2} - a(m-y)\frac{\partial \phi}{\partial y} + a\phi = 0. \tag{5.5}$$

Integrating (5.5) twice *w.r.t.* y, imposing the boundary conditions that $\phi(y)$ is well-behaved as $y \to \pm\infty$, and then finally normalizing so $\int_{-\infty}^{+\infty} \phi(y)dy = 1$ gives the equation for our invariant distribution; this is

$$\phi(y) = \frac{\sqrt{a}}{\beta\sqrt{\pi}} \exp\left(-\frac{a(y-m)^2}{\beta^2}\right), \tag{5.6}$$

which tells us that, in the long run, $\phi(y) \sim \mathcal{N}(m, \nu^2)$ where

$$\nu^2 = \frac{\beta^2}{2a}. \tag{5.7}$$

.

5.4 Deriving the Pricing Equation

The asset price model (5.2) possesses both asset price and volatility risk. While there are two sources of uncertainty, there is only one underlying instrument with which to hedge and the market is therefore incomplete. Simply delta hedging will eliminate the asset price risk, but not the volatility risk as well. To complete the market we need another tradeable asset that is correlated with y_t, but the problem is that volatility is not a directly tradeable quantity. To overcome this problem we choose the second hedging instrument to

be a another derivative contract on the underlying with the same structure as the primary contract, but with a different expiry.

The idea behind deriving the pricing equation is to construct a portfolio of long the primary contract and short a number of both the underlying and the secondary contracts. With the addition of the secondary contract we now possess an instrument that is correlated with the volatility process and therefore should be able to hedge away the risk due to the fluctuations in that process.

As with the Black-Scholes analysis, we choose the number of underlying and secondary contracts in such a way that the risk due to both the asset price movements and the volatility are hedged away; the portfolio will then be riskless and, by no arbitrage, its return must therefore be equal to that of the risk-free rate.

Let $V(t,S,y)$ and $\hat{V}(t,S,y)$ be the values of the primary and secondary derivatives, respectively. If the portfolio consists of long 1 primary derivative, short Δ units of the secondary derivative, and short H units of the underlying, then its value, $\Pi(t,S,y)$ can be written

$$\Pi(t,S,y) = V(t,S,y) - \Delta\hat{V}(t,S,y) - HS, \tag{5.8}$$

and the change in the value of this portfolio over an infinitesimal timestep, dt, is

$$d\Pi(t,S,y) = dV(t,S,y) - \Delta d\hat{V}(t,S,y) - HdS. \tag{5.9}$$

Using Ito's Lemma for the change in value of the primary contract, dV, gives

$$dV = V_t dt + V_S dS + V_y dy + \frac{1}{2}V_{SS}dS^2 + \frac{1}{2}V_{yy}dy^2 + V_{Sy}dSdy + \cdots, \tag{5.10}$$

where $V_t = \partial V/\partial t$, $V_{SS} = \partial^2 V/\partial S^2$, *etc.*, and we have used the relations

$$dS^2 \sim f^2(y)S^2 dt, \tag{5.11}$$

$$dy^2 \sim \beta^2 dt, \tag{5.12}$$

$$dSdy \sim \rho\beta f(y)Sdt, \tag{5.13}$$

for infinitesimal dt and explicit dependencies on t, S, and y have been dropped for simplicity. After a similar application of Ito's Lemma on $d\hat{V}$ and substitution of the results into (5.9)

we are left with

$$d\Pi = (V_t - \Delta\hat{V}_t)dt + (V_S - \Delta\hat{V}_S - H)dS + (V_y - \Delta\hat{V}_y)dy + \frac{1}{2}(V_{SS} - \Delta\hat{V}_{SS})dS^2 + \frac{1}{2}(V_{yy} - \Delta\hat{V}_{yy})dy^2 + (V_{Sy} - \Delta\hat{V}_{Sy})dSdy + \cdots . \quad (5.14)$$

We now choose

$$\Delta = \frac{V_y}{\hat{V}_y}, \qquad \text{and} \qquad H = \frac{V_S\hat{V}_y - V_y\hat{V}_S}{\hat{V}_y}, \quad (5.15)$$

to eliminate the dS and dy terms from equation (5.14). Enforcing $d\Pi = r\Pi dt$ (*i.e.* Π must grow at the risk-free rate) gives

$$V_t - \frac{V_y}{\hat{V}_y}\hat{V}_t + \frac{1}{2}f^2(y)S^2\left(V_{SS} - \frac{V_y}{\hat{V}_y}\hat{V}_{SS}\right) + \frac{1}{2}\beta^2\left(V_{yy} - \frac{V_y}{\hat{V}_y}\hat{V}_{yy}\right) + \rho\beta f(y)S\left(V_{Sy} - \frac{V_y}{\hat{V}_y}\hat{V}_{Sy}\right) = rV - r\frac{V_y}{\hat{V}_y}\hat{V} - rS\left(\frac{V_S\hat{V}_y - V_y\hat{V}_S}{\hat{V}_y}\right), \quad (5.16)$$

which, upon rearranging, leaves us with

$$\frac{V_t + rSV_S + \frac{1}{2}f^2(y)S^2V_{SS} - rV + \rho\beta f(y)SV_{Sy} + \frac{1}{2}\beta^2V_{yy}}{V_y} = \frac{\hat{V}_t + rS\hat{V}_S + \frac{1}{2}f^2(y)S^2\hat{V}_{SS} - r\hat{V} + \rho\beta f(y)S\hat{V}_{Sy} + \frac{1}{2}\beta^2\hat{V}_{yy}}{\hat{V}_y} \quad (5.17)$$

We now note that the left side of (5.17) depends only on V while the right side depends only on $\hat{V}$. Since V and $\hat{V}$ have identical structures except for their expiry dates, each side of (5.17) must be separately equal to some function, ξ, that is independent only of T. For a reason that will become clear in the next section we will choose ξ to be

$$\xi(S, y) = a(m - y) - \beta\Lambda_\sigma(t, S, y), \quad (5.18)$$

where $\Lambda_\sigma(t, S, y)$ is the market price of volatility risk for the asset. The price of a generic vanilla contract with value $V(t, S, y)$ on an asset which is generated by the process (5.2)

must therefore satisfy the equation

$$V_t + rSV_S - rV + (a(m-y) - \beta\Lambda_\sigma(t,S,y))\,V_y + \frac{1}{2}f^2(y)S^2V_{SS} + \rho\beta f(y)SV_{Sy} + \frac{1}{2}\beta^2 V_{yy} = 0, \quad (5.19)$$

with the appropriate boundary conditions on t, S, and y.

At the beginning of this section it was noted that our model of an asset in a market with stochastic volatility is incomplete; while there exist two sources of risk, there is only one underlying instrument with which to hedge. One way of defining a complete market is one in which there exists a *unique* equivalent martingale measure for the underlying stochastic process. It is the uniqueness of the equivalent martingale measure that allows us to define a unique price for a derivative in a complete market model such as the (constant volatility) Black-Scholes model. In an incomplete market, however, while we can still find an equivalent martingale measure, it will not be unique. The presence of the function, Λ_σ, in equation (5.19) is a direct consequence of this incompleteness; it effectively acts as a parameterisation of the continuum of equivalent martingale measures.

In addition to parameterising the equivalent martingale measures, the quantity Λ_σ has an important physical interpretation: It is the market price of volatility risk; that is, the premium that the holder of a contract should expect for holding a contract with the additional volatility risk. Because of its role, Λ_σ is the one quantity in (5.19) not directly observable in the market and it therefore presents a potential problem in using this model to generate real prices. As we will discover below, one of the great advantages of the asymptotic analysis of FPS that will be presented in the next section is that calibrating Λ_σ directly is not necessary to price a contract and so this issue will not prove problematic.

5.5 Asymptotic Analysis of the Pricing Equation

Instead of solving (5.19) directly we will exploit the fact that the timescale for mean-reversion of the volatility process is much smaller than that for the lifetime of a typical derivative contract - *i.e.* $1/a \ll T$. To accomplish this we let ϵ be a small parameter,

redefine a such that

$$a = \frac{1}{\epsilon}, \tag{5.20}$$

and then examine the pricing model in the limit $\epsilon \to 0$.

Recalling the invariant distribution for y, we note that $\phi(y) \sim \mathcal{N}(m, \nu^2)$, where $\nu^2 = \beta^2/2a$. Now, physically, the variance of the distribution should be independent of the rate of mean-reversion; requiring $\nu^2 = O(1)$ implies $\beta = O(\frac{1}{\sqrt{\epsilon}})$, or, more specifically

$$\beta = \frac{\sqrt{2}\nu}{\sqrt{\epsilon}}. \tag{5.21}$$

Rescaling the pricing equation (5.19) with (5.20)-(5.21) gives

$$V_t + rSV_S - rV + \frac{1}{\epsilon}(m-y)V_y - \frac{1}{\sqrt{\epsilon}}\sqrt{2}\nu\Lambda_\sigma(t,S,y)V_y \\ + \frac{1}{2}f^2(y)S^2V_{SS} + \frac{1}{\sqrt{\epsilon}}\sqrt{2}\nu\rho f(y)SV_{Sy} + \frac{1}{\epsilon}\nu^2 V_{yy} = 0, \tag{5.22}$$

If we define the operators

$$L_0 \equiv \nu^2\frac{\partial^2}{\partial y^2} + (m-y)\frac{\partial}{\partial y}, \tag{5.23}$$

$$L_1 \equiv \sqrt{2}\nu\left(\rho f(y)\frac{\partial^2}{\partial S\partial y} - \Lambda_\sigma(t,S,y)\frac{\partial}{\partial y}\right), \quad \text{and} \tag{5.24}$$

$$L_{\mathrm{BS}}(f(y)) \equiv \frac{\partial}{\partial t} + rS\frac{\partial}{\partial S} + \frac{1}{2}f^2(y)S^2\frac{\partial^2}{\partial S^2} - r\cdot, \tag{5.25}$$

(where $L_{\mathrm{BS}}(x)$ is the Black-Scholes operator with volatility x) and multiply the rescaled equation through by ϵ we obtain

$$L_0[V] + \sqrt{\epsilon}L_1[V] + \epsilon L_{\mathrm{BS}}(f(y))[V] = 0. \tag{5.26}$$

To solve (5.26) we first assume the solution can be expressed as an expansion in powers of $\sqrt{\epsilon}$; that is

$$V(t,S,y) \sim V_0(t,S,y) + \sqrt{\epsilon}V_1(t,S,y) + \epsilon V_2(t,S,y) + \epsilon^{\frac{3}{2}}V_3(t,S,y) + \cdots. \tag{5.27}$$

After substituting (5.27) into (5.26) we obtain the following series of problems:

$$\begin{aligned} O(1): &\quad L_0[V_0] = 0, && (5.28)\\ O(\epsilon^{\frac{1}{2}}): &\quad L_0[V_1] + L_1[V_0] = 0, && (5.29)\\ O(\epsilon): &\quad L_0[V_2] + L_1[V_1] + L_{\mathrm{BS}}(f(y))[V_0] = 0, && (5.30)\\ O(\epsilon^{\frac{3}{2}}): &\quad L_0[V_3] + L_1[V_2] + L_{\mathrm{BS}}(f(y))[V_1] = 0, && (5.31)\\ &\quad \vdots \end{aligned}$$

The $O(1)$ Equation Since L_0 is exactly the infinitesimal generator of y, $L_0[V_0] = 0$ simply implies that V_0 is independent of y; that is

$$V_0 = V_0(t, S), \tag{5.32}$$

but we cannot determine the specific form of V_0 at this stage.

The $O(\epsilon^{\frac{1}{2}})$ Equation The operator L_1 takes derivatives only *w.r.t.* y. Since V_0 is independent of y we have $L_1[V_0] = 0$ and so (5.29) reduces to $L_0[V_1] = 0$. By the same reasoning as for V_0, V_1 must also be independent of y; that is

$$V_1 = V_1(t, S). \tag{5.33}$$

The $O(\epsilon)$ Equation Since V_1 is independent of y equation (5.30) reduces to

$$L_0[V_2] + L_{\mathrm{BS}}(f(y))[V_0] = 0. \tag{5.34}$$

At this point we digress to derive a solvability condition for Poisson equations like (5.34) as this will be needed throughout this work. Let $u(y)$, $v(y) \in \mathcal{C}^2$ and $G(y) \in \mathcal{L}^2$; we begin with the more general form of (5.34),

$$L_0[u(y)] = G(y). \tag{5.35}$$

The Fredholm Alternative states that the solution, u, to (5.35) exists if and only

$$\int_{-\infty}^{\infty} v(y)G(y)dy = 0 \tag{5.36}$$

for every $v(y)$ that satisfies the adjoint equation

$$L_0^*[v(y)] = 0, \tag{5.37}$$

with the appropriate adjoint boundary conditions. For the operator, $L_0[\cdot]$ the adjoint operator is defined to be

$$L_0^*[\cdot] \equiv \nu^2 \frac{\partial^2}{\partial y^2} - \frac{\partial}{\partial y}\left((m-y)\cdot\right), \tag{5.38}$$

and so $v(y)$ must satisfy

$$\nu^2 \frac{\partial^2 v}{\partial y^2} - (m-y)\frac{\partial v}{\partial y} + v = 0, \tag{5.39}$$

with boundary conditions $v(y) \to 0$ as $y \to \pm\infty$. But equation (5.39) is exactly the time-independent Kolmogorov equation, (5.5), derived above and so we have that $v(y)$ is exactly the time-invariant distribution, $\phi(y)$, of the OU process. So finally, for a solution to (5.35) to exist we require

$$\int_{-\infty}^{\infty} G(y)\phi(y)dy \equiv \langle G(y)\rangle = 0. \tag{5.40}$$

Returning to equation (5.34), the solvability condition, (5.40) requires

$$\begin{aligned}
\langle L_{\text{BS}}(f(y))[V_0]\rangle &= \langle V_{0_t} + rSV_{0_S} + \frac{1}{2}f^2(y)S^2V_{0_{SS}} - rV_0\rangle, \\
&= V_{0_t} + rSV_{0_S} + \frac{1}{2}\langle f^2(y)\rangle S^2 V_{0_{SS}} - rV_0, \\
&= V_{0_t} + rSV_{0_S} + \frac{1}{2}\bar{\sigma}^2 S^2 V_{0_{SS}} - rV_0, \\
&\equiv L_{\text{BS}}(\bar{\sigma})[V_0] = 0,
\end{aligned} \tag{5.41}$$

where $\bar{\sigma}^2 = \langle f^2(y)\rangle$. Equation (5.41) shows us that the leading-order term in V is simply the Black-Scholes value with volatility equal to the root of the expected square of the process.

To solve the $O(\epsilon^{\frac{3}{2}})$ equation below we will need an expression for V_2; to obtain this we

note that since $\langle L_{\text{BS}}(f(y))[V_0]\rangle = 0$ we can write

$$L_0[V_2] = L_0[V_2] + \langle L_{\text{BS}}(f(y))[V_0]\rangle. \tag{5.42}$$

Substituting (5.34) into (5.42) then gives

$$\begin{aligned} L_0[V_2] &= \langle L_{\text{BS}}(f(y))[V_0]\rangle - L_{\text{BS}}(f(y))[V_0], \\ &= \frac{1}{2}\left(\langle f^2(y)\rangle - f^2(y)\right) S^2 V_{0_{SS}}, \end{aligned} \tag{5.43}$$

which has the solution

$$V_2(t,S,y) = \frac{1}{2}\left(\psi(y) + c(t,S)\right) S^2 V_{0_{SS}}, \tag{5.44}$$

where c is independent of y and $\psi(y)$ solves the equation

$$\nu^2 \frac{\partial^2 \psi}{\partial y^2} + (m-y)\frac{\partial \psi}{\partial y} = \langle f^2(y)\rangle - f^2(y). \tag{5.45}$$

The $O(\epsilon^{\frac{3}{2}})$ Equation Equation (5.31) is again a Poisson equation, but now for V_3 in terms of y; for there to exist a solution we require the solvability condition

$$\langle L_1[V_2] + L_{\text{BS}}(f(y))[V_1]\rangle = 0. \tag{5.46}$$

Using (5.44) we see that

$$\begin{aligned} \langle L_1[V_2] + L_{\text{BS}}(f(y))[V_1]\rangle &= \langle L_1\left[\frac{1}{2}\left(\psi(y) + c(t,S)\right) S^2 V_{0_{SS}}\right] + L_{\text{BS}}(f(y))[V_1]\rangle, \\ &= \frac{1}{2}\langle L_1\left[\psi(y) S^2 V_{0_{SS}}\right]\rangle + L_{\text{BS}}(\bar{\sigma})[V_1], \end{aligned} \tag{5.47}$$

since c is independent of y and L_1 only take derivatives $w.r.t\ y$. Expanding $L_1[\psi(y)S^2V_{0_{SS}}]$ from (5.47) using equation (5.24) gives

$$\langle L_1\left[\psi(y)S^2V_{0_{SS}}\right]\rangle = \langle \sqrt{2}\nu\rho S f(y)\frac{\partial^2}{\partial S \partial y}(\psi(y)S^2V_{0_{SS}}) - \sqrt{2}\nu S^2 V_{0_{SS}}\Lambda_\sigma(t,S,y)\frac{\partial \psi}{\partial y}\rangle. \tag{5.48}$$

We can now integrate by parts to simplify each of the two terms on the right side of (5.48). Firstly,

$$\begin{aligned}\langle f(y)\frac{\partial^2}{\partial S\partial y}(\psi(y)S^2V_{0_{SS}})\rangle &= \frac{\partial}{\partial S}\left(S^2V_{0_{SS}}\right)\langle f(y)\frac{\partial\psi}{\partial y}\rangle,\\ &= \frac{\partial}{\partial S}(S^2V_{0_{SS}})\int f(y)\frac{\partial\psi}{\partial y}\phi(y)dy,\\ &= -\frac{\partial}{\partial S}(S^2V_{0_{SS}})\int f'(y)\psi(y)\phi(y)dy,\\ &= -\frac{\partial}{\partial S}(S^2V_{0_{SS}})\langle f'(y)\psi(y)\rangle, \qquad (5.49)\end{aligned}$$

where $'$ denotes a derivative taken $w.r.t.\ y$. Carrying out a similar calculation on the second term gives

$$\langle\Lambda_\sigma(t,S,y)\psi'(y)\rangle = -\langle\Lambda'_\sigma(t,S,y)\psi(y)\rangle. \qquad (5.50)$$

Substituting (5.49) and (5.50) back into (5.48) gives

$$\langle L_1[\psi(y)S^2V_{0_{SS}}]\rangle = \sqrt{2}\nu\langle\Lambda'_\sigma(t,S,y)\psi(y)\rangle S^2V_{0_{SS}} - \sqrt{2}\nu\rho S\frac{\partial}{\partial S}(S^2V_{0_{SS}})\langle f'(y)\psi(y)\rangle. \qquad (5.51)$$

Finally, after expanding (5.51) and combining with (5.47) we obtain the result

$$\begin{aligned}L_{\mathrm{BS}}(\bar\sigma)[V_1] = \sqrt{2}\nu\left(\rho\langle f'(y)\psi(y)\rangle - \frac{1}{2}\langle\Lambda'_\sigma(t,S,y)\psi(y)\rangle\right)S^2V_{0_{SS}}\\ + \frac{\sqrt{2}}{2}\nu\rho\langle f'(y)\psi(y)\rangle S^3V_{0_{SSS}}. \qquad (5.52)\end{aligned}$$

For convenience we now define the two parameters

$$k_1 = \sqrt{2}\nu\left(\rho\langle f'(y)\psi(y)\rangle - \frac{1}{2}\langle\Lambda'_\sigma(t,S,y)\psi(y)\rangle\right), \quad \text{and} \qquad (5.53)$$

$$k_2 = \frac{\sqrt{2}}{2}\nu\rho\langle f'(y)\psi(y)\rangle, \qquad (5.54)$$

so that equation (5.52) can be written

$$L_{\mathrm{BS}}(\bar\sigma)[V_1] = k_1S^2V_{0_{SS}} + k_2S^3V_{0_{SSS}}. \qquad (5.55)$$

So far no mention has been made about the boundary conditions to be imposed when solving any of the equations that we have encountered. Assuming that the boundary conditions are a-independent they will be $O(1)$ and thus will be satisfied only by the $O(1)$ solution, V_0; for all corrections to V_0, the boundary conditions satisfied will therefore simply be the zero conditions.

There is a simple solution to (5.55) that satisfies the zero boundary conditions; it is

$$V_1(t,S) = -(T-t)\left(k_1 S^2 V_{0_{SS}} + k_2 S^3 V_{0_{SSS}}\right), \tag{5.56}$$

which is obtained using the relations

$$L_{\mathrm{BS}}(\bar{\sigma})\left[-(T-t)B(t,S)\right] = B(t,S) - (T-t)L_{\mathrm{BS}}(\bar{\sigma})\left[B(t,S)\right], \tag{5.57}$$

and

$$L_{\mathrm{BS}}(\bar{\sigma})\left[S^n \frac{\partial^n V_0}{\partial S^n}\right] = S^n \frac{\partial^n}{\partial S^n} L_{\mathrm{BS}}(\bar{\sigma})\left[V_0\right]. \tag{5.58}$$

Substituting (5.41) and (5.56) into (5.27) gives the final result for the value of a generic derivative contract; in full it is

$$V(t,S,y) = V_0(t,S) - \frac{1}{\sqrt{a}}(T-t)\left(k_1 S^2 V_{0_{SS}} + k_2 S^3 V_{0_{SSS}}\right) + O(1/a), \tag{5.59}$$

where

$$L_{\mathrm{BS}}(\bar{\sigma})[V_0(t,S)] = 0.$$

There are several interesting features of equation (5.59). First, it is now clear that the leading-order component of a derivative's value in a market in which the volatility is stochastic and mean reverting is simply the Black-Scholes value of the same contract, but with a modified volatility $\bar{\sigma}$. Second, the dominant correction to this Black-Scholes value is $O(1/\sqrt{a})$ and is independent of y. Furthermore, this correction is composed of two terms; one proportional to $V_{0_{SS}}$ and the other to $V_{0_{SSS}}$. The first of these quantities is the Γ of the Black-Scholes value of the contract; this term will be strictly positive for any contract with a strictly concave payoff function and will also be largest in regions when the contract needs rehedging most frequently. The second quantity is proportional to $\partial\Gamma/\partial S$ and can be both positive and negative for a strictly concave payoff function such as that for a plain vanilla

call option. The total effect of the correction to the Black-Scholes value can therefore be positive or negative depending on the specifics of the contract and the value of S.

5.6 Calibration of the Model

In (5.59) we have an expression for the value of a generic derivative accurate to $O(1/\sqrt{a})$. To use equation (5.59) we must calibrate the parameters k_1 and k_2; there are two possible ways of carrying-out this calibration. The most obvious method is to simply obtain statistical estimates for each of the constituent parameters that make up k_1 and k_2, but there are two problems with this direct approach. The first of these is the fact that two of the parameters will be very difficult to measure; specifically, $f(y)$ and $\Lambda_\sigma(y)$ are not just constants, but entire functions of y and their measurement will not be trivial. Second, the market price of volatility risk, $\Lambda_\sigma(y)$, is contract dependent; not only must $\Lambda_\sigma(y)$ be calibrated from derivatives data, but it will depend on the specific contract being priced and therefore need to be recalibrated for each contract priced.

Comparing it to the exact equation (5.19), there are two aspects of the approximate pricing equation (5.59) that will greatly simplify the calibration procedure. Firstly, in (5.59) we notice that the only three independent parameters in the problem are k_1, k_2, and $\bar{\sigma}$. The seven original parameters in the exact pricing model are not independent to $O(1/\sqrt{a})$; they instead occur as three independent *universal group parameters* and these are all we should need to calibrate. Second, each of these parameters are constants. The y-dependence of f and Λ_σ in the exact model does not take effect to $O(1/\sqrt{a})$; all that matters is the average value of these quantities and so the y-dependent calibration mentioned above will not be necessary.

If we use the concept of implied volatility the calibration procedure can exploit the above aspects of the approximate solution. Recalling that the implied volatility is the value of the volatility parameter that when used in the Black-Scholes formula gives the observed value for a derivative; if $V_{\text{obs}}(t, S, \bar{\sigma})$ is the observed market value of a derivative on an asset with current volatility $\bar{\sigma}$ and $V_{\text{BS}}(t, S, \bar{\sigma})$ is the Black-Scholes value of the same derivative, then the implied volatility, σ_I, is defined such that the relation

$$V_{\text{BS}}(t, S, \sigma_I) = V_{\text{obs}}(t, S, \bar{\sigma}), \tag{5.60}$$

is true.

We now make the assumption that σ_I is very close to the average value of the volatility, $\bar{\sigma}$, so that we can write

$$\sigma_I = \bar{\sigma} + \sqrt{\epsilon}\sigma_1 + \cdots , \tag{5.61}$$

where σ_1 is some unknown $O(1)$ parameter. Using (5.61) we can now expand $V_{\mathrm{BS}}(t, S, \sigma_I)$ about $\bar{\sigma}$; this gives

$$V_{\mathrm{BS}}(t, S, \sigma_I) = V_{\mathrm{BS}}(t, S, \bar{\sigma}) + \sqrt{\epsilon}\sigma_1 \left.\frac{\partial V_{\mathrm{BS}}}{\partial \sigma}\right|_{\bar{\sigma}} + \cdots . \tag{5.62}$$

For the observed value of the derivative on the right side of equation (5.60) we will use our approximate formula for V given in equation (5.59). If we now substitute (5.59), (5.62), and (5.61) into (5.60) and use the fact that $V_0 \equiv V_{\mathrm{BS}}$ we get

$$\sigma_I = \bar{\sigma} - \left[\left.\frac{\partial V_{\mathrm{BS}}}{\partial \sigma}\right|_{\bar{\sigma}}\right]^{-1} \frac{T-t}{\sqrt{a}} \left(k_1 S^2 V_{\mathrm{BS}_{SS}} + k_2 S^3 V_{\mathrm{BS}_{SSS}}\right) + \cdots . \tag{5.63}$$

Now for simple contract structures we have analytical expressions for V_{BS}; for a vanilla call option with strike K this is

$$V_{\mathrm{BS}}(t, S) = SN(d_1) + Ke^{-r(T-t)}N(d_2), \tag{5.64}$$

where

$$d_1 = \frac{\log(S/K) + (r + \frac{1}{2}\bar{\sigma}^2)(T-t)}{\bar{\sigma}\sqrt{T-t}}, \qquad \text{and} \tag{5.65}$$

$$d_2 = \frac{\log(S/K) + (r - \frac{1}{2}\bar{\sigma}^2)(T-t)}{\bar{\sigma}\sqrt{T-t}}. \tag{5.66}$$

Computing the necessary partial derivatives of V_{BS} *w.r.t.* S and σ and then substituting these expressions into (5.63) gives the result

$$\sigma_I = k_3 \left(\frac{\log(K/S)}{T-t}\right) + k_4 + \cdots , \tag{5.67}$$

where

$$k_3 = -\frac{k_2}{\bar{\sigma}^3}, \qquad \text{and} \tag{5.68}$$

$$k_4 = \bar{\sigma} + \frac{k_2}{\bar{\sigma}^3}\left(r + \frac{3}{2}\bar{\sigma}^2\right) - \frac{k_1}{\bar{\sigma}}. \tag{5.69}$$

To leading order we can now see that there is a simple linear relationship between the implied volatility and $\log(K/S)/(T-t)$ (the Log-Moneyness-to-Maturity (LMMR) ratio) for the vanilla call option. The crucial observation in this method of calibration is that our expression (equation (5.59)) for the approximate value of a derivative in a stochastic volatility setting made no use of any contract-specific boundary conditions. This expression and, more importantly, the parameters k_1 and k_2 must therefore hold for *any* derivative written on the underlying. To calibrate the model for a specific underlying we can therefore simply take market data for a vanilla call option for which there is a simple relationship between σ_I and the LMMR ratio; by regressing σ_I against the LMMR ratio we will therefore obtain estimates for the parameters k_3 and k_4 (which give k_1 and k_2 through equations (5.68) and (5.69)) which will then hold for *all* contracts written on this underlying.

5.6.1 Region of Validity

We have found that the approximate pricing equation consists of a leading-order component that is the Black-Scholes price of the contract with the average volatility and a small correction given in equation (5.56). This approximation will be valid while the correction is small compared with the leading-order term, but will break down when they become a similar size. From (5.59), therefore, we can expect the approximation to fail in two regions; when $(T-t)/\sqrt{a} \approx 1$, and when $V_{0_{SS}}$ and/or $V_{0_{SSS}}$ are large. The first of these conditions simply states that we must be far enough away from expiry that the volatility process has enough time to fluctuate to ensure that $\langle f \rangle$ can be used reliably. The second condition states that the second and third derivatives of the leading-order solution *w.r.t.* S must not be too large. Effectively this restricts the model's use close to expiry near any discontinuity in the slope of the payoff function where the contract's Γ becomes very large.

5.7 Conclusions

We have demonstrated the Fouque, Papinicolaou, and Sircar asymptotic analysis to derive an approximation to the value of a plain vanilla option contingent on an asset with a rapidly mean-reverting stochastic volatility process. The main purpose of this work has been to develop the analysis using the relatively straightforward and well-known stochastic volatility problem so that we can apply it in Chapter 6 to derive an approximate pricing framework for a contract on an asset with a stochastic market depth. With this method the problem has been reduced from solving a two spatial dimension PDE to simply solving the standard one-dimensional Black-Scholes equation, but with an averaged volatility parameter. The power of the method, however, is not so much in the reduction of complexity of the pricing equation, but in the reduction in the number and complexity of the model's independent parameters and the fact that these universal parameters are independent of the specific form of the mean-reverting volatility process assumed. Furthermore, with this much simplified pricing framework, a novel method of calibrating the universal parameters using volatility smile data has been demonstrated.

Chapter 6

Derivative Pricing in a Market with Stochastic Liquidity - Part II

In this chapter we will apply the asymptotic techniques developed for the stochastic volatility problem in Chapter 5 to the problem of developing an approximate pricing framework for a derivative on an asset with a stochastic market depth; that is, a derivative on an asset possessing liquidity risk.

The chapter will proceed as follows. In Section 6.1 we will develop an extended version of the Bakstein-Howison liquidity asset price model in which the market depth is modelled as a mean-reverting stochastic variable. In Section 6.2 we will then derive the arbitrage-free price of a derivative contract contingent on this asset. In Section 6.3 we rescale the pricing equation assuming not only that the rate of mean-reversion of the market depth process is large, but also that the overall size of the market depth variable is small. Depending on the relative sizes of these two effects we will find that there exists three regions of the parameter space in which the form of the leading-order solution differs; following the asymptotic analysis demonstrated in Chapter 5 we then find the approximate solution in each of these cases in Section 6.4. In Section 6.5 we then develop the method by which the pricing framework is calibrated and then generate numerical results from the model in Sections 6.6 and 6.7. Specifically, in 6.6 we return to the problem of pricing the American forward developed in Chapter 4, but now do so in a market in which liquidity risk is present. In Section 6.7 we generate prices for a simple European call option and compare them to

those from the constant liquidity model developed in Chapter 2. Finally, in Section 6.8 we make concluding remarks.

6.1 The Stochastic Liquidity Bakstein-Howison (BH) Asset Price Model

The goal of this section is to extend the standard BH model to account for the empirical observation (discussed in Chapter 5) that an asset's liquidity is stochastic and mean-reverts with a daily frequency. As a first step, and to be consistent with the work presented in Chapters 2-4, we will consider the negligible bid-ask spread situation, $\gamma = 0$; our focus will therefore technically be on a stochastic market depth model.

The basis for the asset price model will be the model developed in Chapter 1. We imagine traders in the market as figuratively split into two groups; the first group is simply a single large and influential trader, while the second group, called the background group, consists of the rest of the traders in the market.

Trading is only permitted at discrete times and successive trading dates are separated by a time interval, δt, which is constant over the economy's horizon. To describe the asset price process it is convenient to divide the trading interval into two parts: In the first stage the background agents trade causing a diffusion of the asset price which is modelled as a multiplicative binomial process. The second stage then consists of the large trader readjusting his initial holdings in the underlying in response to the price diffusion due to the background. On physical grounds it is reasonable to expect, since it results from the action of only one trader, that the second stage is much shorter than the first; that is, we assume the first stage is $O(\delta t)$ while the second stage is $o(\delta t)$.

The depth of an asset's market is determined by how much of the asset is available to be bought or sold at a given price. In constructing the original asset price model we assumed that only the large trader's trades were able to move the market; any change in the market depth will therefore occur in the second stage of a trading interval only. The problem, though, with this formulation is that we have already assumed that the large trader's trading activity and the subsequent slippage of the price is $o(\delta t)$. If the change in λ were to be incorporated only into the second stage of the asset price process then we

could not hope to have any mean-reversion effect contribute in the the pricing formulation as $\delta t \to 0$. Instead we will model the change in the market depth as occurring over the entire trading interval, but in deriving the pricing equation, when necessary, we will assume the effect occurs entirely within the second stage.

Following the method of Section 5.3 we now let the market depth parameter be a stochastic variable (*i.e.* $\lambda = \lambda_t$) and specifically that it is some function, f, of a driving variable, y_t, which is, in turn, generated by a discrete-time approximation to an Ornstein-Uhlenbeck process; that is

$$\lambda_t = f(y_t), \tag{6.1}$$

$$\delta y_t = a(m - y_t)\delta t + \beta \delta \hat{W}_t, \tag{6.2}$$

where δt is a non-infinitesimal time interval.

The complete model for the asset price dynamics in a market with a negligible bid-ask spread and a stochastic market depth is then

$$S_{t+\delta t} = \begin{cases} u \cdot S_t e^{\lambda_t(1-\alpha)\delta H_t} & \text{with prob.} \quad p \\ d \cdot S_t e^{\lambda_t(1-\alpha)\delta H_t} & \text{with prob.} \quad 1-p, \end{cases} \tag{6.3}$$

$$\lambda_t = f(y_t), \tag{6.4}$$

$$\delta y_t = a(m - y_t)\delta t + \beta \delta \hat{W}_t. \tag{6.5}$$

6.2 Derivation of the Stochastic Market Depth Pricing Equation

We begin by constructing the standard portfolio of long one derivative (written on n units of the underlying) of value $V(t, S, y)$ and short a number, nH, of the underlying. As in Chapter 2, let $\tilde{V}$ be the value of the contract per unit of underlying so that $V(t, S, y) = n\tilde{V}(t, S, y)$. We know from our previous work that if y is a constant then we can choose H to eliminate the risk in the portfolio as the asset price changes over a timestep. But since y is stochastic and (at least partially) uncorrelated with S there will be an additional component of risk that the choice of H cannot eliminate and, as a result, there should be

some additional premium above the risk-free return to compensate for this unhedged risk.

We start at the beginning of the trading interval indexed by t when the large trader holds the correct hedge quantity of $-nH_t$ units of the underlying. At this moment the background agents begin to trade which causes S to diffuse to $S + \delta S$ over a time interval δt. After these trades the large trader's portfolio will no longer be correctly hedged and he will need to trade $-n\delta H_t$ units in order to rehedge. This trade will cause the asset price to slip to the new level $(S + \delta S)e^{n\lambda_t(1-\alpha)\delta H_t}$ through the removal of layers from the market thereby altering the market depth and causing $y_t \to y_t + \delta y_t$.

In deriving the asset price model in Section 6.1 we wrote

$$\lambda_t = f(y_t),$$
$$\delta y_t = a(m - y)\delta t + \beta\delta\hat{W}_t,$$

implying that the change in y occurs over the entire time interval, but on physical grounds we can see that y can only change during the second stage of the interval as a result of the large trader's trading. When writing down the new price of the asset, $(S_t+\delta S_t)e^{n\lambda_t(1-\alpha)\delta H_t}$, it is important to use $\lambda_t = f(y_t)$ rather than $\lambda_{t+\delta t} = f(y_t+\delta y_t)$ since we have assumed that the large trader's trading takes place very quickly and all at the market depth λ_t before the market has had time to react. Under these conditions, the value, Π, of the above portfolio is

$$\begin{aligned}\Pi_{t+\delta t} &\equiv \Pi(t + \delta t, (S_t + \delta S_t)e^{n\lambda_t(1-\alpha)\delta H_t}, y_t + \delta y_t) \\ &= V(t + \delta t, (S_t + \delta S_t)e^{n\lambda_t(1-\alpha)\delta H_t}, y_t + \delta y_t) - n(H_t + \delta H_t)(S_t + \delta S_t)e^{n\lambda_t(1-\alpha)\delta H_t}.\end{aligned} \tag{6.6}$$

Using the definition

$$\bar{f}(y_t) = nf(y_t), \tag{6.7}$$

and dividing (6.6) through by n gives the result

$$\Pi_{t+\delta t}/n = \tilde{V}(t + \delta t, (S_t + \delta S_t)e^{\bar{f}(y_t)(1-\alpha)\delta H_t}, y_t + \delta y_t) - (H_t + \delta H_t)(S_t + \delta S_t)e^{\bar{f}(y_t)(1-\alpha)\delta H_t}. \tag{6.8}$$

The trade of $n\delta H_t$ assets will incur a cost of

$$n\delta H_t(S_t + \delta S_t)e^{\bar{f}(y_t)\delta H_t}.$$

If y were constant we would expect the new value of the portfolio to be equal to the equivalent cash value of the portfolio, $\Pi(t, S_t, y_t)$, invested at the risk-free rate less the cost of rehedging the portfolio; mathematically

$$\Pi_{t+\delta t}/n = r\Pi(t, S_t, y_t)\delta t/n - \delta H_t(S_t + \delta S_t)e^{\bar{f}(y_t)\delta H_t}. \tag{6.9}$$

But since y is stochastic we should expect a positive premium on $\Pi_{t+\delta t} - r\Pi(t, S_t, y_t)\delta t$. This premium will have a deterministic component, $\beta\Lambda_\lambda(y)\tilde{V}_y$, proportional to the market price of market depth risk, $\Lambda_\lambda(y)$, as well as a component proportional to $\delta\hat{W}_t$ to account for the market depth risk which we will write $\beta\tilde{V}_y$ in an analogous way to Chapter 5. Substituting (6.6) into (6.9) and including the terms that account for the extra market depth risk, the final relation for the portfolio's dynamics over the timestep is

$$\begin{aligned}
\tilde{V}(t + \delta t, (S_t + \delta S_t)e^{\bar{f}(y_t)(1-\alpha)\delta H_t}, y_t + \delta y_t) - (H_t + \delta H_t)(S_t + \delta S_t)e^{\bar{f}(y_t)(1-\alpha)\delta H_t} \\
, -r\tilde{V}(t, S_t, y_t)\delta t + rH_tS_t\delta t + \delta H_t(S_t + \delta S_t)e^{\bar{f}(y_t)\delta H_t}, \\
= \beta\tilde{V}_y\delta\hat{W}_t + \beta\Lambda_\lambda\tilde{V}_y\delta t \\
= \tilde{V}_y\delta y_t - \left(a(m - y_t) - \beta\Lambda_\lambda\right)\tilde{V}_y\delta t
\end{aligned} \tag{6.10}$$

where the second equality sign results from expressing $\beta\tilde{V}_y\delta\hat{W}_t$ in terms of δt and δy_t using relation (6.2).

We now expand $\tilde{V}$ in a Taylor series about the point (t, S, y) (where the subscript t has been dropped for convenience) so that

$$\begin{aligned}
\tilde{V}(t + \delta t, (S + \delta S)e^{\bar{f}(y)(1-\alpha)\delta H}, y + \delta y) = \\
\tilde{V} + \tilde{V}_t\delta t + \tilde{V}_S\left((S + \delta S)e^{\bar{f}(y)(1-\alpha)\delta H} - S\right) + \tilde{V}_y\delta y + \frac{1}{2}\tilde{V}_{yy}\delta y^2 \\
+ \frac{1}{2}\tilde{V}_{SS}\left((S + \delta S)e^{\bar{f}(y)(1-\alpha)\delta H} - S\right)^2 + \tilde{V}_{Sy}\left((S + \delta S)e^{\bar{f}(y)(1-\alpha)\delta H} - S\right)\delta y.
\end{aligned} \tag{6.11}$$

Furthermore we make the assumption (which we will show to be valid below) that $\delta H \ll 1$

so that we can write

$$e^{\bar{f}(y)(1-\alpha)\delta H} = 1 + \bar{f}(y)(1-\alpha)\delta H + \frac{1}{2}\bar{f}^2(y)(1-\alpha)^2\delta H^2 + \cdots .$$

Expanding the exponential terms in both (6.10) and (6.11) and collecting terms reduces equation (6.10) to

$$\begin{aligned}\left(\tilde{V}_t - r\tilde{V} + rSH + (a(m-y) - \beta\Lambda_\lambda)\,\tilde{V}_y\right)\delta t + \left(\tilde{V}_S - H\right)\delta S \\ + \bar{f}(y)(1-\alpha)S(\tilde{V}_S - H)\delta H + \frac{1}{2}\tilde{V}_{SS}\delta S^2 + \bar{f}(y)(1-\alpha)(\tilde{V}_S - H + S\tilde{V}_{SS})\delta S\delta H \\ + \tilde{V}_{Sy}\delta S\delta y + \frac{1}{2}\tilde{V}_{yy}\delta y^2 + \bar{f}(y)(1-\alpha)S\tilde{V}_{Sy}\delta H\delta y \\ + \left(\frac{1}{2}\bar{f}^2(y)(1-\alpha)^2 S(\tilde{V}_S - H + S\tilde{V}_{SS}) + \alpha\bar{f}(y)S\right)\delta H^2 = o(\delta t). \quad (6.12)\end{aligned}$$

We note that the leading-order component of the market depth risk appears in the valuation as $\tilde{V}_y\delta y$ in equation (6.11). The reason for our choice for the form of the stochastic component of the market depth premium, $\beta\tilde{V}_y\delta\hat{W}$, is now clear; it was made specifically to eliminate the leading-order component of the market depth risk. We are now left only with a first order random term proportional to δS which can be eliminated by the usual choice for the hedging quantity

$$H = \tilde{V}_S. \quad (6.13)$$

Furthermore, since $\tilde{V}$ is now function of t, S, and y, the change in the hedge quantity, δH over a timestep is given by

$$\delta H = \tilde{V}_{SS}\delta S + \tilde{V}_{Sy}\delta y + O(\delta t). \quad (6.14)$$

Substituting equations (6.13) and (6.14) into (6.12), letting $\delta t \to dt$, the infinitesimal time interval, and using the continuous-time limit relations

$$dS^2 \sim \sigma^2 S^2 dt, \quad (6.15)$$

$$dy^2 \sim \beta^2 dt, \quad \text{and} \quad (6.16)$$

$$dS\delta y \sim \rho\beta\sigma S dt, \quad (6.17)$$

we find that the value, $\tilde{V}$, of a derivative in this model must satisfy the partial differential equation

$$\begin{aligned}\tilde{V}_t + rS\tilde{V}_S + \frac{1}{2}\sigma^2 S^2 \tilde{V}_{SS} - r\tilde{V} + (a(m-y) - \beta\Lambda_\lambda)\,\tilde{V}_y + \frac{1}{2}\beta^2 \tilde{V}_{yy} + \rho\beta\sigma S\tilde{V}_{Sy} \\ + \bar{f}(y)\sigma^2 S^3 \tilde{V}_{SS}^2 + \frac{1}{2}(1-\alpha)^2\sigma^2 \bar{f}^2(y) S^4 \tilde{V}_{SS}^3 + 2\rho\beta\sigma\bar{f}(y) S^2 \tilde{V}_{SS}\tilde{V}_{Sy} + \beta^2 \bar{f}(y) S\tilde{V}_{Sy}^2 \\ + \rho\beta\sigma(1-\alpha)^2 \bar{f}^2(y) S^3 \tilde{V}_{SS}^2 \tilde{V}_{Sy} + \frac{1}{2}\beta^2(1-\alpha)^2 \bar{f}^2(y) S^2 \tilde{V}_{SS}\tilde{V}_{Sy}^2 = 0. \quad (6.18)\end{aligned}$$

6.3 Rescaling the Stochastic Market Depth Pricing Equation

We now use two of the properties of the process generating the market depth to simplify the analysis of the pricing equation. First, like in the stochastic volatility example, we have $1/a \ll T$ (*i.e.* the market depth mean reverts with a much shorter period than the duration of the contract itself). Second, unlike with the stochastic volatility example where the volatility itself is an $O(1)$ process, the market depth in the present case is a small variable (*i.e.* $|\bar{f}(y_t)| \ll 1$ for all t).

Towards this end we let ϵ be an arbitrary small parameter, B be an $O(1)$ constant, and c and d be non-negative $O(1)$ parameters. To exploit the above two empirical facts about the market depth process, we utilize the rescaling

$$\begin{aligned} a &= \frac{1}{\epsilon^c}, \\ \beta &= \frac{1}{\epsilon^{\frac{c}{2}}} B, \quad \text{and} \\ \bar{f}(y) &= \epsilon^d F(y) \end{aligned} \quad (6.19)$$

where $y = O(1)$. Our goal is now to rescale equation (6.18) and determine the behaviour of the system over the whole of the positive quarter-plane of the (c, d) parameter space.

Substituting (6.19) into (6.18) and multiplying through by ϵ^c gives the rescaled pricing

equation

$$
\begin{aligned}
L_0[\tilde{V}] + \epsilon^{\frac{c}{2}} L_1[\tilde{V}] + \epsilon^c L_{\mathrm{BS}}[\tilde{V}] \\
+ \epsilon^d \cdot B^2 F(y) S \tilde{V}_{Sy}^2 + \epsilon^{2d} \cdot \frac{1}{2} B^2 (1-\alpha)^2 F^2(y) S^2 \tilde{V}_{SS} \tilde{V}_{Sy}^2 \\
+ \epsilon^{\frac{c}{2}+d} \cdot 2\rho B \sigma F(y) S^2 \tilde{V}_{SS} \tilde{V}_{Sy} + \epsilon^{\frac{c}{2}+2d} \cdot \rho B \sigma (1-\alpha)^2 F^2(y) S^3 \tilde{V}_{SS}^2 \tilde{V}_{Sy} \\
+ \epsilon^{c+d} \cdot \sigma^2 F(y) S^3 \tilde{V}_{SS}^2 + \epsilon^{c+2d} \cdot \frac{1}{2} \sigma^2 (1-\alpha)^2 F^2(y) S^4 \tilde{V}_{SS}^3 = 0, \quad (6.20)
\end{aligned}
$$

where

$$
\begin{aligned}
L_0 &\equiv \frac{1}{2} B^2 \frac{\partial^2}{\partial y^2} + (m-y) \frac{\partial}{\partial y}, \\
L_1 &\equiv B \left(\rho \bar{f}(y) \frac{\partial^2}{\partial S \partial y} - \Lambda_\lambda(t,S,y) \frac{\partial}{\partial y} \right), \quad \text{and} \\
L_{\mathrm{BS}} &\equiv \frac{\partial}{\partial t} + rS \frac{\partial}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2}{\partial S^2} - r \cdot .
\end{aligned}
$$

6.4 Analysis of the Rescaled Pricing Equation

Equation (6.20) has terms of many different orders. Depending on the relative size of c and d, the relative importance of these terms will change. For this work we will only focus on the situation when both $a \gg 1$ *and* $|\bar{f}(y)| \ll 1$; when both $c > 0$ and $d > 0$ it is clear that two or more of the terms change relative magnitude at three points; when $c = d$, when $c = 2d$, and when $c = 4d$. Including these three boundaries there are seven distinct regions of the allowable (c, d) space where the behaviour of the system is different. For concreteness these seven regions are shown in Figure 6.1; they are (in (c, d) notation) $(c < d, d)$, $(c = d, d)$, $(d < c < 2d, d)$, $(c = 2d, d)$, $(2d < c < 4d, d)$, $(c = 4d, d)$, and $(c > 4d, d)$. We will now analyze each of these cases.

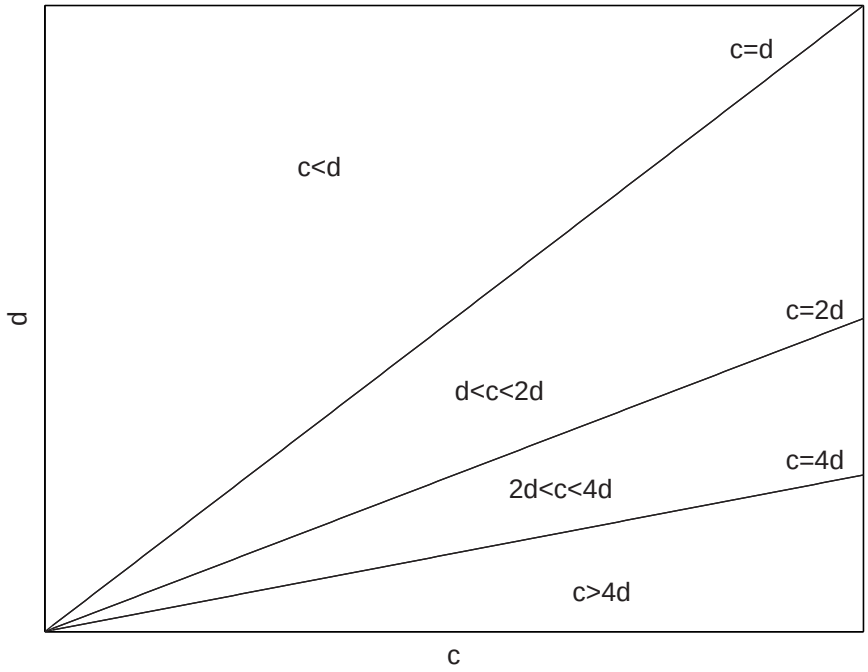


Figure 6.1: The seven regions of the (c, d) space over which the form of the rescaled pricing equation (6.20) varies.

6.4.1 The Case $(c = d, d)$

To treat this case we will use the concrete example of $(c, d) = (2, 2)$; with these values (6.20) becomes

$$\begin{aligned} L_0[\tilde{V}] + \epsilon L_1[\tilde{V}] + \epsilon^2 \left(L_{\mathrm{BS}}[\tilde{V}] + B^2 F(y) S \tilde{V}_{Sy}^2 \right) + \epsilon^3 \cdot 2\rho B\sigma F(y) S^2 \tilde{V}_{SS} \tilde{V}_{Sy} \\ + \epsilon^4 \left(\frac{1}{2} B^2 (1-\alpha)^2 F^2(y) S^2 \tilde{V}_{SS} \tilde{V}_{Sy}^2 + \sigma^2 F(y) S^3 \tilde{V}_{SS}^2 \right) \\ + \epsilon^5 \cdot \rho B\sigma (1-\alpha)^2 F^2(y) S^3 \tilde{V}_{SS}^2 \tilde{V}_{Sy} + \epsilon^6 \cdot \frac{1}{2} \sigma^2 (1-\alpha)^2 F^2(y) S^4 \tilde{V}_{SS}^3 = 0. \end{aligned} \quad (6.21)$$

We now assume that $\tilde{V}$ can be written in the form

$$\tilde{V} \sim h_{1_0} + \epsilon h_{1_1} + \epsilon^2 h_{1_2} + \epsilon^3 h_{1_3} + \cdots . \quad (6.22)$$

After substituting (6.22) into (6.21) and grouping terms of like orders of ϵ we are left with a series of problems for $\tilde{V}$.

$\boldsymbol{O}(1)$: The leading-order component of the contract's value satisfies the equation

$$L_0[h_{1_0}] = 0, \tag{6.23}$$

but since L_0 is exactly the infinitesimal generator of our rescaled OU process, equation (6.23) implies

$$h_{1_0} = h_{1_0}(t, S), \tag{6.24}$$

or, in words, that h_{1_0} is y-independent.

$\boldsymbol{O}(\epsilon)$: The $O(\epsilon)$ equation is

$$L_0[h_{1_1}] + L_1[h_{1_0}] = 0. \tag{6.25}$$

Since $h_{1_0} = h_{1_0}(t, S)$, $L_1[h_{1_0}] = 0$ and so $h_{1_1} = h_{1_1}(t, S)$ also.

$\boldsymbol{O}(\epsilon^2)$: The $O(\epsilon^2)$ equation is

$$L_0[h_{1_2}] + L_1[h_{1_1}] + L_{\text{BS}}[h_{1_0}] + B^2 F(y) S h^2_{1_{0Sy}} = 0, \tag{6.26}$$

but because of the y-independence of both h_{1_0} and h_{1_0} (6.26) reduces to

$$L_0[h_{1_2}] + L_{\text{BS}}[h_{1_0}] = 0. \tag{6.27}$$

From our analysis of the stochastic volatility example we know that a solution to (6.27) exists only if the nonhomogeneous term, $L_{\text{BS}}[h_{1_0}]$, is centered *w.r.t.* the long-run invariant distribution, $\phi(y)$, for the OU process given in equation (5.6); that is

$$\langle L_{\text{BS}}[h_{1_0}] \rangle = 0. \tag{6.28}$$

Since L_{BS} is independent of y, equation (6.28) reduces to

$$L_{\text{BS}}[h_{1_0}] = 0. \tag{6.29}$$

In other words, the leading-order component of the solution is *exactly* the Black-Scholes value of the contract. Furthermore, since $L_{\mathrm{BS}}[h_{1_0}] = 0$, equation (6.27) then reduces to $L_0[h_{1_2}] = 0$, implying

$$h_{1_2} = h_{1_2}(t, S). \tag{6.30}$$

What this last result tells us is that because of the y-independence of the Black-Scholes operator the y-dependence will enter into the solution at a higher order than it otherwise would. This y-dependence is therefore relatively less significant than it is in the stochastic volatility example where the Black-Scholes operator is dependent on the additional stochastic variable.

$\boldsymbol{O}(\epsilon^3)$: The $O(\epsilon^3)$ equation is

$$L_0[h_{1_3}] + L_1[h_{1_2}] + L_{\mathrm{BS}}[h_{1_1}] + 2B^2F(y)Sh_{1_{0Sy}}h_{1_{1Sy}} + 2\rho B\sigma F(y)S^2h_{1_{0SS}}h_{1_{0Sy}} = 0, \tag{6.31}$$

which reduces to

$$L_0[h_{1_3}] + L_{\mathrm{BS}}[h_{1_1}] = 0, \tag{6.32}$$

because of the y-independence of h_{1_0}, h_{1_1}, and h_{1_2}. Using the same reasoning as for the $O(\epsilon^2)$ equation above we require that $\langle L_{\mathrm{BS}}[h_{1_1}]\rangle = 0$ in order for (6.32) to have a solution; the relation satisfied by h_{1_1} is therefore

$$L_{\mathrm{BS}}[h_{1_1}] = 0. \tag{6.33}$$

But for most, if not all, contracts the boundary conditions are $O(1)$. In this case h_{1_1} must satisfy the zero conditions and because its evolution is governed by the (homogeneous) Black-Scholes equation it must therefore be exactly zero for all time; that is

$$h_{1_1}(t, S) = 0. \tag{6.34}$$

From either (6.33) of (6.34) equation (6.32) now reduces to $L_0[h_{1_3}] = 0$ giving

$$h_{1_3} = h_{1_3}(t, S). \tag{6.35}$$

$\boldsymbol{O}(\epsilon^4)$: The $O(\epsilon^4)$ equation is

$$L_0[h_{1_4}] + L_1[h_{1_3}] + L_{\text{BS}}[h_{1_2}] + B^2F(y)S\left(h_{1_{1Sy}}^2 + 2h_{1_{0Sy}}h_{1_{2Sy}}\right) + 2\rho B\sigma F(y)S^2\left(h_{1_{0SS}}h_{1_{1Sy}} + h_{1_{1SS}}h_{1_{0Sy}}\right) + \frac{1}{2}B^2(1-\alpha)^2F^2(y)S^2h_{1_{0SS}}h_{1_{0Sy}}^2 = 0, \quad (6.36)$$

which immediately reduces to

$$L_0[h_{1_4}] + L_{\text{BS}}[h_{1_2}] + \sigma^2F(y)S^3h_{1_{0SS}}^2 = 0, \quad (6.37)$$

because of the y-independence of h_{1_0}, h_{1_2}, and h_{1_3}. Once again, equation (6.37) is a Poisson equation and for a solution to exist the nonhomogeneous term must be centered *w.r.t.* $\phi(y)$; that is

$$\langle L_{\text{BS}}[h_{1_2}] + \sigma^2F(y)S^3h_{1_{0SS}}^2\rangle = 0, \quad (6.38)$$

but since L_{BS} is independent of y (6.38) reduces to

$$L_{\text{BS}}[h_{1_2}] + \sigma^2\langle F(y)\rangle S^3h_{1_{0SS}}^2 = 0, \quad (6.39)$$

which is the relation that the first non-Black-Scholes correction to h_{1_0} must satisfy.

Substituting (6.39) into (6.37) gives

$$L_0[h_{1_4}] = \sigma^2S^3h_{1_{0SS}}^2\left(\langle F(y)\rangle - F(y)\right), \quad (6.40)$$

or

$$h_{1_4}(t,S,y) = L_0^{-1}\left[\sigma^2S^3h_{1_{0SS}}^2\left(\langle F(y)\rangle - F(y)\right)\right]. \quad (6.41)$$

Because L_0 is independent of both t and S we can write (6.41) in the form

$$h_{1_4}(t,S,y) = \sigma^2S^3h_{1_{0SS}}^2\left(\psi(y) + c(t,S)\right), \quad (6.42)$$

where $\psi(y)$ satisfies $L_0[\psi(y)] = \langle F(y)\rangle - F(y)$ and $c(t,S)$ is an arbitrary function of t and S.

$\boldsymbol{O}(\epsilon^5)$: We can use the result (6.42) to take the analysis one step further than the $O(\epsilon^2)$

correction entering through the $O(\epsilon^4)$ equation. The $O(\epsilon^5)$ equation is

$$\begin{aligned} L_0[h_{1_5}] + L_1[h_{1_4}] + L_{\mathrm{BS}}[h_{1_3}] + 2B^2F(y)S\left(h_{1_{0Sy}}h_{1_{3Sy}} + h_{1_{1Sy}}h_{1_{2Sy}}\right) \\ + 2\rho B\sigma F(y)S^2\left(h_{1_{0SS}}h_{1_{2Sy}} + h_{1_{1SS}}h_{1_{1Sy}} + h_{1_{2SS}}h_{1_{0Sy}}\right) \\ + \frac{1}{2}B^2(1-\alpha)^2F^2(y)S^2\left(h_{1_{1SS}}h_{1_{0Sy}}^2 + 2h_{1_{0SS}}h_{1_{0Sy}}h_{1_{1Sy}}\right) \\ + 2\sigma^2F(y)S^3h_{1_{0SS}}h_{1_{1SS}} + \rho B\sigma(1-\alpha)^2F^2(y)S^3h_{1_{0SS}}^2h_{1_{0Sy}} = 0, \end{aligned} \tag{6.43}$$

which reduces to

$$L_0[h_{1_5}] + L_1[h_{1_4}] + L_{\mathrm{BS}}[h_{1_3}] = 0, \tag{6.44}$$

due to the y-independence of h_{1_3}, h_{1_2}, and h_{1_0} and the fact that $h_{1_1} \equiv 0$. Using the usual existence condition for a solution to this Poisson equation, we can write

$$\langle L_1[h_{1_4}] + L_{\mathrm{BS}}[h_{1_3}]\rangle = 0, \tag{6.45}$$

which, after substituting (6.42) and using the fact that both L_{BS} and $c(t,S)$ are independent of y, further reduces to

$$L_{\mathrm{BS}}[h_{1_3}] + \langle L_1\left[\sigma^2S^3h_{1_{0SS}}^2\psi(y)\right]\rangle = 0. \tag{6.46}$$

Using the definition for L_1 (5.24), we can write

$$\begin{aligned} \langle L_1\left[S^3h_{1_{0SS}}^2\psi(y)\right]\rangle &= \langle \rho B\sigma S\frac{\partial^2}{\partial S\partial y}\left(S^3h_{1_{0SS}}^2\psi(y)\right) - B\Lambda_\lambda(y)\frac{\partial}{\partial y}\left(S^3h_{1_{SS}}^2\psi(y)\right)\rangle, \\ &= \left(3\rho B\sigma\ \langle\psi'(y)\rangle - B\langle\psi'(y)\Lambda_\lambda(y)\rangle\right)S^3h_{1_{0SS}}^2 + 2\rho B\sigma\langle\psi'(y)\rangle S^4h_{1_{0SS}}h_{1_{0SSS}}, \end{aligned} \tag{6.47}$$

(where we have explicitly shown the dependence of Λ_λ on y alone and where $'$ denotes a derivative taken *w.r.t.* y) which tells us the $O(\epsilon^3)$ correction to h_0^1 must satisfy

$$L_{\mathrm{BS}}[h_{1_3}] + \kappa_1S^3h_{1_{0SS}}^2 + \kappa_2S^4h_{1_{0SS}}h_{1_{0SSS}} = 0, \tag{6.48}$$

where

$$\kappa_1 = 3\rho B\sigma^3\langle\psi'(y)\rangle - B\sigma^2\langle\psi'(y)\Lambda_\lambda(y)\rangle, \tag{6.49}$$

$$\kappa_2 = 2\rho B\sigma^3\langle\psi'(y)\rangle. \tag{6.50}$$

The goal of the analysis is to determine the form of the solution (or, more precisely, the form of the equations that the solution must satisfy) up to the level at which the y-dependence enters. Since we know from (6.42) that $h_{1_4} = h_{1_4}(t, S, y)$ we will stop the analysis at this point.

The form of the solution (which we mean to be the leading-order solution plus non-Black-Scholes corrections) is determined by equations (6.29), (6.39), and (6.48) when $c = d$, but actually remains the same for the entire region $0 < c < 2d$. To see why this is so we will briefly summarize the results of the above analysis and then examine how the system changes due to changes in the relative order of terms within equation (6.20) as c varies between 0 and $2d$.

From the analysis above, the leading-order term in the expansion was determined from the $O(\epsilon^2)$ equation (or, more generally, from the $O(\epsilon^c)$ equation) when the L_{BS} operator first entered. Similarly, the first non-Black-Scholes correction was determined by the $O(\epsilon^4)$ equation ($O(\epsilon^{c+d})$ in more general terms) when the $\sigma^2F(y)S^3h^2_{1_{0SS}}$ term first enters. The leading-order term will always be $O(1)$ and the first non-BS correction will therefore be $O(\epsilon^d)$ which is the difference in orders between the equations that determine the leading-order and first correction solutions.

Now, since

$$\left\langle\sigma^2F(y)S^3h^2_{1_{0SS}}\right\rangle \neq \sigma^2F(y)S^3h^2_{1_{0SS}},$$

it was found that $L_0[h_{1_4}] \neq 0$ from the $O(\epsilon^4)$ equation (in general this first y-dependent term, $h_{1_{c+d}}$ is determined through $L_0[h_{1_{c+d}}] \neq 0$ from the $O(\epsilon^{c+d})$ equation) which, in turn, forces the second correction to be determined by the next equation in the series.

At this point it is necessary to examine the relative order of terms in equation (6.20). When $0 < c < d$:

$$\begin{aligned} O(1) > O(\epsilon^{\frac{c}{2}}) > O(\epsilon^c) > O(\epsilon^d) > O(\epsilon^{\frac{c}{2}+d}) > O(\epsilon^{c+d}) \\ > O(\epsilon^{2d}) > O(\epsilon^{\frac{c}{2}+2d}) > O(\epsilon^{c+2d}). \end{aligned} \tag{6.51}$$

When $c = d$:

$$O(1) > O(\epsilon^{\frac{c}{2}}) > \Big(O(\epsilon^c) = O(\epsilon^d)\Big) > O(\epsilon^{\frac{c}{2}+d}) \\ > \Big(O(\epsilon^{c+d}) = O(\epsilon^{2d})\Big) > O(\epsilon^{\frac{c}{2}+2d}) > O(\epsilon^{c+2d}). \quad (6.52)$$

When $d < c < 2d$:

$$O(1) > O(\epsilon^{\frac{c}{2}}) > O(\epsilon^d) > O(\epsilon^c) > O(\epsilon^{\frac{c}{2}+d}) > O(\epsilon^{2d}) \\ > O(\epsilon^{c+d}) > O(\epsilon^{\frac{c}{2}+2d}) > O(\epsilon^{c+2d}). \quad (6.53)$$

In all three of these regions the next-smallest term after the $O(\epsilon^{c+d})$ term of $\sigma^2 \bar{f}(y) S^3 \tilde{V}_{SS}^2$ is either $\frac{1}{2}\beta^2(1-\alpha)^2 \bar{f}^2(y) S^2 \tilde{V}_{SS}^2 \tilde{V}_{Sy}$ or $\rho\beta\sigma(1-\alpha)^2 \bar{f}^2(y) S^3 \tilde{V}_{SS}^2 \tilde{V}_{Sy}$. In the equation immediately following the $O(\epsilon^{c+d})$ equation these terms will first appear, but they will only act on h_{1_0}. Since h_{1_0} is independent of y neither will contribute to the second correction. In the region $0 < c < 2d$ this second correction will therefore be completely determined by h_{1_0} and so will remain constant in this range.

As c varies throughout the region $0 < c < 2d$, the solution will actually vary somewhat. If we examine (6.51)-(6.53) again we see that there are two terms separating $L_{\text{BS}}[\tilde{V}]$ and $\sigma^2 F(y) S^3 \tilde{V}_{SS}^2$ for all of the region except for $c = d$ when there is only one term. While this has no effect on the *form* of either the first or second corrections, it does affect the *order* at which these terms enter the analysis and therefore their size relative to the leading-order term.

Summary and Comments

In summary, the approximate form of the pricing problem in this region is

$$\begin{gathered}\tilde{V} \sim h_{1_0} + \epsilon h_{1_1} + \epsilon^2 h_{1_2} + \epsilon^3 h_{1_3} + \cdots, \quad \text{where} \\ L_{\mathrm{BS}}[h_{1_0}] = 0, \\ h_{1_1} = 0, \\ L_{\mathrm{BS}}[h_{1_2}] + \sigma^2 \langle F(y) \rangle S^3 h_{1_{0SS}}^2 = 0, \quad \text{and} \\ L_{\mathrm{BS}}[h_{1_3}] + \kappa_1 S^3 h_{1_{0SS}}^2 + \kappa_2 S^4 h_{1_{0SS}} h_{1_{0SSS}} = 0.\end{gathered} \tag{6.54}$$

As expected, the leading-order component of the solution is exactly the Black-Scholes value of the contract. The first correction to this value enters at $O(\epsilon^2)$ and, furthermore, it satisfies the same form of equation as the first correction in the outer region of the constant-λ problem as we found in Chapter 3, except with the long-run distributional average of the market depth, $\langle F(y) \rangle$.

The equation for the second correction of $O(\epsilon^3)$ is driven by two nonhomogeneous terms. As both of these terms contain B their presence is directly attributable to the stochastic nature of the market depth and they are therefore the lowest order corrections to the contract's value due to the liquidity risk in the underlying.

The two terms have interesting interpretations. Firstly, as the κ_1 term is driven by $h_{1_{0SS}}^2$ it behaves like a transaction cost effect. The presence of an effect of this type is not surprising as it always increases hedging costs and is largest where the contract's Γ is largest; definitely a reasonable quality for a liquidity risk term. The κ_2 term similarly contains $h_{1_{0SS}}$, but it is also driven by $h_{1_{0SSS}}$. As $h_{1_{0SS}}$ is a measure of how frequently the contract (in the Black-Scholes world) will need to be rehedged, $h_{1_{0SSS}}$ is therefore a measure of how this frequency varies with S. For instance, for a call option $h_{1_{0SSS}} > 0$ for $S < K$ (where K is the strike) and $h_{1_{0SSS}} < 0$ for $S > K$; for $S < K$ an increase in S will make it necessary to rehedge more often and the converse for $S > K$. For an asset with a finite and uncertain level of liquidity, this effect is clearly important..

6.4.2 The Case $(c = 2d, d)$

When $c = 2d$ a new effect enters and there is a significant change in the form of the solution. For this case we will choose the concrete example of $c = 2$ and $d = 1$; with this choice, equation (6.20) becomes

$$L_0[\tilde{V}] + \epsilon\left(L_1[\tilde{V}] + B^2F(y)S\tilde{V}_{Sy}^2\right) + \epsilon^2\Bigg(L_{\text{BS}}[\tilde{V}] + \frac{1}{2}B^2(1-\alpha)^2F^2(y)S^2\tilde{V}_{SS}\tilde{V}_{Sy}^2 + 2\rho B\sigma F(y)S^2\tilde{V}_{SS}\tilde{V}_{Sy}\Bigg) + \epsilon^3\left(\rho B\sigma(1-\alpha)^2F^2(y)S^3\tilde{V}_{SS}^2\tilde{V}_{Sy} + \sigma^2F(y)S^3\tilde{V}_{SS}^2\right) + \epsilon^4\cdot\frac{1}{2}\sigma^2(1-\alpha)^2F^2(y)S^4\tilde{V}_{SS}^3 = 0. \quad (6.55)$$

Now assuming $\tilde{V}$ can be written in the form

$$\tilde{V} \sim h_{2_0} + \epsilon h_{2_1} + \epsilon^2 h_{2_2} + \cdots, \quad (6.56)$$

then we have a series of problems in powers of ϵ.

$\boldsymbol{O}(1)$: The $O(1)$ problem is simply

$$L_0[h_{2_2}] = 0, \quad (6.57)$$

and since L_0 is the infinitesimal generator of the rescaled market depth process, we know that h_{2_0} must be independent of y; that is

$$h_{2_0} = h_{2_0}(t, S). \quad (6.58)$$

$\boldsymbol{O}(\epsilon)$: The $O(\epsilon)$ problem is

$$L_0[h_{2_1}] + L_1[h_{2_0}] + B^2F(y)Sh_{2_{0Sy}}^2 = 0, \quad (6.59)$$

but due to the y-independence of h_{2_0}, this immediately reduces to

$$\begin{aligned} & L_0[h_{2_1}] = 0, \\ \implies & h_{2_1} = h_{2_1}(t, S). \end{aligned} \tag{6.60}$$

$\boldsymbol{O}(\epsilon^2)$: The $O(\epsilon^2)$ equation is

$$\begin{aligned} & L_0[h_{2_2}] + L_1[h_{2_1}] + 2B^2F(y)Sh_{2_{0Sy}}h_{2_{1Sy}} + L_{\mathrm{BS}}[h_{2_0}] \\ & \quad + \frac{1}{2}B^2(1-\alpha)^2F^2(y)S^2h_{2_{0SS}}h_{2_{0Sy}}^2 + 2\rho B\sigma F(y)S^2h_{2_{0SS}}h_{2_{0Sy}} = 0. \end{aligned} \tag{6.61}$$

Again using the y-independence of both h_{2_0} and h_{2_1}, (6.61) reduces to

$$L_0[h_{2_2}] + L_{\mathrm{BS}}[h_{2_0}] = 0. \tag{6.62}$$

With the same reasoning as was used with both the case $(c < 2d, d)$ and the stochastic volatility example, a solution to the Poisson equation (6.62) can only exist if $\langle L_{\mathrm{BS}}[h_{2_0}]\rangle = 0$. Because of the y-independence of L_{BS} we therefore have

$$L_{\mathrm{BS}}[h_{2_0}] = 0, \tag{6.63}$$

or, as was the case when $(c < 2d, d)$, the leading-order component of the contract's value is the Black-Scholes value in this region of the parameter space. Substituting (6.63) into (6.62) gives $L_0[h_{2_2}] = 0$, or

$$h_{2_2} = h_{2_2}(t, S). \tag{6.64}$$

$\boldsymbol{O}(\epsilon^3)$: The $O(\epsilon^3)$ equation is

$$\begin{aligned} & L_0[h_{2_3}] + L_1[h_{2_2}] + L_{\mathrm{BS}}[h_{2_1}] + B^2F(y)S\left(h_{2_{1Sy}}^2 + 2h_{2_{0Sy}}h_{2_{2Sy}}\right) \\ & \quad + \sigma^2F(y)S^3h_{2_{0SS}}^2 + \frac{1}{2}B^2(1-\alpha)^2F^2(y)S^2\left(h_{2_{1SS}}h_{2_{0Sy}}^2 + 2h_{2_{0SS}}h_{2_{0Sy}}h_{2_{1Sy}}\right) \\ & + 2\rho B\sigma F(y)S^2\left(h_{2_{0SS}}h_{2_{1Sy}} + h_{2_{1SS}}h_{2_{0Sy}}\right) + \rho B\sigma(1-\alpha)^2F^2(y)S^3h_{2_{0SS}}^2h_{2_{0Sy}} = 0. \end{aligned} \tag{6.65}$$

Using the y-independence of h_{2_0}, h_{2_1}, and h_{2_2}, (6.65) reduces to

$$L_0[h_{2_3}] + L_{\text{BS}}[h_{2_1}] + \sigma^2 F(y) S^3 h_{2_{1SS}}^2 = 0, \tag{6.66}$$

and the usual existence condition for a solution to this Poisson equation dictates that the first correction to the solution in this region of the parameter space must satisfy

$$L_{\text{BS}}[h_{2_1}] + \sigma^2 \langle F(y) \rangle S^3 h_{2_{0SS}}^2 = 0. \tag{6.67}$$

Substituting (6.67) into (6.66) results in the relation

$$\begin{aligned} L_0[h_{2_3}(t,S,y)] &= \sigma^2 S^3 h_{2_{0SS}}^2 \left(\langle F(y) \rangle - F(y) \right), \\ \implies h_{2_3}(t,S,y) &= \sigma^2 S^3 h_{2_{0SS}}^2 \left(\psi(y) + c(t,S) \right), \end{aligned} \tag{6.68}$$

where, as before, $\psi(y)$ is defined as the solution to $L_0[\psi(y)] = \langle F(y) \rangle - F(y)$ and $c(t,S)$ is some function independent of y.

$\boldsymbol{O}(\epsilon^4)$: From (6.68) we know that the y-dependence will enter the solution, (6.56), at $O(\epsilon^3)$ through h_{2_3}. We have the relation, (6.67), that the first-order correction must satisfy, but we can go one step further and derive a relation for h_{2_2} using the $O(\epsilon^4)$ equation. This equation is

$$\begin{aligned} &L_0[h_{2_4}] + L_1[h_{2_3}] + L_{\text{BS}}[h_{2_2}] + 2B^2 F(y) S \left(h_{2_{0Sy}} h_{2_{3Sy}} + h_{2_{1Sy}} h_{2_{2Sy}} \right) \\ &\quad + \frac{1}{2} B^2 (1-\alpha)^2 F^2(y) S^2 \left(h_{2_{2SS}} h_{2_{0Sy}}^2 + 2 h_{2_{1SS}} h_{2_{0Sy}} h_{2_{1Sy}} + h_{2_{2SS}}^2 h_{2_{0Sy}}^2 \right) \\ &\quad + 2\rho B \sigma F(y) S^2 \left(h_{2_{0SS}} h_{2_{2Sy}} + h_{2_{1SS}} h_{2_{1Sy}} + h_{2_{0SS}} h_{2_{2Sy}} \right) \\ &\quad + \rho B \sigma (1-\alpha)^2 F^2(y) S^3 \left(h_{2_{0SS}}^2 h_{2_{1Sy}} + 2 h_{2_{0SS}} h_{2_{1SS}} h_{2_{0Sy}} \right) \\ &\quad + 2\sigma^2 F(y) S^3 h_{2_{0SS}} h_{2_{1SS}} + \frac{1}{2} \sigma^2 (1-\alpha)^2 F^2(y) S^4 h_{2_{0SS}}^3 = 0, \end{aligned} \tag{6.69}$$

which reduces to

$$L_0[h_{2_4}] + L_1[h_{2_3}] + L_{\mathrm{BS}}[h_{2_2}] + 2\sigma^2 F(y) S^3 h_{2_{0SS}} h_{2_{1SS}} + \frac{1}{2}\sigma^2(1-\alpha)^2 F^2(y) S^4 h_{2_{0SS}}^3 = 0, \quad (6.70)$$

once the y-independence of h_{2_0}, h_{2_1}, and h_{2_2} is taken into account. The existence condition for a solution to (6.70) states that

$$\left\langle L_{\mathrm{BS}}[h_{2_2}] + L_1[h_{2_3}] + 2\sigma^2 F(y) S^3 h_{2_{0SS}} h_{2_{1SS}} + \frac{1}{2}\sigma^2(1-\alpha)^2 F^2(y) S^4 h_{2_{0SS}}^3 \right\rangle = 0. \quad (6.71)$$

Comparing equation (6.71) with (6.45) it is now clear how $\tilde{V}$ differs in the two regions $c < 2d$ and $c = 2d$. While the form for the equations determining the first-order corrections in the two regions are identical, those for the equations determining the second-order corrections differ by the $h_{2_{0SS}}^3$ and $h_{2_{0SS}} h_{2_{1SS}}$ terms which arise when $c = 2d$. The presence of these terms is a result of the fact that the $\tilde{V}_{SS}^2$ and $\tilde{V}_{SS}^3$ terms are consecutive in equation (6.55) while they are separated by the $\tilde{V}_{SS}^2 \tilde{V}_{Sy}$ term (which, crucially takes a derivative *w.r.t.* y) in equation (6.21). Substituting (6.68) into (6.71) and using (6.47), (6.49), and (6.50) we get

$$L_{\mathrm{BS}}[h_{2_2}] + \kappa_1 S^3 h_{2_{0SS}}^2 + \kappa_2 S^4 h_{2_{0SS}} h_{2_{0SSS}} + 2\sigma^2 \langle F(y) \rangle S^3 h_{2_{0SS}} h_{2_{1SS}} + \frac{1}{2}\sigma^2(1-\alpha)^2 \left\langle F^2(y) \right\rangle S^4 h_{2_{0SS}}^3 = 0, \quad (6.72)$$

which has the same form as equation (6.48) for the second correction in the region $c < 2d$ except for the additional $h_{2_{0SS}}^3$ and $h_{2_{0SS}} h_{2_{1SS}}$ terms.

Summary and Comments

The approximate solution in this region is therefore

$$\tilde{V} \sim h_{2_0} + \epsilon h_{2_1} + \epsilon^2 h_{2_2} + \cdots, \quad \text{where}$$
$$L_{\text{BS}}[h_{2_0}] = 0,$$
$$L_{\text{BS}}[h_{2_1}] + \sigma^2 \langle F(y) \rangle S^3 h_{2_{0SS}}^2 = 0, \quad \text{and}$$
$$L_{\text{BS}}[h_{2_2}] + \kappa_1 S^3 h_{2_{0SS}}^2 + \kappa_2 S^4 h_{2_{0SS}} h_{2_{0SSS}} + 2\sigma^2 \langle F(y) \rangle S^3 h_{2_{0SS}} h_{2_{1SS}}$$
$$+ \frac{1}{2}\sigma^2 (1-\alpha)^2 \left\langle F^2(y) \right\rangle S^4 h_{2_{0SS}}^3 = 0,$$

The solution in this region of the (c, d) space differs from that when $0 < c < 2d$ in two ways. First, while the first and second corrections were $O(\epsilon^2)$ and $O(\epsilon^3)$, respectively, when $0 < c < 2d$ they are now $O(\epsilon)$ and $O(\epsilon^2)$. As the liquidity risk terms enter through the second-order correction they are effectively more important when $c = 2d$ than in the previous case. Second, we notice that in the second correction to the leading-order Black-Scholes value there is now an extra contribution due to the $h_{2_{0SS}}^3$ and $h_{2_{0SS}} h_{2_{1SS}}$ terms which are simply the standard $O(\bar{\lambda}^2)$ corrections in the outer solution of the constant-λ problem. Now that $c = 2d$ the magnitude of $\bar{\lambda}$ is increased relative to a and the introduction of these constant-liquidity term is understandable as they now have a greater relative importance to the terms originating from the stochastic nature of the market depth parameter.

6.4.3 The Case $(c > 2d, d)$

We now move onto the third interior region of the parameter space where $c > 2d$. For this case we will examine the specific example $c = 4, d = 1$; with this choice of (c, d) equation (6.20) becomes

$$L_0[\tilde{V}] + \epsilon \cdot B^2 F(y) S \tilde{V}_{Sy}^2 + \epsilon^2 \left(L_1[\tilde{V}] + \frac{1}{2} B^2 (1-\alpha)^2 F^2(y) S^2 \tilde{V}_{SS} \tilde{V}_{Sy}^2 \right)$$
$$+ \epsilon^3 \cdot 2\rho B \sigma F(y) S^2 \tilde{V}_{SS} \tilde{V}_{Sy} + \epsilon^4 \left(L_{\text{BS}}[\tilde{V}] + \rho B \sigma (1-\alpha)^2 F^2(y) S^3 \tilde{V}_{SS}^2 \tilde{V}_{Sy} \right)$$
$$+ \epsilon^5 \cdot \sigma^2 F(y) S^3 \tilde{V}_{SS}^2 + \frac{1}{2} \sigma^2 (1-\alpha)^2 F^2(y) S^4 \tilde{V}_{SS}^3 = 0. \quad (6.73)$$

We now assume that $\tilde{V}$ can be written in the form

$$\tilde{V} \sim h_{3_0} + \epsilon h_{3_1} + \epsilon^2 h_{3_2} + \epsilon^3 h_{3_3} + \cdots, \tag{6.74}$$

which then results in the following series of problems:

$\boldsymbol{O}(1)$: As usual the $O(1)$ problem is

$$L_0[h_{3_0}] = 0, \tag{6.75}$$

which implies the y-independence of h_{3_0}; that is

$$h_{3_0} = h_{3_0}(t, S). \tag{6.76}$$

$\boldsymbol{O}(\epsilon)$: The $O(\epsilon)$ equation is

$$L_0[h_{3_1}] + B^2 F(y) S h^2_{3_{0Sy}} = 0, \tag{6.77}$$

but since h_{3_0} is independent of y this simplifies to

$$L_0[h_{3_1}] = 0, \tag{6.78}$$

giving the result that

$$h_{3_1} = h_{3_1}(t, S). \tag{6.79}$$

$\boldsymbol{O}(\epsilon^2)$: The $O(\epsilon^2)$ equation is

$$L_0[h_{3_2}] + 2B^2 F(y) S h_{3_{0Sy}} h_{3_{1Sy}} + L_1[h_{3_0}] + \frac{1}{2} B^2 (1-\alpha)^2 F^2(y) S^2 h_{3_{0SS}} h^2_{3_{0Sy}} = 0, \tag{6.80}$$

but due to the y-independence of h_{3_0} and h_{3_1} this reduces to simply

$$L_0[h_{3_2}] = 0 \tag{6.81}$$

giving the result that the $O(\epsilon^2)$ term is y-independent also; that is

$$h_{3_2} = h_{3_2}(t, S). \tag{6.82}$$

$\boldsymbol{O}(\epsilon^3)$: The $O(\epsilon^3)$ equation is

$$L_0[h_{3_3}] + L_1[h_{3_1}] + B^2F(y)S\left(h_{3_{1Sy}}^2 + 2h_{3_{0Sy}}h_{3_{2Sy}}\right) + 2\rho B\sigma F(y)S^2 h_{3_{0SS}}h_{3_{0Sy}} + \frac{1}{2}B^2(1-\alpha)^2F^2(y)S^2\left(h_{3_{1SS}}h_{3_{0Sy}}^2 + 2h_{3_{0SS}}h_{3_{0Sy}}h_{3_{1Sy}}\right) = 0, \tag{6.83}$$

which reduces to

$$L_0[h_{3_3}] = 0, \tag{6.84}$$

when the y-independence of h_{3_0}, h_{3_1}, and h_{3_2} is taken into account. At $O(\epsilon^3)$ we therefore have

$$h_{3_3} = h_{3_3}(t, S). \tag{6.85}$$

$\boldsymbol{O}(\epsilon^4)$: The $O(\epsilon^4)$ equation is

$$\begin{aligned} &L_0[h_{3_4}] + L_1[h_{3_2}] + L_{\mathrm{BS}}[h_{3_0}] + 2B^2F(y)S\left(h_{3_{0Sy}}h_{3_{3Sy}} + h_{3_{1Sy}}h_{3_{2Sy}}\right) \\ &+ \frac{1}{2}B^2(1-\alpha)^2F^2(y)S^2\left(h_{3_{2SS}}h_{3_{0Sy}}^2 + h_{3_{0SS}}h_{3_{1Sy}}^2 + 2h_{3_{0Sy}}\left(h_{3_{0SS}}h_{3_{2Sy}} + h_{3_{1SS}}h_{3_{1Sy}}\right)\right) \\ &+ 2\rho B\sigma F(y)S^2\left(h_{3_{0SS}}h_{3_{1Sy}} + h_{3_{1SS}}h_{3_{0Sy}}\right) + \rho B\sigma(1-\alpha)^2F^2(y)S^3h_{3_{0SS}}^2h_{3_{0Sy}} = 0, \end{aligned} \tag{6.86}$$

which reduces to

$$L_0[h_{3_4}] + L_{\mathrm{BS}}[h_{3_0}] = 0 \tag{6.87}$$

when the y-independence of h_{3_0}, h_{3_1}, h_{3_2}, and h_{3_3} is taken into account. The existence condition for a solution to (6.87) requires the nonhomogeneous term to be centered *w.r.t.* $\phi(y)$, or

$$\langle L_{\mathrm{BS}}[h_{3_0}]\rangle = 0. \tag{6.88}$$

Since L_{BS} is independent of y we once again find

$$L_{\text{BS}}[h_{3_0}] = 0, \tag{6.89}$$

as the equation that the leading-order component of the derivative's value must satisfy for this case.

Substituting (6.89) back into (6.87) gives $L_0[h_{3_4}] = 0$, or

$$h_{3_4} = h_{3_4}(t, S). \tag{6.90}$$

$\boldsymbol{O}(\epsilon^5)$: The $O(\epsilon^4)$ equation is large and cumbersome and we expect the $O(\epsilon^5)$ equation to be even more so. Instead of writing the $O(\epsilon^5)$ equation out in full we will use the fact that h_{3_0}-h_{3_4} are y-independent so that we can eliminate terms that take derivatives *w.r.t.* y. This then immediately leaves us with the equation

$$L_0[h_{3_5}] + L_{\text{BS}}[h_{3_1}] + \sigma^2 F(y) S^3 h^2_{3_{0SS}} = 0, \tag{6.91}$$

which has the usual condition

$$\left\langle L_{\text{BS}}[h_{3_1}] + \sigma^2 F(y) S^3 h^2_{3_{0SS}} \right\rangle = 0, \tag{6.92}$$

required for the existence of a solution. Taking into account the y-independence of L_{BS} leaves us with

$$L_{\text{BS}}[h_{3_1}] + \sigma^2 \left\langle F(y) \right\rangle S^3 h^2_{3_{0SS}} = 0, \tag{6.93}$$

as the equation that the first correction to the Black-Scholes value must satisfy.

Substituting (6.93) back into (6.91) gives the relation

$$L_0[h_{3_5}] = \sigma^2 S^3 h^2_{3_{0SS}} \left(\left\langle F(y) \right\rangle - F(y) \right), \tag{6.94}$$

or

$$h_{3_5}(t, S, y) = \sigma^2 S^3 h^2_{3_{0SS}} \left(\psi(y) + c(t, S) \right), \tag{6.95}$$

where, as usual, $\psi(y)$ is defined by the equation $L_0[\psi(y)] = \langle F(y) \rangle - F(y)$ and $c(t, S)$ is

some arbitrary function independent of y.

$\boldsymbol{O}(\epsilon^6)$: So far we have found h_{3_0}, h_{3_1}, h_{3_2}, h_{3_3}, and h_{3_4} all to be functions independent of y and h_{3_5} to be the lowest-order y-dependent term in the expansion (6.74). Examining equation (6.4.2) it is easy to see that the $O(\epsilon^6)$ equation will reduce to

$$L_0[h_{3_6}] + L_{\mathrm{BS}}[h_{3_2}] + 2\sigma^2 F(y) S^3 h_{3_{0SS}} h_{3_{1SS}} + \frac{1}{2}\sigma^2(1-\alpha)^2 F^2(y) S^4 h_{3_{0SS}}^3 = 0. \tag{6.96}$$

For a solution to exist to this Poisson equation we require that the nonhomogeneous term is centered *w.r.t.* the invariant distribution, $\phi(y)$. Once the y-independence of the Black-Scholes operator is accounted for we are left with

$$L_{\mathrm{BS}}[h_{3_2}] + 2\sigma^2 \langle F(y) \rangle S^3 h_{3_{0SS}} h_{3_{1SS}} + \frac{1}{2}\sigma^2(1-\alpha)^2 \left\langle F^2(y) \right\rangle S^4 h_{3_{0SS}}^3 = 0. \tag{6.97}$$

as the equation that the second correction to the Black-Scholes value of the derivative must satisfy.

The form of the equation satisfied by the second correction in the region $c > 2d$ has a much different form to that in either of the regions $c = 2d$, or $c < 2d$. As in Case 2 the $\tilde{V}_{SS}^2$ and $\tilde{V}_{SS}^3$ terms remain consecutive when $c > 2d$ and the presence of the $h_{3_{0SS}}^3$ term in equation (6.97) is therefore expected. The difference results from the fact that L_0 and L_1 are not consecutive in this region causing L_1 not to act on h_{3_5} in the $O(\epsilon^6)$ equation resulting in both the $L_1[\tilde{V}]$ and $\tilde{V}_{Sy}^2$ terms being identically zero.

$\boldsymbol{O}(\epsilon^7)$: When $c > 2d$ we can take the analysis one step further without too much difficulty. Ignoring all terms of equation (6.4.2) that take derivatives *w.r.t.* y of terms that we have already shown to be independent of y results in the $O(\epsilon^7)$ equation

$$\begin{aligned} &L_0[h_{3_7}] + L_1[h_{3_5}] + L_{\mathrm{BS}}[h_{3_3}] \\ &\qquad + \sigma^2 F(y) S^3 \left(2 h_{3_{0SS}} h_{3_{2SS}} + h_{3_{1SS}}^2\right) + \frac{3}{2}\sigma^2(1-\alpha)^2 F^2(y) S^4 h_{3_{0SS}}^2 h_{3_{1SS}} = 0. \end{aligned} \tag{6.98}$$

Imposing the usual existence condition for a solution to this Poisson equation and simulta-

neously using equation (6.95) to eliminate h_{3_5} results in the equation

$$L_{\mathrm{BS}}[h_{3_3}] + \left\langle L_1\left[\sigma^2 S^3 h_{3_{0SS}}^2\left(\psi(y) + c(t,S)\right)\right]\right\rangle + \sigma^2 S^3\left(2h_{3_{0SS}}h_{3_{2SS}} + h_{3_{1SS}}^2\right)\langle F(y)\rangle + \frac{3}{2}\sigma^2(1-\alpha)^2 S^4 h_{3_{0SS}}^2 h_{3_{1SS}}\left\langle F^2(y)\right\rangle = 0. \quad (6.99)$$

Finally, substituting equation (6.47) into (6.99) and using the definitions of κ_1 and κ_2 from (6.49) and (6.50) gives us the relation that the third correction, h_{3_3}, must satisfy; this is

$$L_{\mathrm{BS}}[h_{3_3}] + \kappa_1 S^3 h_{3_{0SS}}^2 + \kappa_2 S^4 h_{3_{0SS}} h_{3_{0SSS}} + \sigma^2\langle F(y)\rangle S^3 h_{3_{1SS}}^2 + 2\sigma^2\langle F(y)\rangle S^3 h_{3_{0SS}} h_{3_{2SS}} + \frac{3}{2}\sigma^2(1-\alpha)^2\left\langle F^2(y)\right\rangle S^4 h_{3_{0SS}}^2 h_{3_{1SS}} = 0. \quad (6.100)$$

Summary and Comments

When $c > 2d$ the approximate solution is

$$\begin{gathered}\tilde{V} \sim h_{3_0} + \epsilon h_{3_1} + \epsilon^2 h_{3_2} + \epsilon^3 h_{3_3} + \cdots, \\ L_{\mathrm{BS}}[h_{3_0}] = 0, \\ \left\langle L_{\mathrm{BS}}[h_{3_1}] + \sigma^2 F(y) S^3 h_{3_{0SS}}^2\right\rangle = 0, \\ L_{\mathrm{BS}}[h_{3_2}] + 2\sigma^2\langle F(y)\rangle S^3 h_{3_{0SS}} h_{3_{1SS}} + \frac{1}{2}\sigma^2(1-\alpha)^2\left\langle F^2(y)\right\rangle S^4 h_{3_{0SS}}^3 = 0, \quad \text{and} \\ L_{\mathrm{BS}}[h_{3_3}] + \kappa_1 S^3 h_{3_{0SS}}^2 + \kappa_2 S^4 h_{3_{0SS}} h_{3_{0SSS}} + \sigma^2\langle F(y)\rangle S^3 h_{3_{1SS}}^2 \\ +2\sigma^2\langle F(y)\rangle S^3 h_{3_{0SS}} h_{3_{2SS}} + \frac{3}{2}\sigma^2(1-\alpha)^2\left\langle F^2(y)\right\rangle S^4 h_{3_{0SS}}^2 h_{3_{1SS}} = 0.\end{gathered} \quad (6.101)$$

In this region of the (c,d) space the market depth mean-reverts *very* rapidly. The first and second corrections to the leading-order Black-Scholes value of the contract now obey the exact equations as the first and second corrections to the leading-order outer solution in the constant-λ model, but, again, with the distributional averages of both $F(y)$ and $F^2(y)$. Because of this we now have that the liquidity risk corrections all enter into the solution through the third correction and have even less relative significance than they did when $c = 2d$.

6.5 Calibration of the Stochastic Market Depth Model

We have so far derived a pricing equation for a generic derivative contract in a market in which the market depth is generated by a mean-reverting stochastic process. Furthermore, by assuming both the market depth is small and that its rate of mean-reversion is large we derived a rescaled pricing equation and found that the model has three distinct types of behaviour depending on the relative sizes of the market depth process and its rate of mean-reversion. Finally, in each of these regions we have found the leading-order solution to the pricing equation along with the first two or three corrections to this value.

Comparing the analysis of the stochastic market depth model with the stochastic volatility example of Chapter 5 we can see that the asymptotic technique of Fouque, Papanicolaou, and Sircar has adapted well to our nonlinear model. While we have not been able to calculate an explicit solution to the first correction as was done with the stochastic volatility case, we have succeeded in reducing a parabolic PDE nonlinear in both of its two spatial dimensions to a series of simple nonhomogeneous Black-Scholes equations for each of the corrections to the leading-order Black-Scholes value.

While this linearization of the pricing problem is a useful simplification, in order to implement the model we must obtain numerical estimates for each of its parameters and the analysis has not greatly reduced the number of these. For this analysis to be of any great practical use we must therefore check whether a similar method of calibration to that developed in Section 5.6 for the stochastic volatility model can be adapted to the present model.

Towards this goal we introduce the *implied market depth*, λ_I, defined as the market depth that when used as the parameter λ in the Bakstein-Howison equation gives the observed market value of the contract; that is the relation

$$\tilde{V}_{\mathrm{BH}}(t, S, \lambda_I) = \tilde{V}_{\mathrm{obs}}(t, S, \lambda) \tag{6.102}$$

holds true.

With this definition we now proceed with calibrating the stochastic market depth model in an analogous way to that for the stochastic volatility model. We will focus only on the case $c = 2d$ as the analysis for both the regions $c < 2d$ and $c > 2d$ is the same.

For the observed market value of a derivative contract on the right hand side of equation (6.102) we use the asymptotic approximation derived for the case $c = 2d$ above; that is

$$\tilde{V}_{\text{obs}}(t, S, \lambda) = h_{2_0} + \epsilon h_{2_1} + \epsilon^2 h_{2_2} + \cdots , \tag{6.103}$$

where h_{2_0}, h_{2_1}, and h_{2_2} are defined by the relations (6.63), (6.67), and (6.72), respectively.

For the left-hand side we first make the assumption that the implied market depth differs from the average market depth, $\langle \bar{f}(y) \rangle$, by a small amount. Under this assumption we can then write λ_I as the expansion

$$\lambda_I = \langle \bar{f}(y) \rangle + \delta \eta(t, S) + \cdots , \tag{6.104}$$

where δ is some small parameter that will be determined in terms of ϵ (specifically $\delta = o(\epsilon)$ since $\langle \bar{f}(y) \rangle = O(\epsilon)$) below and $\eta(t, S)$ is an $O(1)$ function.

In the same way that $\tilde{V}_{\text{BS}}(t, S, \sigma_I)$ was expanded about $\bar{\sigma}$ in the stochastic volatility example, we now expand $\tilde{V}_{\text{BH}}(t, S, \lambda_I)$ about $\langle \bar{f}(y) \rangle$ to get

$$\tilde{V}_{\text{BH}}(t, S, \lambda_I) = \tilde{V}_{\text{BH}}(t, S, \langle \bar{f}(y) \rangle) + \delta \eta(t, S) \left. \frac{\partial \tilde{V}_{\text{BH}}}{\partial \lambda} \right|_{\langle \bar{f}(y) \rangle} + \frac{1}{2} \delta^2 \eta(t, S)^2 \left. \frac{\partial^2 \tilde{V}_{\text{BH}}}{\partial \lambda^2} \right|_{\langle \bar{f}(y) \rangle} + \cdots . \tag{6.105}$$

The $O(1)$ term, $\tilde{V}_{\text{BH}}(t, S, \langle \bar{f}(y) \rangle)$, on the left-hand side of equation (6.103) satisfies the standard $\gamma = 0$ equation, while on the right hand side, h_{2_0} satisfies the Black-Scholes equation. We can put $\tilde{V}_{\text{BH}}(t, S, \langle \bar{f}(y) \rangle)$ into a more useful form if we use the results of our asymptotic analysis of the constant liquidity $\gamma = 0$ equation. From this work we know that $\tilde{V}_{\text{BH}}$ has two asymptotic forms depending on whether the time to expiry, $T - t$, is large or small compared with $\langle \bar{f}(y) \rangle^2$. But the fast-timescale asymptotic analysis of this chapter is only valid for $T - t \gg 1/a$; since $\langle \bar{f}(y) \rangle^2 = O(\epsilon^2)$ and $1/a = O(\epsilon)$, if we use the expansion valid for $T - t \gg \langle \bar{f}(y) \rangle^2$ then we are guaranteed to be in the correct region. In this large-time region, $\tilde{V}_{\text{BH}}(t, S, \langle \bar{f}(y) \rangle)$ can be written as the regular expansion

$$\tilde{V}_{\text{BH}}(t, S, \langle \bar{f}(y) \rangle) = \tilde{V}_{\text{BS}}(t, S) + \langle \bar{f}(y) \rangle \tilde{V}_1(t, S) + \langle \bar{f}(y) \rangle^2 \tilde{V}_2(t, S) + \cdots , \tag{6.106}$$

where

$$L_{\mathrm{BS}}[\tilde{V}_1] + \sigma^2 S^3 \tilde{V}^2_{\mathrm{BS}_{SS}} = 0, \qquad \text{and} \tag{6.107}$$

$$L_{\mathrm{BS}}[\tilde{V}_2] + 2\sigma^2 S^3 \tilde{V}_{\mathrm{BS}_{SS}} \tilde{V}_{1_{SS}} + \frac{1}{2}\sigma^2 S^4 \tilde{V}^3_{\mathrm{BS}_{SS}} = 0. \tag{6.108}$$

Combining (6.106) and (6.105) gives

$$\tilde{V}_{\mathrm{BH}}(t,S,\lambda_I) = \tilde{V}_{\mathrm{BS}}(t,S) + \epsilon\langle F(y)\rangle \tilde{V}_1(t,S) + \epsilon^2 \langle F(y)\rangle^2 \tilde{V}_2(t,S) + \delta\eta(t,S)\tilde{V}_1(t,S) + \cdots, \tag{6.109}$$

where we have used the fact that $\langle \bar{f}(y)\rangle = \epsilon\langle F(y)\rangle$.

Examining equations (6.103) and (6.109) it is clear that for the two sides of (6.102) to properly match we must have

$$\delta = \epsilon^2. \tag{6.110}$$

With this condition, equating the two sides of (6.102) gives the following comparison conditions:

$\boldsymbol{O}(1)$: At $O(1)$ we have the relation

$$\tilde{V}_{\mathrm{BS}}(t,S) = h_{2_0}(t,S), \tag{6.111}$$

which is trivially satisfied as both $\tilde{V}_{\mathrm{BS}}$ and h_{2_0} satisfy the Black-Scholes equation and the same boundary conditions.

$\boldsymbol{O}(\epsilon)$: At $O(\epsilon)$ the matching condition is

$$\langle F(y)\rangle \tilde{V}_1(t,S) = h_{2_1}(t,S). \tag{6.112}$$

If we operate on both sides of this equation by L_{BS} and then substitute (6.107) and (6.67) for the left and right sides, respectively, we find

$$\sigma^2 \langle F(y)\rangle S^3 \tilde{V}^2_{\mathrm{BS}_{SS}} = \sigma^2 \langle F(y)\rangle S^3 h^2_{2_{0SS}},$$

which, by equation (6.111) is also therefore trivially satisfied.

$\boldsymbol{O}(\epsilon^2)$: At $O(\epsilon^2)$ the matching condition is

$$\eta(t,S)\tilde{V}_1(t,S) + \langle F(y)\rangle^2 \tilde{V}_2(t,S) = h_{2_2}(t,S). \tag{6.113}$$

The main goal of the calibration process can be thought of as finding an expression for $\eta(t,S)$, but with the stochastic market depth model the situation is more difficult as we no longer have analytical expressions for the correction terms $\tilde{V}_1(t,S)$, $\tilde{V}_2(t,S)$, and $h_{2_2}(t,S)$. To find an expression for $\eta(t,S)$ we will need to use a different approach.

To find $\eta(t,S)$ we first introduce the Green's function, $G(t,S,t',S')$, for the Black-Scholes operator:

$$G(t,S,t',S') = \frac{e^{-r(t'-t)}}{\sigma S'\sqrt{2\pi(t'-t)}} \exp\left(-\frac{\left(\log(S/S') + \left(r - \frac{1}{2}\sigma^2\right)(t'-t)\right)^2}{2\sigma^2(t'-t)}\right). \tag{6.114}$$

Using $G(t,S,t',S')$ along with equations (6.72), (6.107), and (6.108) allows us to write

$$\tilde{V}_1(t,S) = \sigma^2 \int_0^\infty \int_T^t G(t,S,t',S') S'^3 \tilde{V}^2_{\mathrm{BS}_{S'S'}} dt' dS' \tag{6.115}$$

$$\tilde{V}_2(t,S) = \sigma^2 \int_0^\infty \int_T^t G(t,S,t',S') \left(2S'^3 \tilde{V}_{\mathrm{BS}_{S'S'}} \tilde{V}_{1_{S'S'}} + \frac{1}{2}(1-\alpha)^2 S'^4 \tilde{V}^3_{\mathrm{BS}_{S'S'}}\right) dt' dS' \tag{6.116}$$

$$\begin{aligned} h_{2_2}(t,S) = \sigma^2 \int_0^\infty \int_T^t G(t,S,t',S') \Big(&\kappa_1 a^2 S'^3 h^2_{2_{0S'S'}} + \kappa_2 a^2 S'^4 h_{2_{0S'S'}} h_{2_{0S'S'S'}} \\ &+ 2\langle F(y)\rangle S'^3 h_{2_{0S'S'}} h_{2_{1S'S'}} + \frac{1}{2}(1-\alpha)^2 \langle F^2(y)\rangle S'^4 h^3_{2_{0S'S'}}\Big) dt' dS'. \end{aligned} \tag{6.117}$$

Using the $O(1)$ and $O(\epsilon)$ matching conditions to convert $h_{2_0} \to \tilde{V}_{\mathrm{BS}}$ and $h_{2_1} \to \tilde{V}_1$, then substituting (6.115), (6.116), and (6.117) into (6.113), and finally solving for $\eta(t,S)$ gives

$$\begin{aligned} \eta(t,S) = \kappa_1 &+ \kappa_2 \frac{\int_0^\infty \int_T^t G(t,S,t',S') S'^4 \tilde{V}_{\mathrm{BS}_{S'S'}} \tilde{V}_{\mathrm{BS}_{S'S'S'}} dt' dS'}{\int_0^\infty \int_T^t G(t,S,t',S') S'^3 \tilde{V}^2_{\mathrm{BS}_{S'S'}} dt' dS'} \\ &+ \frac{1}{2}(1-\alpha)^2 \left(\langle F^2(y)\rangle - \langle F(y)\rangle^2\right) \frac{\int_0^\infty \int_T^t G(t,S,t',S') S'^4 \tilde{V}^3_{\mathrm{BS}_{S'S'}} dt' dS'}{\int_0^\infty \int_T^t G(t,S,t',S') S'^3 \tilde{V}^2_{\mathrm{BS}_{S'S'}} dt' dS'}. \end{aligned} \tag{6.118}$$

Defining the integrals

$$I_0 = \int_0^\infty \int_T^t G(t,S,t',S') S'^3 \tilde{V}^2_{\mathrm{BS}_{S'S'}} dt' dS', \tag{6.119}$$

$$I_1 = \int_0^\infty \int_T^t G(t,S,t',S') S'^4 \tilde{V}_{\mathrm{BS}_{S'S'}} \tilde{V}_{\mathrm{BS}_{S'S'S'}} dt' dS', \quad \text{and} \tag{6.120}$$

$$I_2 = \int_0^\infty \int_T^t G(t,S,t',S') S'^4 \tilde{V}^3_{\mathrm{BS}_{S'S'}} dt' dS', \tag{6.121}$$

then the implied market depth, λ_I, takes the form

$$\lambda_I = \langle \bar{f}(y) \rangle + \tilde{\kappa_1} + \tilde{\kappa_2} \frac{I_1}{I_0} + \frac{1}{2}(1-\alpha)^2 \left(\langle \bar{f}^2(y) \rangle - \langle \bar{f}(y) \rangle^2 \right) \frac{I_2}{I_0} + O(1/a^3), \tag{6.122}$$

where $\tilde{\kappa_1} = a^2 \kappa_1$ and $\tilde{\kappa_2} = a^2 \kappa_2$.

In order to obtain numerical estimates for $\tilde{\kappa_1}$ and $\tilde{\kappa_2}$ we need to know the dependence of I_1/I_0 on t and S. To this goal we will first convert (6.119) and (6.120) back into their differential forms. By rescaling each of these equations and focusing on a restricted region of the (t,S) space we will then be able to derive approximate solutions valid in this restricted domain. While we will be restricting the validity of our model using these approximations, we will show that this restriction is both consistent with previous approximations within the analysis and not limiting to the applicability of the model.

Following the example of the stochastic volatility model, since there exists a simple analytical formula for the value of a European call option we will calibrate the model using this contract.[1] The Black-Scholes value of a vanilla call option with strike, K, and expiry, T, was given in equation (5.64); from this $\tilde{V}_{\mathrm{BS}_{SS}}$ and $\tilde{V}_{\mathrm{BS}_{SSS}}$ are easily found to be

$$\tilde{V}_{\mathrm{BS}_{SS}}(t,S) = \frac{1}{\sigma S \sqrt{2\pi(T-t)}} \exp\left(-\frac{1}{2} d_1^2 \right), \quad \text{and} \tag{6.123}$$

$$\tilde{V}_{\mathrm{BS}_{SSS}}(t,S) = -\frac{d_1 + \sigma\sqrt{T-t}}{\sqrt{2\pi}\sigma^2 S^2 (T-t)} \exp\left(-\frac{1}{2} d_1^2 \right), \tag{6.124}$$

where $'$ denotes a derivative taken *w.r.t.* d_1, and d_1 and d_2 are defined in (5.65) and (5.66).

[1] However, since (6.122) was derived independent of any contract-specific boundary conditions it will hold generally for all contracts written on the underlying.

We begin with the equation for I_0; in differential form I_0 satisfies

$$\begin{aligned} I_{0_t} + rSI_{0_S} + \frac{1}{2}\sigma^2 S^2 I_{0_{SS}} - rI_0 + \sigma^2 S^3 \tilde{V}^2_{\mathrm{BS}_{SS}} = 0, \\ I_0(T,S) = 0, \quad I_0(t,0) = 0, \text{ and } I_0(t,S) \to 0 \text{ as } S \to \infty. \end{aligned} \tag{6.125}$$

We now introduce the change of variables

$$\tau = \frac{1}{2}\sigma^2(T-t), \quad X = \frac{S-K}{K}, \quad \text{and} \quad I_0 = KH_0, \tag{6.126}$$

and the rescaling

$$\tau = \varepsilon\hat{\tau}, \quad X = \sqrt{\varepsilon}\hat{X}, \quad \text{and} \quad H_0 = H_{0_0} + \sqrt{\varepsilon}H_{0_1} + \cdots, \tag{6.127}$$

where ε is an arbitrary small parameter unrelated to ϵ. Under (6.126) and (6.127) H_{0_0} will then satisfy

$$\begin{aligned} H_{0_{0\hat{\tau}}} = H_{0_{0\hat{X}\hat{X}}} + \frac{1}{2\pi\hat{\tau}} e^{-\frac{\hat{X}^2}{2\hat{\tau}}}, \\ H_{0_0}(0,\hat{X}) = 0, \quad \text{and} \quad H_{0_0}(\hat{\tau},\hat{X}) \to 0 \text{ as } \hat{X} \to \pm\infty, \end{aligned} \tag{6.128}$$

which will be a valid approximation to the problem (6.125) in the region $\tau \ll 1$ and $X = O(\sqrt{\tau})$. Now, under the similarity transformation

$$H_{0_0}(\hat{\tau}) = W_0(\xi), \quad \text{where} \quad \xi = \frac{\hat{X}}{\sqrt{\hat{\tau}}}, \tag{6.129}$$

equation (6.128) reduces to

$$\begin{aligned} \frac{d^2W_0}{d\xi^2} + \frac{1}{2}\xi\frac{dW_0}{d\xi} + \frac{1}{2\pi}e^{-\frac{1}{2}\xi^2} = 0, \\ W_0(\xi) \to 0 \quad \text{as} \quad \xi \to \pm\infty. \end{aligned} \tag{6.130}$$

Integrating (6.130) twice *w.r.t.* ξ and imposing both boundary conditions gives the solution

$$W_0(\xi) = \frac{1}{2\sqrt{\pi}}\left(1 - \mathrm{erf}\,(\xi/2)^2\right), \tag{6.131}$$

or, in terms of τ and X

$$I_0(\tau, X) = \frac{K}{2\sqrt{\pi}} \left(1 - \operatorname{erf}\left(\frac{X}{2\sqrt{\tau}} \right)^2 \right) + \cdots . \tag{6.132}$$

The process for finding an approximation to I_1 is similar. The differential form of (6.120) is

$$\begin{aligned} &I_{1_t} + rSI_{1_S} + \frac{1}{2}\sigma^2 S^2 I_{1_{SS}} - rI_1 + \sigma^2 S^4 \tilde{V}_{\mathrm{BS}_{SS}} \tilde{V}_{\mathrm{BS}_{SSS}} = 0, \\ &I_1(T, S) = 0, \quad I_1(t, 0) = 0, \text{ and } I_1(t, S) \to 0 \text{ as } S \to \infty. \end{aligned} \tag{6.133}$$

We now use the change of variables, (6.126), but a slightly different rescaling

$$\tau = \varepsilon\hat{\tau}, \quad X = \sqrt{\varepsilon}\hat{X}, \quad \text{and} \quad H_1 = \frac{1}{\sqrt{\varepsilon}} H_{1_0} + H_{1_1} + \cdots , \tag{6.134}$$

and the leading-order behaviour of (6.133) becomes

$$\begin{aligned} H_{1_{0\hat{\tau}}} &= H_{1_{0\hat{X}\hat{X}}} + \frac{1}{2\pi} \frac{\hat{X}}{\hat{\tau}^2} e^{-\frac{\hat{X}^2}{2\hat{\tau}}}, \\ H_{1_0}(0, \hat{X}) = 0, \quad &\text{and} \quad H_{1_0}(\hat{\tau}, \hat{X}) \to 0 \text{ as } \hat{X} \to \pm\infty. \end{aligned} \tag{6.135}$$

If we seek a similarity solution of the form

$$H_{1_0}(\hat{\tau}, \hat{X}) = \frac{1}{\sqrt{\hat{\tau}}} W_1(\xi), \tag{6.136}$$

then (6.135) reduces to

$$\begin{aligned} &\frac{d^2 W_1}{d\xi^2} + \frac{1}{2}\xi \frac{dW_1}{d\xi} + \frac{1}{2} W_1 + \frac{1}{4\pi} \xi e^{-\frac{1}{2}\xi^2}, \\ &W_1(\xi) \to 0 \quad \text{as} \quad \xi \to \pm\infty. \end{aligned} \tag{6.137}$$

Again, integrating (6.137) twice *w.r.t.* ξ and imposing the boundary conditions gives the solution

$$W_1(\xi) = \frac{1}{4\sqrt{\pi}} \left(1 - \operatorname{erf}(\xi/2) \right) e^{-\frac{1}{4}\xi^2} \tag{6.138}$$

or,

$$I_1(\tau, X) = \frac{K}{4\sqrt{\pi\tau}} \left(1 - \text{erf}\left(\frac{X}{2\sqrt{\tau}}\right)\right) e^{-\frac{X^2}{4\tau}} + \cdots . \tag{6.139}$$

Combining (6.132) and (6.139) we finally find

$$\frac{I_1}{I_0} = \frac{1}{2\sqrt{\tau}} \left(\frac{\exp(-\frac{X^2}{4\tau})}{1 + \text{erf}\left(\frac{X}{2\sqrt{\tau}}\right)} \right). \tag{6.140}$$

At this point we could continue and attempt to derive an expression for I_2/I_0, but we will not do so for two reasons. First, while it is possible to find an ODE in terms of a similarity variable for the leading-order term in I_2, finding a closed-form solution to this equation is much more difficult than was the case for either I_0 or I_1.

Second, from equation (6.122) it is clear that the I_2/I_0 term is independent of both $\tilde{\kappa_1}$ and $\tilde{\kappa_2}$. To simplify notation, let

$$\Omega(t, S) = \frac{1}{2}(1-\alpha)^2 \left(\langle \bar{f}^2(y) \rangle - \langle \bar{f}(y) \rangle^2 \right) \frac{I_2}{I_0}. \tag{6.141}$$

Then, given empirical estimates for the parameters r, σ, α, $\langle \bar{f}(y) \rangle$, and $\langle \bar{f}^2(y) \rangle$, Ω will be completely determined in terms of t and S. By moving $\langle \bar{f}(y) \rangle$ and $\Omega(t, S)$ to the *l.h.s.* of (6.122) and substituting (6.140) for I_1/I_0 we get

$$\lambda_I - \langle \bar{f}(y) \rangle - \Omega(t, S) = \tilde{\kappa_1} + \tilde{\kappa_2} \frac{1}{2\sqrt{\tau}} \left(\frac{\exp(-\frac{X^2}{4\tau})}{1 + \text{erf}\left(\frac{X}{2\sqrt{\tau}}\right)} \right), \tag{6.142}$$

and it is now clear that by regressing the *l.h.s.* of (6.142) against τ and X we can fairly simply obtain estimates for $\tilde{\kappa_1}$ and $\tilde{\kappa_2}$.

Validity of Calibration By rescaling equations (6.125) and (6.133) with the arbitrary small parameter, ε we have limited the region of the (τ, X) space where the model can be calibrated to

$$\tau \ll 1, \qquad\qquad X = O(\sqrt{\tau}). \tag{6.143}$$

We note that this restriction is firstly consistent with our use of the regular expansion form of $\tilde{V}_{\text{BH}}$ (equation (6.106)) which imposed the constraint $\tau \gg \langle \bar{f}(y) \rangle^2$ and, secondly, that

it should not actually prove too great of a actual restriction as for a value of $\sigma = 0.3$ the above condition is easily satisfied even for contracts with $T - t = 1$ year.

6.5.1 Calibration Procedure

To clarify the above analysis we will now give a step-by-step procedure for how to calibrate our stochastic market depth model.

1. Measure r, σ, α, $\langle \bar{f}(y) \rangle$, and $\langle \bar{f}^2(y) \rangle$ using historical values of the asset price and market depth processes.

2. Using the constant-liquidity Bakstein-Howison equation, calculate the implied market depth for the necessary range of τ and X values for a simple vanilla call option. Furthermore, care must be taken to ensure that $\langle \bar{f}(y) \rangle^2 \ll \tau \ll 1$ and $X \gg \langle \bar{f}(y) \rangle$ so that both (6.106) and (6.143) are valid.

3. For these same values of τ and X numerically compute the integrals I_2 and I_0. Using these values compute $\Omega(\tau, X)$ and then finally the *l.h.s.* of equation (6.119).

4. Fit the values of the *l.h.s.* of equation (6.119) against τ and X of the form on the *r.h.s.* of the equation to get estimates for $\tilde{\kappa_1}$ and $\tilde{\kappa_2}$.

5. For the desired contract, calculate its leading-order Black-Scholes component using equation (6.63). For contracts with simple payoff structures it is possible that this component has an analytical form, but for more complex payoff structures this step will most likely require a numerical calculation.

6. When we go to assemble the expansion solution (given in equation (6.56) we will need a value for ϵh_{2_1}. If we multiply equation (6.67) by ϵ we obtain

$$L_{\text{BS}} \left[\epsilon h_{2_1} \right] + \sigma^2 \langle \bar{f}(y) \rangle S^3 h^2_{2_{0SS}} = 0, \tag{6.144}$$

   which, since all quantities within the nonhomogeneous term are known, can be simply solved for ϵh_{2_1}.

7. To calculate the second-order correction we multiply equation (6.72) by ϵ^2 to get

$$L_{\text{BS}}\left[\epsilon^2 h_{2_2}\right] + \tilde{\kappa_1}\sigma^2 S^3 h_{2_{0SS}}^2 + \tilde{\kappa_2}\sigma^2 S^4 h_{2_{0SS}} h_{2_{0SSS}} + 2\sigma^2\langle\bar{f}(y)\rangle S^3 h_{2_{0SS}} \cdot \epsilon h_{2_{1SS}} + \frac{1}{2}\sigma^2(1-\alpha)^2\langle\bar{f}^2(y)\rangle S^4 h_{2_{0SS}}^3 = 0. \quad (6.145)$$

Once again all quantities within the nonhomogeneous terms of (6.145) are known and so we simply need to numerically solve this equation to find $\epsilon^2 h_{2_2}$.

6.6 The American Forward with Liquidity Risk

In Chapter 4 we presented the American forward as a possible hedging instrument against liquidity risk in the underlying. At that time the contract was priced using the standard BH model in which the liquidity parameters are constant, but pricing a contract to be used as a liquidity hedge with a constant liquidity pricing model is contradictory as there can be no liquidity risk in an asset with constant liquidity.

In this chapter we have developed a pricing framework for a contract contingent on an asset with stochastic liquidity. We will now use this framework to reprice the American forward and, in doing so, quantify the effect of the underlying's liquidity risk on the value of the forward contract.

From the asymptotic analysis of Section 6.4 we know that the leading-order component of a derivative's value in a stochastic liquidity environment is the Black-Scholes value and the first correction is entirely driven by the *average* market depth; it is not until the second correction that effects due to the stochastic nature of the market depth enter the solution. Given that the first correction is $O(\bar{\lambda})$ and therefore small (this is the effect studied in Chapter 4) we do not expect the stochastic liquidity correction to have a significant impact on the contract's value, but the *form* of the is correction is still of interest. For this reason, we will not concern ourselves with calibrating the model to market data in this study, only using physically reasonable approximate values for the parameters. The specific parameter values used will be discussed below.

The pricing framework developed in Section 6.4 is only valid for European vanilla contracts; to price the American forward we must extend the framework to account for the early exercise feature of the contract. While this extension will introduce some fundamen-

tal changes into the model, much of the analysis will remain the same; to avoid too much repetition we will therefore skip over many of the steps in the derivation.

The formulation of the pricing problem for the ask price of an American call forward contingent on an asset with a stochastic market depth is

$$\begin{aligned}
&\tilde{V}_t + (r-D)S\tilde{V}_S + \frac{1}{2}\sigma^2 S^2 \tilde{V}_{SS} - r\tilde{V} + (a(m-y) - \beta\Lambda_\lambda)\tilde{V}_y + \frac{1}{2}\beta^2 \tilde{V}_{yy} + \rho\beta\sigma S\tilde{V}_{Sy} \\
&+\bar{f}(y)\sigma^2 S^3 \tilde{V}_{SS}^2 + \frac{1}{2}(1-\alpha)^2\sigma^2 \bar{f}^2(y) S^4 \tilde{V}_{SS}^3 + 2\rho\beta\sigma\bar{f}(y)S^2\tilde{V}_{SS}\tilde{V}_{Sy} + \beta^2\bar{f}(y)S\tilde{V}_{Sy}^2 \\
&+\rho\beta\sigma(1-\alpha)^2\bar{f}^2(y)S^3\tilde{V}_{SS}^2\tilde{V}_{Sy} + \frac{1}{2}\beta^2(1-\alpha)^2\bar{f}^2(y)S^2\tilde{V}_{SS}\tilde{V}_{Sy}^2 = 0: \qquad S < S_f(t,y), \\
&\qquad\qquad \tilde{V}(t,S,y) = S - \tilde{F}: \qquad\qquad S \geq S_f(t,y), \\
&\qquad\qquad \tilde{V}(T,S,y) = S - \tilde{F}, \\
&\tilde{V}(t,S_f(t,y),y) = S_f(t,y) - \tilde{F}, \qquad \tilde{V}_S(t,S_f(t,y),y) = 1, \qquad \tilde{V}_y(t,S_f(t,y),y) = 0,
\end{aligned} \tag{6.146}$$

for a forward price, $\tilde{F}$, dividend yield, D, expiry, T, and free boundary, $S_f(t,y)$, which is assumed to depend on both t and y. The formulation of this problem is fairly self-explanatory with the exception of the new free boundary condition on $\tilde{V}_y$. It is not immediately obvious that $\tilde{V}_y = 0$ across the free boundary as it is completely reasonable that the exercise value of the contract should change as the market depth changes, but for now we will simply assume that this condition holds and show later that this is true at least to the order of accuracy of our approximate solution.

We will only analyze the case $(c,d) = (1,1)$ for the American forward; the analysis for the other two regions of the (c,d) space is similar, but will not be explicitly demonstrated to avoid repetition. We therefore rescale (6.146) according to

$$\begin{aligned}
a &= \frac{1}{\epsilon}, \\
\beta &= \frac{1}{\epsilon^{\frac{1}{2}}}B, \\
\bar{f}(y) &= \epsilon F(y),
\end{aligned} \tag{6.147}$$

which results in the rescaled free boundary problem

$$\begin{aligned}
&L_0[\tilde{V}] + \epsilon^{\frac{1}{2}} L_1[\tilde{V}] + \epsilon \left(L_{\mathrm{BS}}[\tilde{V}] + B^2 F(y) S \tilde{V}_{Sy}^2 \right) \\
&+\epsilon^{\frac{3}{2}} \cdot 2\rho B\sigma F(y) S^2 \tilde{V}_{SS} \tilde{V}_{Sy} + \epsilon^2 \left(\sigma^2 F(y) S^3 \tilde{V}_{SS}^2 + \frac{1}{2} B^2 (1-\alpha)^2 F^2(y) S^2 \tilde{V}_{SS} \tilde{V}_{Sy}^2 \right) \\
&+\epsilon^{\frac{5}{2}} \cdot \rho B\sigma (1-\alpha)^2 F^2(y) S^3 \tilde{V}_{SS}^2 \tilde{V}_{Sy} + \epsilon^3 \cdot \frac{1}{2} \sigma^2 (1-\alpha)^2 F^2(y) S^4 \tilde{V}_{SS}^3 = 0, \quad S < S_f(t,y), \\
&\tilde{V}(t,S,y) = S - \tilde{F}, \qquad S \geq S_f(t,y), \\
&\tilde{V}(T,S,y) = S - \tilde{F}, \qquad (6.148) \\
&\tilde{V}(t, S_f(t,y), y) = S_f(t,y) - \tilde{F}, \qquad \tilde{V}_S(t, S_f(t,y), y) = 1, \qquad \tilde{V}_y(t, S_f(t,y), y) = 0.
\end{aligned}$$

Note that because of our specific scaling $\bar{f}(y) = \epsilon F(y)$, ϵ is precisely $\bar{\lambda}$.

To solve this problem we will follow the method of FPS [23]. We write both the function, $\tilde{V}$, and the free boundary, $S_f(t,y)$, as expansions in powers of $\epsilon^{\frac{1}{2}}$; that is

$$\tilde{V}(t,S,y) \sim h_{10}(t,S,y) + \epsilon^{\frac{1}{2}} h_{11}(t,S,y) + \epsilon h_{12}(t,S,y) + \cdots, \qquad (6.149)$$

$$S_f(t,y) \sim S_0(t,y) + \epsilon^{\frac{1}{2}} S_1(t,y) + \epsilon S_2(t,y) + \cdots. \qquad (6.150)$$

We now substitute (6.149) and (6.150) into (6.148) and solve the resulting series of problems; at $O(1)$ this is

$$\begin{aligned}
&L_0[h_{10}(t,S,y)] = 0, \qquad S < S_0(t,y), \\
&h_{10}(t,S,y) = S - \tilde{F}, \qquad S \geq S_0(t,y), \\
&h_{10}(T,S,y) = S - \tilde{F}, \qquad (6.151) \\
&h_{10}(t, S_0(t,y), y) = S_0(t,y) - \tilde{F}, \qquad h_{10_S}(t, S_(t,y), y) = 1, \qquad h_{10_y}(t, S_0(t,y), y) = 0.
\end{aligned}$$

In the hold region we have $L_0[h_{10}(t,S,y)] = 0$; by the usual argument that since L_0 is the infinitesimal generator of y_t we have the result that h_{10} must be independent of y; that is

$$h_{10} = h_{10}(t,S), \qquad S < S_f(t,y). \qquad (6.152)$$

As $h_{10}(t,S,y) = S - \tilde{F}$ is y-independent in the exercise region as well, however, it must also be y-independent *on* the free boundary, but the only way that h_{10} can be independent of y

on the free boundary is for the free boundary itself to be independent of y; that is

$$S_0 = S_0(t). \tag{6.153}$$

The plan for the higher-order problems is to solve for the higher-order corrections to the contract value, but only retain the leading-order term in the free-boundary expansion. In making this approximation we will introduce an $O(\epsilon^{\frac{1}{2}})$ error into the position of the free boundary. Within this $O(\epsilon^{\frac{1}{2}})$ band the solution will only be accurate to $O(1)$ and will therefore ignore liquidity effects (as we have already found $h_{10} = h_{10}(t, S)$). Outside of this band our approximation will hold; this should present no problem in pricing the American forward as the position of the forward price (the zero point of $\tilde{V}$) should be well away from the free boundary for large enough times to expiry when the asymptotic approximation is valid to begin with.

At $O(\epsilon^{\frac{1}{2}})$ the problem is

$$\begin{aligned} L_0[h_{11}(t, S,y)] + L_1[h_{10}(t, S)] = 0, \qquad & S < S_0(t), \\ h_{11}(t, S, y) = 0, \qquad & S \geq S_0(t), \\ h_{11}(T, S, y) = 0 \qquad & , \end{aligned} \tag{6.154}$$

with boundary conditions

$$h_{11}(t, S_0(t), y) = 0, \tag{6.155}$$

$$\begin{aligned} h_{11_S}(t, S_0(t), y) + S_1(t, y)h_{10_{SS}}(t, S_0(t), y) = 0, \\ \implies h_{11_S}(t, S_0(t), y) = 0, \end{aligned} \tag{6.156}$$

and

$$\begin{aligned} h_{11_y}(t, S_0(t), y) + S_1(t, y)h_{10_{Sy}}(t, S_0(t), y) = 0, \\ \implies h_{11_y}(t, S_0(t), y) = 0. \end{aligned} \tag{6.157}$$

In the second line of both (6.156) and (6.157) we used properties of $h_{10}(t, S_0(t), y)$ determined from the $O(1)$ problem. As h_{10} is y-independent we have $L_1[h_{10}(t, S, y)] = 0 \implies$

$L_0[h_{11}(t, S_0(t), y)] = 0$ and so

$$h_{11} = h_{11}(t, S), \tag{6.158}$$

in the hold region. This situation is similar to what was found in the $O(1)$ problem; from (6.154) and (6.158) we know the $O(\epsilon^{\frac{1}{2}})$ correction is y-independent on both sides of the free boundary (in fact it is precisely zero in the exercise region) and it must therefore also be *on* the boundary implying that the boundary itself is y-independent up to $O(\epsilon^{\frac{1}{2}})$.

At $O(\epsilon)$ the problem is

$$\begin{aligned} L_0[h_{12}(t,S,y)] + L_1[h_{11}(t,S)] + L_{\mathrm{BS}}[h_{10}(t,S)] + B^2F(y)Sh^2_{10_{Sy}} = 0, \qquad & S < S_0(t), \\ h_{12}(t,S,y) = 0, & S \geq S_0(t), \\ h_{12}(t,S_0(t),y) = 0, \quad h_{12_S}(t,S_0(t),y) = 0, \quad h_{12_S}(t,S_0(t),y) = 0, & \end{aligned} \tag{6.159}$$

where we have omitted the obvious zero final condition and have used the same reasoning to derive the free boundary conditions as was used in (6.155), (6.156), and (6.157) for the $O(\epsilon^{\frac{1}{2}})$ problem. Because of the y-independence of h_{10} and h_{11} the equation for h_{12} in the hold region reduces to

$$L_0[h_{12}(t,S,y)] + L_{\mathrm{BS}}[h_{10}(t,S)] = 0, \tag{6.160}$$

which has the solvability condition

$$\langle L_{\mathrm{BS}}[h_{10}(t,S)] \rangle = L_{\mathrm{BS}}[h_{10}(t,S)] = 0, \tag{6.161}$$

and tells us that the leading-order component of the solution in the hold region satisfies the Black-Scholes equation. Combining (6.161) with the boundary conditions of (6.151) gives the full problem that h_{10} satisfies; this is

$$\begin{aligned} L_{\mathrm{BS}}[h_{10}(t,S)] = 0, & \qquad S < S_0(t), \\ h_{10}(t,S,y) = S - \tilde{F}, & \qquad S \geq S_0(t), \\ h_{10}(t,S_0(t),y) = S_0(t) - \tilde{F}, \quad h_{10_S}(t,S_0(t),y) = 1, \quad h_{10_S}(t,S_0(t),y) = 0. & \end{aligned} \tag{6.162}$$

The leading-order solution is therefore exactly the Black-Scholes value of the American call forward.

It should be fairly clear from the $O(1)$, $O(\epsilon^{\frac{1}{2}})$, and $O(\epsilon)$ problems that the analysis of Section 6.4 for a European contract will be the same for an American contract in its hold region, but where the right-hand boundary is now $S_0(t)$ instead of $+\infty$. Omitting the calculations, the American call forward has an ask price, $\tilde{V}$, given by

$$\tilde{V}(t,S,y) \sim h_{10}(t,S,y) + \epsilon^{\frac{1}{2}} h_{11}(t,S,y) + \epsilon h_{12}(t,S,y) + \cdots,$$

where $h_{10}(t,S)$ satisfies (6.162), $h_{11} \equiv 0$ for all S, $h_{12}(t,S)$ satisfies

$$\begin{aligned} L_{\mathrm{BS}}[h_{12}(t,S)] + \sigma^2 \langle F(y) \rangle S^3 h_{10_{SS}}^2 = 0, & \qquad S < S_0(t), \\ h_{12}(t,S,y) = 0, & \qquad S \geq S_0(t), \\ h_{12}(t,S_0(t),y) = 0, \quad h_{12_S}(t,S_0(t),y) = 0, \quad h_{12_S}(t,S_0(t),y) = 0, & \end{aligned} \tag{6.163}$$

and $h_{13}(t,S)$ satisfies

$$\begin{aligned} L_{\mathrm{BS}}[h_{1_3}] + \kappa_1 S^3 h_{1_{0SS}}^2 + \kappa_2 S^4 h_{1_{0SS}} h_{1_{0SSS}} = 0, & \qquad S < S_0(t), \\ h_{13}(t,S,y) = 0, & \qquad S \geq S_0(t), \\ h_{13}(t,S_0(t),y) = 0, \quad h_{13_S}(t,S_0(t),y) = 0, \quad h_{13_S}(t,S_0(t),y) = 0, & \end{aligned} \tag{6.164}$$

where κ_1 and κ_2 are defined in equations (6.49) and (6.50). With the appropriate alterations to the payoff function and corresponding boundary conditions and the transformation $\tilde{V} \to -\tilde{V}$ we can also value the bid side as well as the put forward contract.

6.6.1 Solving for the American forward Price

The method for solving the approximate pricing problem for the American forward is as follows:

1. We numerically solve the free-boundary problem (6.162) for the leading-order component of the contract's value. This is exactly the American forward free-boundary problem in the Black-Scholes model solved in Section 4.4.

2. Using the leading-order solution and the free boundary determined in the first step, we then compute the first correction from (6.163). Note that this is a fixed-boundary

problem as we have only approximated the free boundary by its leading-order term and this was determined in the first step.

3. Using the leading-order solution, first correction and free boundary determined in the first two steps we then compute the second correction from (6.164). Again, as the free boundary was calculated in step 1, this is a fixed boundary problem.

Solutions of the approximate pricing problem for the American call and put forward using the above method have been generated with an explicit finite difference routine. While not very elegant, the algorithm is simple and stable for a small enough timestep; as only a small number of simulations have been necessary, this timestep restriction has not been a concern.

To compare results with those of Chapter 4 we will use $r = 0.10$, $D = 0.09$, $\sigma = 0.3$, $T - t = 0.75$, and $\epsilon = 10^{-2}$ (corresponding to $\bar{\lambda} = 10^{-2}$). To fully characterize the problem, however, we also need values for $\langle F(y) \rangle$, κ_1, and κ_2. Since we have already defined $\epsilon = \bar{\lambda}$ we can simply choose $\langle F(y) \rangle = 1$.

To approximate κ_1 we note firstly that, as mentioned previously, it contains the parameters B and ρ and is therefore a direct result of the liquidity risk of the asset. Secondly, we note that $h_{1_{0SS}}^2$ is effectively a transaction cost term (*i.e.* Γ^2). If the $h_{1_{0SS}} h_{1_{0SSS}}$ term was negligible then h_{1_2} would be entirely driven by the $h_{1_{0SS}}^2$ term, but since it originates from the liquidity risk of the asset we expect it to increase the spread on the derivative (*i.e.* raise the ask price and lower the bid price) and the only means by which this is guaranteed is if $\kappa_1 > 0$. As $\kappa_1 = O(1)$, for simplicity we will simply use $\kappa_1 = 1$ in our numerical results.

Finally, returning to the definitions of κ_1 and κ_2 in (6.49) and (6.50), we note that if $\rho = 0$ then $\kappa_1 = -B\sigma^2 \langle \psi'(y) \Lambda_\lambda(y) \rangle$. But since $\kappa_1 > 0$ this implies $\langle \psi'(y) \Lambda_\lambda(y) \rangle < 0$. Now since the market price of market depth risk must be strictly positive this further implies that $\psi'(y)$ must be negative for at least some y. While this fact does not guarantee $\langle \psi'(y) \rangle < 0$ it does indicate that it may be either positive or negative. Furthermore, due to the σ^3 factor within κ_2, we will use slightly smaller values for this parameter; specifically $\kappa_2 \in [-0.1, 0.1]$.

In Figure 6.2 we show the first correction, ϵh_{1_2}, for the bid and ask prices of both the call and put forwards. As we would expect, the results shown in Figure 6.2 are virtually identical to those in Figure 4.4 as both show the leading-order liquidity correction to the Black-Scholes value of the contract. This result makes sense since, as we have shown in this

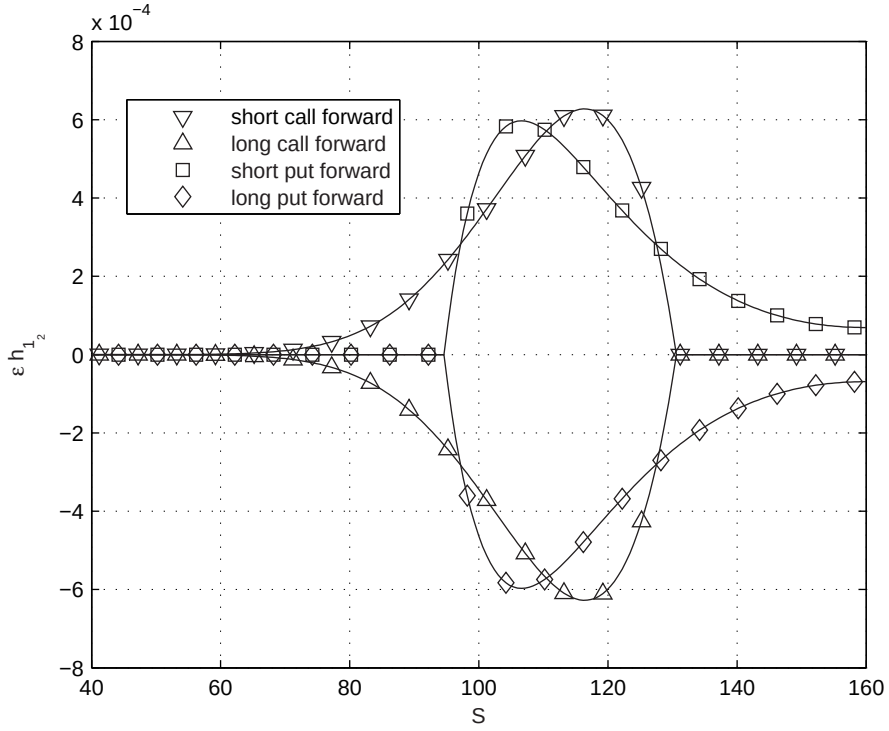


Figure 6.2: The first correction, ϵh_{1_2}, to the bid and ask prices of both the call and put forward contracts. Parameter values are $r = 0.10$, $D = 0.09$, $\sigma = 0.3$, $T - t = 0.75$, $\tilde{F} = 100$, $\epsilon = 0.01$, and $\langle F(y) \rangle = 1.0$.

chapter, the leading-order correction is an average liquidity effect.

In Figure 6.3 we show the second correction, $\epsilon^{\frac{3}{2}} h_{1_3}$ to the Black-Scholes value of the bid and ask prices of both the call and put forwards. As we determined in the derivation of the pricing framework of this chapter, the effects of the underlying's liquidity risk first enter the valuation through h_{1_3}. From Figure 6.3 we can now see how the liquidity risk affects the price. Firstly, we notice that the shape of the second correction is very similar to the corresponding first correction. At the free boundary the correction is zero and then increases rapidly as we move into the hold region until its peak around the value of S where the Γ of the contract is largest. As we move further into the hold region the correction then decreases to zero. Because of the strong presence of $h_{1_{0SS}}$ in (6.164) this similarity in form between the first two corrections is understandable and, furthermore, indicates the importance of the $h^2_{1_{0SS}}$ contribution as a source for the second correction for the present

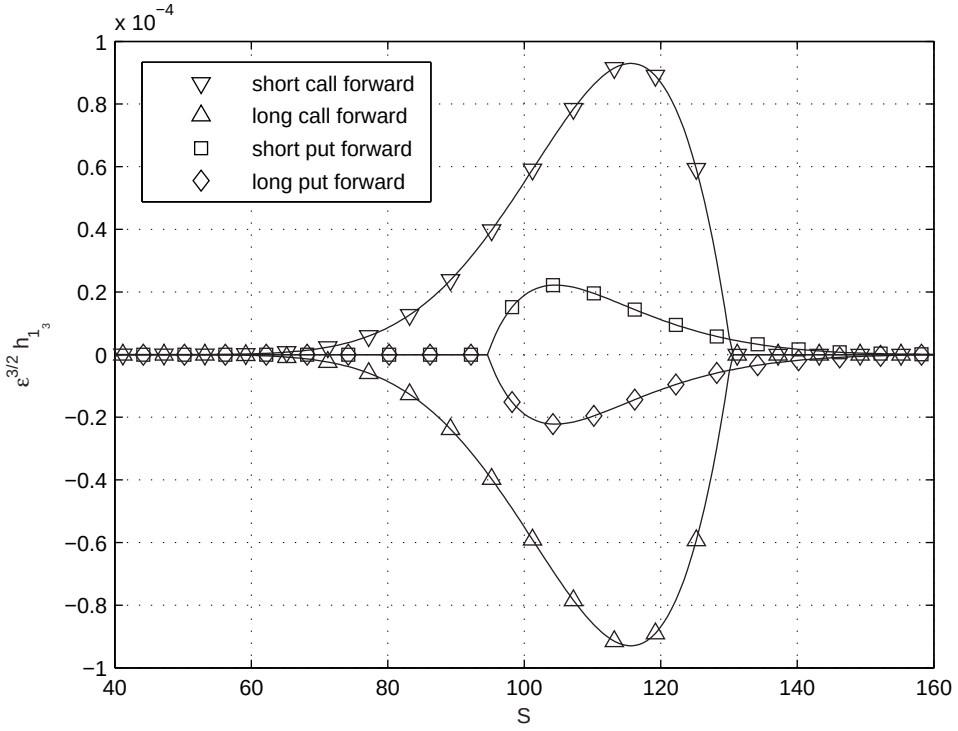


Figure 6.3: The second correction, $\epsilon^{\frac{3}{2}} h_{1_3}$, to the bid and ask prices of both the call and put forward contracts. Parameter values are $r = 0.10$, $D = 0.09$, $\sigma = 0.3$, $T - t = 0.75$, $\tilde{F} = 100$, $\epsilon = 0.01$, $\langle F(y) \rangle = 1.0$, $\kappa_1 = 1.0$, and $\kappa_2 = 0.1$.

choice of the parameter values. It is also clear that, as expected, the bid-ask spread in the American forward widens in the presence of liquidity risk in the underlying. Furthermore, this increase in the spread is largest where the Γ of the contract is largest. This result is intuitive as liquidity risk should impact the price most in areas where rehedging is most frequent.

Another noticeable feature of Figure 6.3 is the difference in size between the second correction to the value of the call forward and that to the put forward. To understand this disparity we need to examine the dependence on κ_2 of the second correction. In Figures 6.4 and 6.5 we show the correction to both the bid and ask prices of the call and put forwards, respectively, for several values of κ_2. It is clear from these figures that while the magnitude of the liquidity risk correction increases with κ_2 for the call forward, it is a decreasing function of the parameter for the put forward. This asymmetric behaviour is a result of the

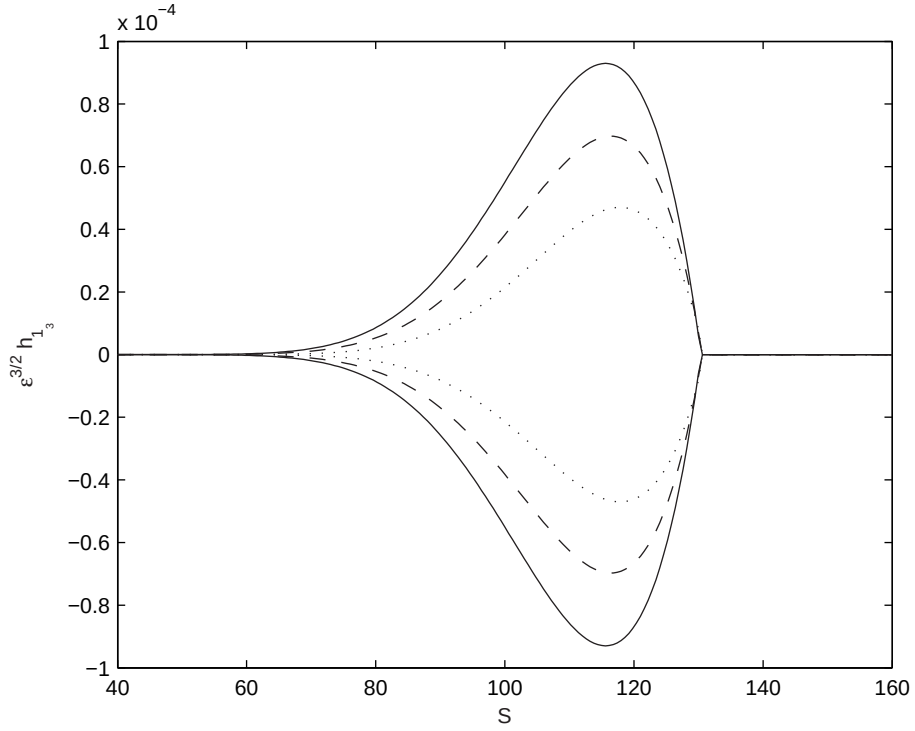


Figure 6.4: Dependence of the second correction, $\epsilon^{\frac{3}{2}}h_{1_3}$, to the bid and ask prices of the call forward on the parameter κ_2. Parameter values are $r = 0.10$, $D = 0.09$, $\sigma = 0.3$, $T - t = 0.75$, $\tilde{F} = 100$, $\epsilon = 0.01$, $\langle F(y) \rangle = 1.0$, and $\kappa_1 = 1.0$. Dotted lines represent $\kappa_2 = -0.1$, dashed lines represent $\kappa_2 = 0$, and solid lines represent $\kappa_2 = 0.1$. Curves above $h_{1_3} = 0$ are ask price corrections and below are bid price corrections.

$h_{1_{0SSS}}$ component of the κ_2 term which quantifies the rate of change of the contract's Γ, or curvature, with S. For both the call and put forward, deep in the hold region the contract will behave very much like the European equivalent and will be nearly linear. Near the free boundary, however, there will be a significant amount of curvature in the solution to match up with the form of the payoff function at the free boundary. For the call forward, as S increases towards the free boundary the curvature will therefore steadily increase and $h_{1_{0SSS}}$ will be (at least mostly) a positive function. For the put forward, however, as S increases away from the free boundary the curvature of the solution will decrease and $h_{1_{0SSS}}$ will, in general, be negative. Alternatively $h_{1_{0SSS}}$ can be thought of as the skew of the leading-order component of the contract's value. In this case, because of the call forward's free boundary

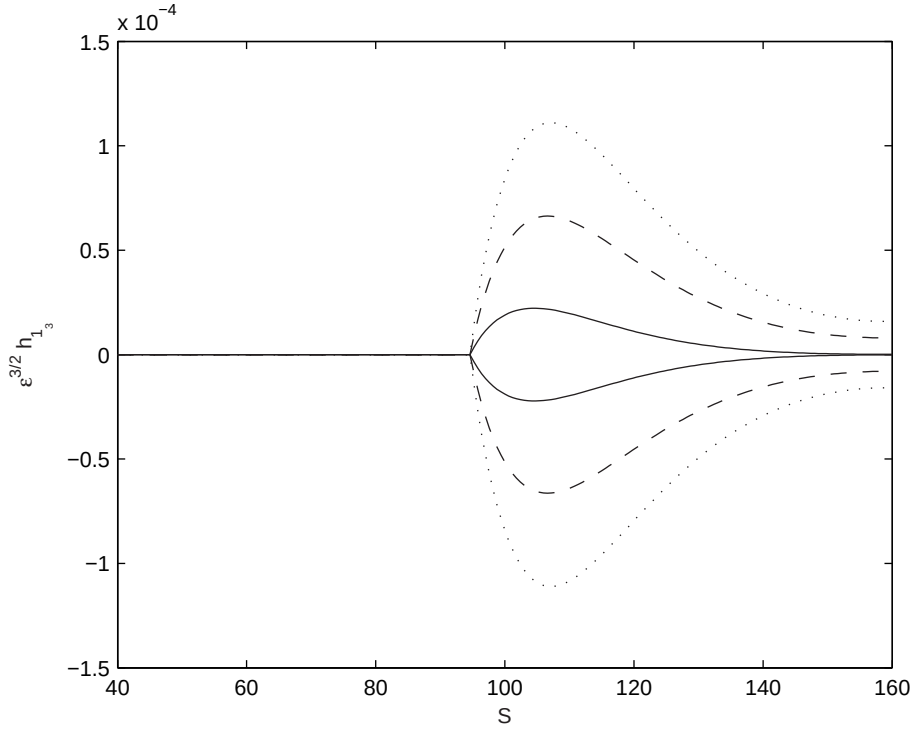


Figure 6.5: Dependence of the second correction, $\epsilon^{\frac{3}{2}} h_{1_3}$, to the bid and ask prices of the put forward on the parameter κ_2. Parameter values are $r = 0.10$, $D = 0.09$, $\sigma = 0.3$, $T - t = 0.75$, $\tilde{F} = 100$, $\epsilon = 0.01$, $\langle F(y) \rangle = 1.0$, and $\kappa_1 = 1.0$. Dotted lines represent $\kappa_2 = -0.1$, dashed lines represent $\kappa_2 = 0$, and solid lines represent $\kappa_2 = 0.1$. Curves above $h_{1_3} = 0$ are ask price corrections and below are bid price corrections.

being on the right hand side of the hold region, the leading-order solution is very heavily positively skewed and the converse for the put forward.

The final notable aspect of Figures 6.3, 6.4, and 6.5 is that the liquidity risk correction is symmetric for the bid and ask positions. As the two terms driving this correction depend on the leading-order solution only (*i.e.* through $h_{1_{0SS}}^2$ and $h_{1_{0SS}} h_{1_{0SSS}}$) and since no bid-ask spread exists in the leading-order solution, this result is understandable. For the second correction in the other two regions of the (c, d) space, however, there is a term in the equations for the second correction (see equations (6.72) and (6.100)) driven by the first correction. As a bid-ask spread exists in the first correction we expect that this symmetry between the bid and ask positions will be broken in this regions.

Using the pricing framework we have solved for the bid and ask forward prices of both the call and put forwards; results of these calculations along with the premium above the constant liquidity prices (calculated as the forward price using only the leading-order solution and first correction) are shown in Table 6.1. Given the above discussion, the

	$\tilde{F}$(call forward)	$\tilde{F}$(put forward)
bid (long)	100.173340430 $(-2.8120\,10^{-5})$	99.1934853270 $(1.05672\,10^{-5})$
ask (short)	100.1740031110 $(2.8123\,10^{-5})$	99.1923976112 $(-1.05669\,10^{-5})$

Table 6.1: Short and long forward prices, $\tilde{F}$, per share for the American call and put forwards with $r = 0.10$, $D = 0.09$, $\sigma = 0.3$, $T - t = 0.75$, $\langle F(y) \rangle = 1.0$, $\kappa_1 = 1.0$, $\kappa_2 = 0.1$, and $S_0 = 99.2528055$. The premium above the corresponding constant-liquidity forward is shown in brackets.

results are fairly predictable; in the presence of liquidity risk the spread on the forward price increases by a small correction above the constant-liquidity correction. Furthermore, for the values of κ_1 and κ_2 used, the premium for the call forward is noticeably larger than that for the put forward in keeping with the results shown in Figure 6.3.

6.7 European Call Option with Liquidity Risk

In the previous section we presented numerical results for the pricing of the American forward on an asset with liquidity risk. While these results gave a good indication of how liquidity risk impacts the pricing of a contingent claim, the American forward is not a standard contract and we have very little to reference the results to. In this section we will give a very brief treatment of the application of the stochastic liquidity pricing framework to a far more common contract, a European call option.

As was the case with the American forward, we will not calibrate the model to market data, but simply use the approximate and reasonable model parameter values used in the previous section. We will also only focus on the ask price of the contract (*i.e.* the short position); the properties of the bid-ask spread generated by the liquidity risk model have been thoroughly analyzed with the American forward and we will not concern ourselves with the subject again.

As with the American forward we will only analyze the case $c = 1$, $d = 1$; for the short position of a European call option the pricing problem is therefore given in (6.54) with boundary conditions

$$\begin{aligned} \tilde{V}(T,S) &= \max(S-K,0), \\ \tilde{V}(t,S) &\sim S, \qquad\qquad S \to \infty, \\ \tilde{V}(t,0) &= 0, \end{aligned} \tag{6.165}$$

where K is the contract's strike.

We have solved for the price in the stochastic liquidity model using our standard explicit finite-difference algorithm. In Figure 6.8 we show the first correction, ϵh_{1_2}, to the Black-Scholes price in the stochastic liquidity model and, for comparison, the total correction to the Black-Scholes price, $\tilde{V}_{\text{BH}} - V_{\text{BS}}$, for the same contract in the constant liquidity model which has been calculated numerically from equation (2.18) using an explicit finite-difference routine and the boundary conditions (6.165). As expected, the first correction of the stochastic liquidity model is sharply peaked near strike where the call option's Γ is largest. Furthermore, the correction falls to zero much more sharply as we move out of the money than as we move into the money. In general we see that the first correction reproduces the total correction of the constant liquidity to the Black-Scholes model well.

The major difference between the two curves in Figure 6.6 occurs near strike where their effects are largest. To explain this we show the liquidity risk (second) correction to the Black-Scholes price, $\epsilon^{\frac{3}{2}} h_{1_3}$, in Figure 6.7. As has been discussed previously, the second correction is also driven by terms proportional to Γ which explains the fact that $\epsilon^{\frac{3}{2}} h_{1_3}$ is sharply peaked near strike and therefore why ϵh_{1_2} differs most significantly from the constant liquidity model near its peak.

Finally, we would like to quantify the difference between the price of the European call in the stochastic liquidity model and that from the constant liquidity model. To do this we return to the concept of implied market depth introduced in Section 6.5. To reiterate, the implied market depth is the market depth parameter, $\bar{\lambda}$, required in the constant liquidity pricing equation (2.18) to reproduce the market price of the contract. We will now assume the the stochastic liquidity model generates the true market price and find what market

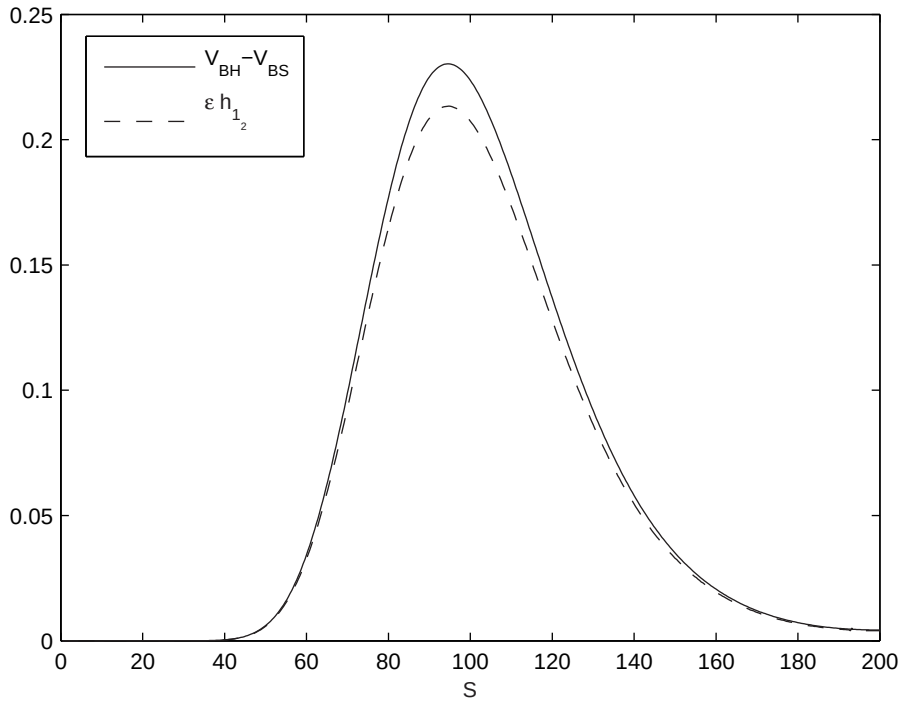


Figure 6.6: The dominant liquidity correction to the Black-Scholes ask price of a European call option with $r = 0.10$, $\sigma = 0.3$, $D = 0.0$, $T - t = 0.75$, $\langle \bar{f}(y) \rangle = 0.01$, $\alpha = 0.5$, $\kappa_1 = 1.0$, and $\kappa_2 = 0.1$. The solid line is the total correction for the constant liquidity model. The dashed line is the first correction for the stochastic liquidity model.

depths are implied by the constant liquidity model for a range of strikes; that is, we will find the implied market depth curve from the stochastic liquidity model.

To generate the "market" prices we used the same parameter values as have been used for the other results in this section. Solving the system (6.54) we obtain an ask price for the call option. For an initial guess of $\bar{\lambda}$ the $\gamma = 0$ pricing equation was then solved for the same parameter values. Comparing the two prices, the value of $\bar{\lambda}$ was then perturbed and the $\gamma = 0$ equation was then solved again. This procedure was repeated until a value of $\bar{\lambda}$ was chosen that generated a price from the constant liquidity model which agreed with the market price to six decimal places; this is the implied market depth. This process was then repeated for a range of strikes with a fixed spot price; the implied market depth curve generated by this procedure is shown in Figure 6.8. There are two interesting

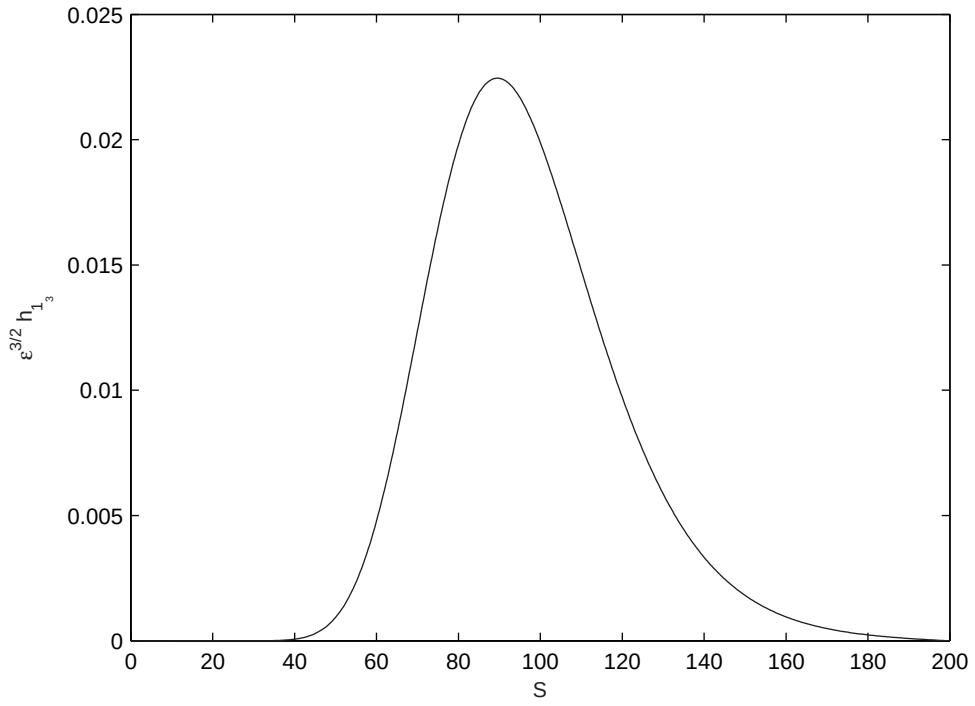


Figure 6.7: The second (liquidity risk) correction to the Black-Scholes ask price of a European call option with $r = 0.10$, $\sigma = 0.3$, $D = 0.0$, $T - t = 0.75$, $\langle \bar{f}(y) \rangle = 0.01$, $\kappa_1 = 1.0$, and $\kappa_2 = 0.1$.

characteristics to the curve in Figure 6.8. First is the fact that implied market depths range from approximately 0.011 to 0.012 for an average market depth, $\langle \bar{f}(y) \rangle = 0.01$; this result tells us that the liquidity risk always adds a positive premium to the price of the contract equivalent to 10% to 20% of the market depth. Second, if we regard the stochastic liquidity prices as "real" then not only does the constant liquidity model underprice the call option, but this underpricing becomes progressively worse as we move out of the money. A possible reason for this result is that, as can be seen in Figure 6.6, the call's Γ is much smaller out of the money than in the money. Since liquidity effects are closely linked to the contract's Γ they will have less of an impact out of the money than in. But since liquidity costs will be less out of the money, it is reasonable to expect a *change* in the level of the liquidity costs should have a proportionately greater affect in this region than in the money. The upward sloping nature of the implied market depth curve may be reflecting this effect.

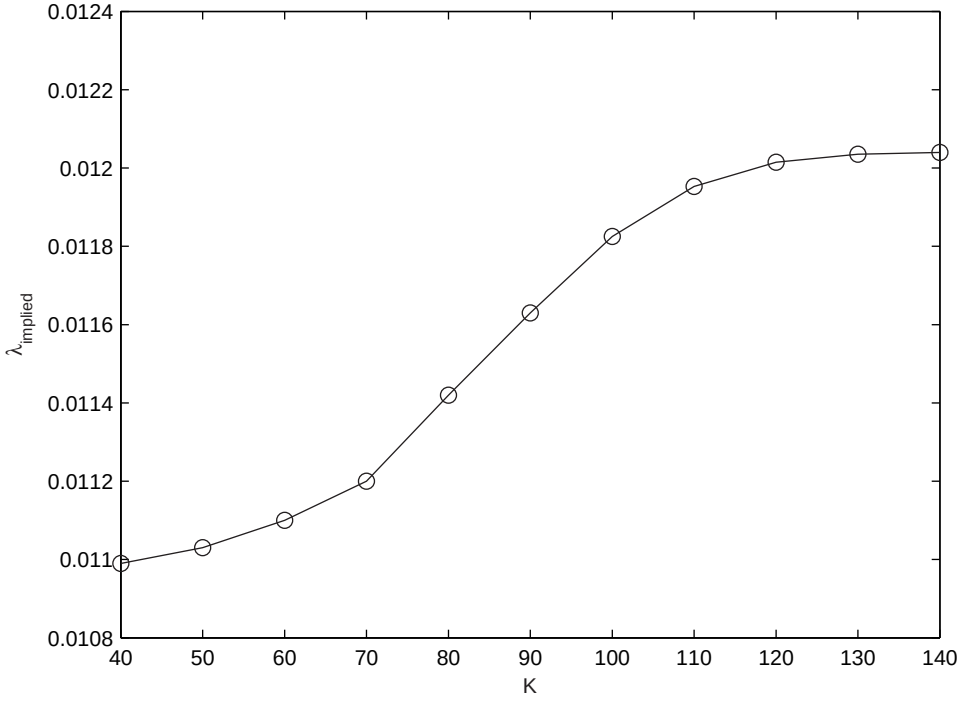


Figure 6.8: The implied market depth as a function of strike, K, for the ask price of a European call option with spot price $S_{\text{spot}} = 100$ and $r = 0.10$, $\sigma = 0.3$, $D = 0.0$, $T - t = 0.75$, $\langle \bar{f}(y) \rangle = 0.01$, $\kappa_1 = 1.0$, and $\kappa_2 = 0.1$.

6.8 Conclusions

In this chapter we have extended the asymptotic methods developed by Fouque, Papanicolaou, and Sircar to analyze the pricing problem for a European vanilla derivative contract contingent on an asset possessing liquidity risk. In keeping with the work of previous chapters, we have assumed a negligible bid-ask spread on the underlying and instead focussed only on the effects of a finite market depth.

To incorporate liquidity risk into the model we have exogenously specified the market depth as being generated by an Ornstein-Uhlenbeck process which crucially incorporates the mean-reverting behaviour of the parameter observed in real markets. By assuming not only that the size of the market depth is small, but also that its period of mean-reversion is small compared to the time to maturity of the derivative, we were able to rescale the pricing

equation and apply asymptotic techniques to simplify the pricing framework. Because of the interplay between the two small parameters in the problem our first significant result was the discovery that there actually exist three regions of the parameter space in which the solution differs to the level of the first liquidity-dependent correction.

In all three regions it was found that the leading-order component of the solution is exactly the Black-Scholes value of the contract. Furthermore, the first correction is always the typical $O(\lambda)$ constant-liquidity first correction, but with the market depth given by the distributional average of the market depth process.

The variability of the solution across the three regions of the parameter space enters through the second correction when $c \leq 2d$ and the third correction when $c > 2d$ when the liquidity risk contribution becomes significant. In all cases this liquidity risk enters through two terms; one driven by $h_{0_{SS}}^2$ and one by $h_{0_{SS}} h_{0_{SSS}}$, where h_0 is the leading-order solution. The $h_{0_{SS}}^2$ term effectively acts as a transaction cost term and always acts to increase the spread on the price of the contract. The $h_{0_{SS}} h_{0_{SSS}}$ term is driven not only by the Γ of the contract, but also by its skew; depending on the payoff structure of the contract this term may therefore be positive or negative thus adding to or detracting from the increase in the spread caused by the $h_{0_{SS}}^2$ term.

The difference in the solution between the three regions occurs mainly as a result of the relative size of the second (liquidity risk) correction to the leading-order solution. As the rate of mean-reversion of the market depth process increases relative to its size (that is, c grows relative to d) we find that the liquidity risk correction decreases in significance relative to the first correction caused by constant-liquidity effects. This result is intuitive as a higher rate of mean reversion effectively makes the process behave more like a constant, averaged process and so the effects due to the fluctuations become less important.

Extending the work of Fouque, Papanicolaou, and Sircar further, we then developed a relatively simple method of calibrating the pricing framework to market data. Introducing the concept of the implied market depth and using further asymptotic analysis, we showed that, by regressing the implied market depth against both the moneyness and time to expiry of a derivative contingent on the underlying, we can obtain estimates for the two universal liquidity risk parameters, κ_1 and κ_2, generated by the approximate pricing framework.

Finally, we ended the chapter by applying the pricing framework to two examples.

In the first of these we repriced the American forward contract introduced in Chapter 4. Extending the stochastic liquidity model to accommodate the early exercise feature, it was found that the same pricing equations that hold for a European contract also hold for an American contract in its hold region. Performing numerical calculations, we have found firstly that the dominant correction to the Black-Scholes value is virtually identical to that in the constant-liquidity model; this is to be expected as the dominant correction is the same (with $\bar{\lambda} = \langle \bar{f}(y) \rangle$) in both models. Secondly, in general, the liquidity risk (second) correction acted to increase the spread between the bid and ask prices of both the put and call contracts. Interestingly, while this spread was found to increase with increasing κ_2 for the call forward, it was found to be a decreasing function of the parameter for the put forward. This result was explained in terms of the free-boundary drive skew of the contract's value function which is positive for the call forward, yet negative for the put forward.

In the second numerical example we priced a simple European call option to determine the effects of stochastic liquidity on a well-known contract. As was the case with the American forward we found that both the first and second correction to the Black-Scholes price of the contract are heavily peaked near the strike of the contract and they are both positive for all values of the asset price. Isolating the liquidity risk component of the model by comparing the results of the stochastic liquidity model to those of the constant liquidity model we found, in general, that the constant liquidity model underprices the contract and this effect increases the further we move out of the money. This last result has been explained using the fact that the contract's Γ is much smaller out of the money and so liquidity risk will impact proportionately more in this region.

Chapter 7

Minimizing Transaction Costs in a Finitely Liquid Market

We will now change our focus and investigate how to optimally execute a portfolio transaction in a finitely liquid market in a finite horizon economy. By a portfolio transaction we simply mean a trade that alters the composition of a portfolio of assets from one state to some desired final state.

There are many ways of carrying out a portfolio transaction; with each of these is associated a particular *strategy* which will be defined formally below, but for now we will simply take to mean a sequence of traded quantities for each asset that comprise the total transaction. When we consider how to optimally move a portfolio from one state to another we must immediately ask what we mean by an 'optimal' strategy. As the transaction is to be carried out over a finite time interval and each trade during this time will incur a fixed cost and will affect all future asset prices through the market impact effect, there are many possible ways for the transaction to be executed and each one will have a unique cost. One seemingly obvious measure of a strategies' performance is therefore the cost of the transaction and the associated optimal strategy might be the one that minimizes this cost.

While the strategies' cost is a good measure of its performance, it is not ideal as it neglects a key aspect of the trade. To see this we will briefly consider two strategies for executing a sell order. In the first of these we simply submit a market order for the entire

quantity that we wish to sell (*i.e.* we sell everything at the beginning of the interval). With this strategy the assets are sold as quickly as possible, but if the number of assets is large then the depression of the selling price caused by market impact can be significant and the amount recovered from the trade might be quite low. For the second strategy we split the trading horizon into a finite number of intervals and then execute the transaction in equal parcels over each of these intervals. Because of the smaller quantity sold during each of these intervals and the convex nature of the price impact effect, the total depression of the asset price over these multiple smaller trades will be less than that caused by the large trade of the first strategy which will therefore result in a lower 'cost' of this sell strategy. (*i.e.* more of the book value of the portfolio will be recovered).

With the transaction's cost as the sole measure of its associated strategies' performance it is clear that the second strategy will be more nearly optimal than the first. But often these transactions are carried out by traders (for instance, fund managers) who must be very careful of the risk that they take on. While we expect the second strategy to have a lower cost than the first, there is much more uncertainty, and therefore risk, associated with the second strategy as there is much more uncertainty in the asset price during the latter trades of the strategy. To incorporate this *time risk* aspect of the strategy a better way of optimizing the total transaction's performance is therefore to choose a strategy that minimizes the expected transaction costs while simultaneously minimizing its uncertainty, or risk. To investigate the structure of the optimal strategy when this time risk is taken into account is one of our main goals and will be the focus of Chapters 8 and 9.

7.1 Previous Literature

Compared with the very similar problem of optimal portfolio selection (see, for example [38] and [35]), the problem of how to optimally execute a portfolio transaction has received very little attention in academic circles.

In [6], Bertsimas and Lo ignore the time risk of the transaction and derive, through the use of stochastic dynamic programming, trading strategies that minimize the expected cost of the transaction. Even though they focus only on cost-minimization, in order to generate explicit, closed-form solutions they need to make simplifying assumptions such as a linear market impact function and only either a temporary or and permanent market impact

effect, but never both. Even with this relatively simple set-up they find some interesting results. Firstly, when the the market impact of a trade only affects the asset price through a permanent and linear mechanism and the dynamics of the asset price are generated by an Arithmetic Brownian Motion the optimal strategy is then simply what they refer to as the *naive strategy*; in this case the total transaction is broken-up into equal sized parcels and traded homogeneously over time.

Secondly, they examine the situation in which the asset price is generated by a Geometric Brownian Motion and the trading activity, through a temporary impact mechanism, has a direct effect on the cost of the strategy, but no affect on the permanent dynamics of the asset price. In this situation it is found that the optimal amount traded during each period is the parcel size of the naive strategy with an additional constant that is independent of the asset price. In fact they find that the optimal strategy is *dynamic* (*i.e.* depends on quantities that can only be determined at the time of a trade) only when serial correlations exist between successive asset prices. In the absence of these correlations they find that the optimal strategy is *static* and can therefore be determined with all information available before the transaction begins.

In [2], Almgren and Chriss use an asset price process generated by an Arithmetic Brownian Motion (ABM) with independent increments and linear forms for both the temporary and permanent market impact functions. In doing so they are able to derive closed-form solutions to the optimal mean-variance problem (*i.e.* the problem of simultaneously minimizing both the expectation and variance of the cost of the transaction) and show that for each unique value of the trader's level of risk-aversion there is a unique optimal trading strategy. Plotting the expectation of the cost vs. its variance for each of these optimal strategies, they have been able to assemble the *efficient frontier* for the transaction. Furthermore, because of the independent increments of the price process and the symmetry of their risk function each increment of their cost function is independent and can be optimized independently; as a result they generate optimal trading strategies that are strictly static. Finally, for a 'sell' ('buy') transaction it is shown that the optimal strategy is monotonically decreasing; that is, it consists entirely of individual sell orders.

7.2 Outline of work

The aim of this and the following two chapters is to develop analytic, closed-form solutions to the mean-variance optimal portfolio problem using the technique of stochastic dynamic programming. As with all previous work in this thesis, we will model the dynamics of the underlying asset price using the Bakstein-Howison liquidity model and this will allow us to extend the previous literature in several critical ways.

Firstly, whereas the works by both Bertsimas and Lo [6] and Almgren and Chriss [2] were based upon an asset price model generated by an Arithmetic Brownian Motion, our discrete-time, binomial model will be calibrated to a Geometric Brownian Motion thereby eliminating the possibility of negative prices and making our results more applicable to longer trading horizons.

Secondly, the market impact function of our asset price model contains a nonlinear temporary impact mechanism, a nonlinear permanent impact mechanism, and a fixed-cost bid-ask spread effect. While we will ignore the bid-ask spread effect and linearize the impact functions in order to focus on leading-order solutions in this work, we will see that the method developed here to solve for the leading-order solutions will also work with the more general forms of these trading effects.

Finally, and most importantly, the feedback introduced into the asset price process through our trading mechanism automatically creates strong serial correlations between asset prices. While we will see that this effect causes difficulties in solving for the optimal trading strategies in some situations, it is precisely the quality claimed in [2] necessary to generate a dynamic strategy. With our model and the use of dynamic programming we will be able to test this claim.

So as to focus on the qualitative features of the results, we will only examine the optimal *liquidation* problem (that is, a complete sell order) and for one type of asset only. While this may seem restrictive, it will become clear that, with only a small amount of extra work, the analysis easily generalizes in both of these respects.

We will find that solving for optimal trading strategies in the full mean-variance problem involves several significant difficulties compared to the problem of minimizing only the expected cost (and ignoring the variance, or time risk) of the transaction. To set up the more difficult mean-variance problem (which will be treated in chapters 8 and 9), we will

devote the rest of this chapter to formulating and solving the simpler problem of finding the optimal trading strategies when only minimizing expected cost.

The rest of this chapter is organized as follows. In Section 7.3 we briefly restate and review some of the key features of the Bakstein-Howison liquidity asset price model. In Section 7.4 we will then formulate the cost-minimizing optimization problem and using this we will then derive the equivalent Bellman equation formulation to the problem in Section 7.5. In Section 7.6 we will solve the Bellman equation for the leading-order component of the optimal trading strategy and then finally conclude in Section 7.7 with a discussion of qualitative features of the solution and a comparison of the results with those in [6] and [2].

7.3 The Bakstein-Howison Liquidity Asset Price Model

In this section we will rederive the $\gamma = 0$ BH asset price model in order to review several of its key features that will play an important role in the optimal execution problem. This review is not exhaustive; for a complete description of the model we refer the reader back to Section 1.5.

We take ourselves to be the large trader described in Section 1.5 and imagine that we are at a specific node of an asset price tree at time t when the asset price is $S_t(\omega_t)$ and we hold $H_t(\omega_{t-\delta t})$ assets. At this moment we execute a trade of $\delta H_t(\omega_t) \equiv H_{t+\delta t}(\omega_t) - H_t(\omega_{t-\delta t})$ assets where ω_t is the state of the system at the end of the previous trading interval (*i.e.* the specific path the asset price had taken to arrive at its current value) and $\delta H_t(\omega_t)$ is predictable *w.r.t.* ω_t. For a general trading strategy in our model the asset price tree is, in general, non-recombining; the quantities S_t, H_t, $H_{t+\delta t}$, and δH_t are therefore path-dependent and the state of the system that generated them is crucial. We note that we are being somewhat 'loose' with our notation for the state of the system as to be completely accurate a quantity generated by the information at t is actually dependent on the entire sequence $\omega_1, \ldots \omega_{t-\delta t}, \omega_t$, but as we will show in the next section, this path-dependency can be ignored when solving for the optimal trading strategy.

As the binomial component of our asset price process is calibrated to a Geometric Brownian Motion we model the market impact function as a multiplicative effect. The temporary market impact is shown through the average price per asset, $\bar{S}_t$ which, for a

trade in a market where the bid-ask spread is negligible, is written as

$$\bar{S}_t(\omega_t) = S_t(\omega_{t-\delta t})e^{\lambda \delta H_t(\omega_t)}, \tag{7.1}$$

λ, as usual, is a measure for the depth of the asset's market. In response to this trade, the background traders adjust their outstanding orders which results in a permanent shift in the asset price so that

$$S_t(\omega_{t-\delta t}) \to S_t(\omega_{t-\delta t})e^{\lambda(1-\alpha)\delta H_t(\omega_t)} \tag{7.2}$$

after our trade is completed. As before, α is a parameter that measures the 'believability' of our trade by the background and thereby determines how much permanent impact on the asset price that the trade will have. Finally, the background traders complete their trades for the trading interval. The price change due to this activity we model as a binomial process; the asset price moves up by a multiplicative factor $u > 1$ with probability p and down by a factor $0 < d < 1$ with probability $1 - p$ over a time of δt. The asset price dynamics over an entire trading interval are therefore

$$S_{t+\delta t}(\omega_t) = \begin{cases} u \cdot S_t(\omega_{t-\delta t})e^{\lambda(1-\alpha)\delta H_t(\omega_t)} & \text{probability } p, \\ d \cdot S_t(\omega_{t-\delta t})e^{\lambda(1-\alpha)\delta H_t(\omega_t)} & \text{probability } 1-p. \end{cases} \tag{7.3}$$

With respect to the problem of determining the optimal strategy for a portfolio transaction, there are three important characteristics of the asset price process. First, because of the way in which the market impact has been modeled both the temporary and permanent effects are simply included in the one impact term. Second, the trade size affects the cost of the trade (through the temporary impact mechanism) and the future asset prices through the convex exponential market impact function. Finally, serial correlations are introduced into the price process as it is clear that $S_{t+\delta t}(\omega_t)$ is strongly dependent on $\delta H_t(\omega_t)$. If our trade begins at $t = 1$ when $S = S_1$ and we assume that $\delta H_1 = \delta H_1(S_1)$, then this asset price dependence propagates through the entire asset price tree; we will see that this is indeed the case when we attempt to solve the mean-variance problem.

7.4 Formulation of the Cost-Minimization Problem

We will concern ourselves with finding the optimal trading strategy over a finite horizon of length T. We divide this interval into N trading periods each of length δt so that $N\delta t = T$ and index these periods with the variable $n \in \{1, \ldots, N\}$. Furthermore, for ease of notation we will drop the δt from all quantities' notation so that, for example, $S_{n\delta t}(\omega_{n-1}) = S_n(\omega_{n-1})$, $\delta H_{n\delta t}(\omega_n)) = \delta H_n(\omega_n)$, etc. We now formally define a *trading strategy* as the sequence, $\{\delta H_1, \delta H_2(\omega_1), \ldots, \delta H_N(\omega_{N-1})\}$, of traded quantities that make up the entire portfolio transaction.

As we showed in the previous section the average price per asset for a trade of size $\delta H_n(\omega_n)$ is $S_n(\omega_{n-1})e^{\lambda \delta H_n(\omega_n)}$; since the trade consists of $\delta H_n(\omega_n)$ assets the total 'cost' (which will be negative for a sell order), $\tilde{C}_n$, will be

$$\tilde{C}_n = \delta H_n(\omega_n) S_n(\omega_{n-1}) e^{\lambda \delta H_n(\omega_n)}.$$

Simply summing each of these costs then gives us the entire cost, $\tilde{C}_{\text{total}}$, of the strategy, $\{\delta H_n(\omega_{n-1})\}$ for $n = 1, \ldots, N$; this is

$$\tilde{C}_{\text{total}} = \sum_{n=1}^{N} \tilde{C}_n. \tag{7.4}$$

But as it stands, equation (7.4) contains N cash streams all generated at separate times; to give each an equal weight *w.r.t.* its time value of money we discount each stream to the first period when $n = 1$. Let $C_n = \tilde{C}_n e^{-(n-1)r\delta t}$ be the discounted cost of the n^{th} trade; the total discounted cost of the strategy, C_{total}, is then

$$C_{\text{total}} = \sum_{n=1}^{N} C_n = \sum_{n=1}^{N} \delta H_n(\omega_n) S_n(\omega_{n-1}) e^{\lambda \delta H_n(\omega_n) - (n-1)r\delta t}. \tag{7.5}$$

It is this *objective function*, C_{total}, that we wish to minimize; the expected-cost minimization problem is therefore simply

$$\min \sum_{\substack{\{\delta H_n(\omega_n)\} \\ n=1,\ldots,N}} E_1 \left[C_{\text{total}} \right], \tag{7.6}$$

where E_i is the expectation operator taken *w.r.t.* the information available at the beginning

of the i^{th} trading period and the control variables are each of the δH_n.

7.5 Derivation of the Equivalent Bellman Equation Formulation

We wish to solve equation (7.6). The argument, $E_1\left[C_{\text{total}}\right]$ is convex in each of its N control variables and so we should be able to find the minimum simply by setting the partial derivative of this argument *w.r.t.* each of these control variable equal to zero and solving each of the resulting N equations. The problem with the direct approach is that each $E_1\left[C_n\right]$ depends on all previous δH (that is, all δH_i where $i = 1, \ldots, n$); finding a closed-form solution to the problem will therefore involve solving N coupled equations which will become very difficult for large N.

To overcome this difficulty we need to remove the cost-minimization problem's explicit dependence on all N control variables which we accomplish by re-expressing it in terms of its equivalent Bellman equation. Towards this end, imagine we are at the beginning of the n^{th} $(n = 2, \ldots, N)$ trading period when the asset price is $S_n(\omega_{n-1})$. We now invoke the *principle of optimality* (see, for example, [17]) which states that any globally optimal solution to the full problem, equation (7.6), must also optimize the system when beginning at any intermediate interval regardless of how the system arrived in that state. For our purpose, the key aspect to this statement is the final requirement; in order to apply the principle of optimality the cost of the total transaction leading up to the n^{th} trading period (where $S_n = S_n(\omega_{n-1})$) must be independent of the path taken to arrive at that point. As our asset price process is, in general, non-recombining, it is not immediately obvious that this requirement holds. But we are concerned with how the system arrives at the *specific* state $(S_n(\omega_{n-1}), H_n(\omega_{n-1}))$. Since the asset price process does not recombine there can have been only *one* path that led to this state and so the requirement must hold and the principle of optimality is valid in this example. As a result we can eliminate the dependence on the path-dependent state of all quantities in the system. Because of this the optimal

solution that we seek must also solve the n^{th} sub-problem which is written as

$$\min_{\substack{\{\delta H_m\}\\ m=n,\ldots,N}} \left\{ E_n \left[\sum_{m=n}^{N} \delta H_m S_m e^{\lambda \delta H_m - (m-1) r \delta t} \right] \right\}. \tag{7.7}$$

We now define the n^{th} *optimal value function*, J_n, as

$$J_n = \min_{\substack{\{\delta H_m\}\\ m=n,\ldots,N}} \left\{ E_n \left[\sum_{m=n}^{N} \delta H_m S_m e^{\lambda \delta H_m - (m-1) r \delta t} \right] \right\}, \tag{7.8}$$

which we can then expand to give

$$\begin{aligned} J_n &= \min_{\substack{\{\delta H_m\}\\ m=n,\ldots,N}} \left\{ E_n \left[\delta H_n S_n e^{\lambda \delta H_n - (n-1) r \delta t} + \sum_{m=n+1}^{N} \delta H_m S_m e^{\lambda \delta H_m - (m-1) r \delta t} \right] \right\}, \\ &= \min_{\substack{\{\delta H_m\}\\ m=n,\ldots,N}} \left\{ \delta H_n S_n e^{\lambda \delta H_n - (n-1) r \delta t} + E_n \left[\sum_{m=n+1}^{N} \delta H_m S_m e^{\lambda \delta H_m - (m-1) r \delta t} \right] \right\}, \end{aligned} \tag{7.9}$$

where the second equality results from the fact that δH_n and S_n are known *w.r.t.* the information available at the beginning of the n^{th} interval. As the first term on the right hand side of (7.9) is independent of δH_m for all $m > n$ we can expand the objective function and write

$$J_n = \min_{\delta H_n} \left\{ \delta H_n S_n e^{-\lambda \delta H_n - (n-1) r \delta t} + \min_{\substack{\{\delta H_m\}\\ m=n+1,\ldots,N}} E_n \left[\sum_{m=n+1}^{N} \delta H_m S_m e^{\lambda \delta H_m - (m-1) r \delta t} \right] \right\}. \tag{7.10}$$

Finally we utilize iterative property of the expectation operator; that is $E_i[E_j[y_k]] = E_i[y_k]$ for any random variable y generated at time k and for any $k \geq j > i$. We can now write the second term on the right hand side of (7.10) as

$$\min_{\substack{\{\delta H_m\}\\ m=n+1,\ldots,N}} E_n \left[E_m \left[\sum_{m=n+1}^{N} \delta H_m S_m e^{\lambda \delta H_m - (m-1) r \delta t} \right] \right] = \min_{\substack{\{\delta H_m\}\\ m=n+1,\ldots,N}} E_n \left[E_m \left[\sum_{m=n+1}^{N} C_m \right] \right],$$

$$= E_n\left[\min_{\substack{\{\delta H_m\}\\ m=n+1,\ldots,N}} E_m\left[\sum_{m=n+1}^{N} C_m\right]\right]$$
$$= E_n[J_{n+1}]. \tag{7.11}$$

Combining (7.11) and (7.10) gives

$$J_n = \min_{\delta H_n}\Big\{\delta H_n S_n e^{\lambda\delta H_n - (n-1)r\delta t} + E_n[J_{n+1}]\Big\}, \qquad n = 1,\ldots,N, \tag{7.12}$$

which is the Bellman equation for the cost-minimization problem. It should now be clear why we sought to re-express the original problem, (7.6), in terms of its equivalent Bellman equation. Instead of an optimization problem explicitly dependent on all N control variables, we now have a series of N optimization problems, each of which depends only on quantities known during the respective period thus eliminating the strongly coupled nature of the problem. How the optimal strategy is derived from the Bellman equation (7.12) will be dealt with in depth in the following section.

For future reference, we note that the existence of a Bellman equation for the cost-minimization problem depended on two aspects of the problem. First, as we have already shown the applicability of the principle of optimality was crucial, but just as important was the *additive separability* of the problem's objective function. In our context, additive separability refers to the ability to separate the objective function (in this case C_{total}) into a sum of N functions that all have a similar functional form and only depend on quantities during their respective time period. For our example this characteristic holds as, by definition, $C_{\text{total}} = C_1 + C_2 + \ldots + C_N$.

As was stated above, we will seek the solution to the problem of minimizing the transaction costs specifically for a liquidation order. Let the initial number of assets in the portfolio be H_1; for a liquidation order we also have, by definition, the final condition $H_{N+1} = 0$. The final relation needed to complete the minimization problem is the *continuity condition*

$$H_{n+1} = H_n + \delta H_n, \tag{7.13}$$

which simply relates the number of assets remaining in the portfolio between two consecutive trading periods. The complete cost-minimization problem for the liquidation transaction is

therefore

$$\begin{aligned} J_n &= \min_{\delta H_n}\Big\{\delta H_n S_n e^{\lambda\delta H_n-(n-1)r\delta t} + E_n[J_{n+1}]\Big\}, \\ H_{n+1} &= H_n + \delta H_n, \qquad\qquad H_{N+1} = 0, \\ S_{n+1} &= \begin{cases} u\cdot S_n e^{\lambda(1-\alpha)\delta H_n} & \text{probability } p, \\ d\cdot S_n e^{\lambda(1-\alpha)\delta H_n} & \text{probability } 1-p, \end{cases} \end{aligned} \tag{7.14}$$

which must hold for all $n = 1, \ldots, N$, and where H_1 is specified.

7.6 Solving for the Optimal Trading Strategy

The system (7.14) is relatively straightforward to solve. We begin at the final trading period when $n = N$; the Bellman equation for this period is

$$J_N = \min_{\delta H_N}\Big\{\delta H_N S_N e^{\lambda\delta H_N-(N-1)r\delta t} + E_N[J_{N+1}]\Big\}. \tag{7.15}$$

The final boundary condition requires $H_{N+1} = 0$ and this immediately implies that the optimal traded quantity during the final period, δH_N^*, is

$$\delta H_n^* = -H_N, \tag{7.16}$$

where H_N, as usual, is the number of assets remaining in the portfolio at the beginning of the N^{th} period. As the portfolio will be liquidated at the end of the N^{th} period there will be no assets remaining in the portfolio after this time and we must therefore have $J_{N+1} = 0$; equation (7.15) therefore reduces to

$$\begin{aligned} J_N &= \min_{\delta H_N}\Big\{\delta H_N S_N e^{\lambda\delta H_N-(N-1)r\delta t}\Big\}, \\ &= \delta H_N^* S_N e^{\lambda\delta H_N^*-(N-1)r\delta t}, \\ &= -H_N S_N e^{-\lambda H_N-(N-1)r\delta t}. \end{aligned} \tag{7.17}$$

We now move to the previous trading period. When $n = N-1$ equation (7.12) becomes

$$\begin{aligned} J_{N-1} &= \min_{\delta H_{N-1}} \Big\{ \delta H_{N-1} S_{N-1} e^{\lambda \delta H_{N-1} - (N-2) r \delta t} + E_{N-1}[J_N] \Big\}, \\ &= \min_{\delta H_{N-1}} \Big\{ \delta H_{N-1} S_{N-1} e^{\lambda \delta H_{N-1} - (N-2) r \delta t} - H_N E_{N-1}[S_N] e^{-\lambda H_N - (N-1) r \delta t} \Big\}, \end{aligned} \tag{7.18}$$

as H_N is predictable with the information known at the beginning of period $N-1$. We now evaluate the expectation and implement the continuity condition, $H_N = H_{N-1} + \delta H_{N-1}$, to give

$$\begin{aligned} J_{N-1} = \min_{\delta H_{N-1}} \Big\{ & \delta H_{N-1} S_{N-1} e^{\lambda \delta H_{N-1} - (N-2) r \delta t} \\ & - (H_{N-1} + \delta H_{N-1}) S_{N-1} e^{-\lambda H_{N-1} - \alpha \lambda \delta H_{N-1} + (\mu - (N-1) r) \delta t} \Big\}. \end{aligned} \tag{7.19}$$

Equation (7.19) is a minimization in one variable; the minimum is therefore simply the solution of the equation

$$\left. \frac{\partial J_{N-1}}{\partial (\delta H_{N-1})} \right|_{\delta H^*_{N-1}} = 0, \tag{7.20}$$

where δH^*_{N-1} is the optimal value of the traded quantity during the $N-1^{\text{th}}$ trading period. The minimization condition for δH_{N-1} is therefore

$$(1+\lambda \delta H^*_{N-1}) e^{\lambda \delta H^*_{N-1} - (N-2) r \delta t} - (1 - \alpha \lambda (H_{N-1} + \delta H^*_{N-1})) e^{-\alpha \lambda \delta H^*_{N-1} - \lambda H_{N-1} + (\mu - (N-1) r) \delta t} = 0. \tag{7.21}$$

To derive an explicit, analytical expression for δH^*_{N-1} we must use the fact that $r\delta t$, $\mu \delta t$, and λH_1 are all small so that we can expand the exponential terms in (7.21). More specifically, though, we will make the assumption that $r\delta t,\ \mu \delta t \ \approx\ \lambda H_1$.[1] Since we are interested in the leading-order form of the optimal strategy and it is clear from (7.21) that the $O(1)$ terms in that expression will cancel, the above assumption dictates that we retain up to the linear terms in both the δt and λ expansions. Under this assumption the leading-order minimization condition is

$$2(1+\alpha) \lambda \delta H^*_{N-1} + (1+\alpha) \lambda H_{N-1} + (\mu - r) \delta t = 0. \tag{7.22}$$

[1] This assumption is based on a daily trading frequency, $\delta t \approx 0.004$, a portfolio size, $H_1 = 10^4 - 10^6$, and typical values of λ of $10^{-7} - 10^{-9}$.

Finally, solving (7.22) for δH^*_{N-1} gives

$$\delta H^*_{N-1} = -\frac{1}{2}H_{N-1} + \frac{(\mu - r)\delta t}{2(1+\alpha)\lambda}, \tag{7.23}$$

which is the leading-order component of the optimal traded quantity during the $N-1^{\text{th}}$ trading period. Note that because we have assumed that $r\delta t$ and $\mu\delta t$ are approximately the same size as λH_1 both terms in (7.23) are $O(1)$.

Substituting (7.23) back into (7.19), expanding all exponential terms (except $e^{-r(T-2\delta t)}$ as this is the discount factor), and retaining the $O(1)$ and linear terms of the expansion gives the optimal value function for period $N-1$ as

$$J_{N-1} = -H_{N-1}S_{N-1}e^{-r(T-2\delta t)}\left(1 - \frac{1}{4}(3-\alpha)\lambda H_{N-1} + \frac{1}{2}(\mu - r)\delta t\right). \tag{7.24}$$

Repeating this same procedure for $n = N-2$ and then $n = N-3$ gives

$$\delta H^*_{N-2} = -\frac{1}{3}H_{N-2} + \frac{(\mu - r)\delta t}{(1+\alpha)\lambda}, \tag{7.25}$$

$$J_{N-2} = -H_{N-2}S_{N-2}e^{-r(T-3\delta t)}\left(1 - \frac{1}{3}(2-\alpha)\lambda H_{N-2} + (\mu - r)\delta t\right), \tag{7.26}$$

$$\delta H^*_{N-3} = -\frac{1}{4}H_{N-3} + \frac{3(\mu - r)\delta t}{2(1+\alpha)\lambda}, \tag{7.27}$$

$$J_{N-3} = -H_{N-3}S_{N-3}e^{-r(T-4\delta t)}\left(1 - \frac{1}{8}(5-3\alpha)\lambda H_{N-3} + \frac{3}{2}(\mu - r)\delta t\right). \tag{7.28}$$

Extrapolating for general $m = 1, \ldots, N-1$ it is clear that

$$\delta H^*_{N-m} = -\frac{1}{m+1}H_{N-m} + \frac{m(\mu - r)\delta t}{2(1+\alpha)\lambda}, \tag{7.29}$$

and

$$J_{N-m} = -H_{N-m}S_{N-m}e^{-r(T-(m+1)\delta t)}\Bigg(1 + \frac{m}{2}(\mu - r)\delta t - \frac{1}{2(m+1)}((m+2) - m\alpha)\lambda H_{N-m}\Bigg). \tag{7.30}$$

In the first trading period ($m = N - 1$) the optimal value function, J_1, is

$$J_1 = -H_1 S_1 \left(1 - \frac{1}{2N}(N + 1 - (N-1)\alpha)\lambda H_1 + \frac{1}{2}(N-1)(\mu - r)\delta t\right). \tag{7.31}$$

The above expression tells us that the expected revenue generated from the liquidation transaction is, to leading order, $H_1 S_1$, which is simply the book value of the portfolio. The first-order correction to the book value arises from two sources. First, because of our trading activity we will continually depress the permanent value of the asset price and receive less per asset than if there was an infinite market depth; as a result this activity contributes $-\frac{1}{2N}(N + 1 - (N-1)\alpha)\lambda H_1$ which is always negative as $0 \leq \alpha \leq 1$. Second, the longer an asset remains in our portfolio before being sold, the more we expect it to earn compared to if it were invested at the risk-free rate; this expected rate of return above the risk-free rate will contribute positively to our expected revenue in the amount $\frac{1}{2}(N-1)(\mu - r)\delta t$.

In general we expect the liquidity term in (7.31) to dominate the expected return term so that the expected revenue generated by the portfolio will be less than its book value. But if $T = O(1)$ and N is very large then the situation where the growth of the asset over the lifetime of the transaction dominates the trading effect is possible and we then expect to receive a revenue from the sale of the portfolio greater than its book value.

We can put the optimal trading strategy, (7.29) in a more useful form. Beginning at $m = N - 1$ (7.29) becomes

$$\delta H_1^* = -\frac{1}{N} H_1 + (N-1)\frac{(\mu - r)\delta t}{2(1+\alpha)\lambda}.$$

Using the continuity relation, $H_2 = H_1 + \delta H_1$, gives

$$H_2 = \frac{N-1}{N} H_1 + (N-1)\frac{(N-1)(\mu - r)\delta t}{2(1+\alpha)\lambda}.$$

Once again using equation (7.29), but now with $m = N - 2$ gives

$$\begin{aligned}\delta H_2^* &= -\frac{1}{N-1} H_2 + (N-2)\frac{(N-2)(\mu - r)\delta t}{2(1+\alpha)\lambda}, \\ &= -\frac{1}{N} H_1 + (N-3)\frac{(\mu - r)\delta t}{2(1+\alpha)\lambda}.\end{aligned}$$

Repeating this method reveals the final expression for the optimal trading strategy when minimizing costs only; this is

$$\delta H_n^* = -\frac{1}{N}H_1 + (N-(2n-1))\,\frac{(\mu-r)\delta t}{2(1+\alpha)\lambda} + O(\{r,\mu\}\delta t) + O(\lambda H_1), \qquad n = 1,\ldots,N, \tag{7.32}$$

where the $O(\delta t)$ and $O(\lambda H_1)$ terms have been explicitly included to remind us that this is strictly the leading-order component of the solution.

7.7 Discussion

The first noticeable feature of the optimal trading strategy, (7.32), is that it is static; all trades are completely characterizable in terms of quantities known before the trade begins. We will find that this property is not preserved when we attempt to solve the mean-variance optimization problem.

There are two components to the cost-minimization optimal trading strategy. First, the term $-H_1/N$ is the same for all trading periods; this is exactly the naive strategy found by Bertsimas and Lo [6] and Almgren and Chriss [2] when the trader's risk-aversion is identically zero. This component simply states that we should sell an equal number of assets each period.

Second, in addition to the $-H_1/N$ assets of the naive strategy, we should trade a further $(N-(2n-1))\frac{(\mu-r)\delta t}{2(1+\alpha)\lambda}$ assets to maximize the transaction's revenue. This liquidity-modified *cost-of-carry* term is present due to our expectations that an amount of cash invested in the asset will grow in value at a faster rate than if invested in a risk-free asset and was not present in either [6] or [2] as interest rates were neglected in those papers.

The magnitude of this second term depends on the relative magnitude of the two effects discussed towards the end of the previous section. For a relatively illiquid market with a small market depth (large λ) and δt very small, this term will be small and the naive strategy will be approximately correct; this agrees with our intuition as a large trade will impact much more in an illiquid market. When the market is very liquid, on the other hand, λ will be very small and this term can become very significant.

The sign of the second term is positive for the first half of the transaction when $n < N/2$ and negative for the second half of the transaction. The effect is also largest at the endpoints

of the transaction which means that we should sell the fewest assets during the first trading interval and the most during the last.

Interestingly, for the first trade we have

$$\delta H_1^* = -\frac{1}{N}H_1 + (N-1)\frac{(\mu - r)\delta t}{2(1+\alpha)\lambda}; \tag{7.33}$$

if the condition

$$\lambda < \frac{N(\mu - r)(T - \delta t)}{2(1+\alpha)} \tag{7.34}$$

is ever satisfied (which is very likely for reasonable values of the parameters) then we find that this *unconstrained* optimization states that $\delta H_1^* > 0$; the revenue will be maximized for the liquidation order if we *buy* assets at the beginning of the transaction.

Finally, one further interesting aspect of our result is that if we take the continuous-time limit of the optimal strategy where $\delta t \to 0$, $N \to \infty$, such that $N\delta t = T$ (and $n\delta t = t$) we have

$$dH_t^* = (T - 2t)\frac{\mu - r}{(1+\alpha)\lambda}.$$

This interesting result tells us that if we intend to optimize the liquidation process in continuous time we should buy an infinite amount of the asset at the beginning of the interval and then sell it all off when $t = T/2$.

Chapter 8

Optimal Liquidation with Small Risk Aversion

In this chapter we extend the analysis developed in the previous chapter to the full mean-variance formulation of the optimal portfolio transaction problem. That is, we will attempt to find the trading strategy that simultaneously minimizes both the expectation and the variance of the cost of the transaction. Toward this goal we will find there are several characteristics of the mean-variance problem that make it very difficult to solve, but that approximate solutions *can* be found and that these reveal some interesting qualitative features of the system.

8.1 Outline of Chapter

We will begin in Section 8.2 by formulating the exact mean-variance stochastic optimization problem. In Section 8.3 we then demonstrate how the inclusion of a variance component to the optimization problem drastically increases its difficulty and suggest a modified version of the problem that partially overcomes these problems. In Section 8.4 we present a novel method of solving the modified problem and prove two theorems to show that the solution is indeed the one we seek. In Section 8.5 we derive the equivalent Bellman formulation of the modified problem and then solve it in Section 8.6, but find that there are many regions of the solution space across which the solution varies greatly in form. The rest of the chapter

will focus on the optimal strategy in the limit of a very small level of risk-aversion by the trader. We finally conclude the chapter in Section 8.7 with a discussion of the results.

8.2 Formulation of the Mean-Variance Problem

In this chapter we will again focus on finding the optimal trading strategy for a liquidation transaction over a finite horizon, T. Once again we divide the interval into N trading periods of equal length, δt, such that $N\delta t = T$. Let δH_n be the number of assets traded during the n^{th} period; a trading strategy is then defined to be $\{\delta H_n\}$ where $n = 1, \ldots, N$.

The mean-variance problem consists of finding a strategy that minimizes the expected total cost of the trasaction while simultaneously minimizing the variance of this cost. As in the previous chapter, let C_n be the cost of the trade during the n^{th} trading period discounted to the beginning of the first trading period; that is

$$C_n = \delta H_n S_n e^{\lambda \delta H_n - (n-1) r \delta t}. \tag{8.1}$$

The total cost of the strategy associated with the sequence of these N trades is defined as C_{total} and is then simply

$$C_{\text{total}} = \sum_{n=1}^{N} C_n = \sum_{n=1}^{N} \delta H_n S_n e^{\lambda \delta H_n - (n-1) r \delta t}. \tag{8.2}$$

The problem we will aim to solve can now be stated mathematically as

$$\begin{aligned} &\min_{\substack{\{\delta H_n\} \\ n=1,\ldots,N}} E_1[C_{\text{total}}], \\ s.t. \qquad &\text{Var}_1[C_{\text{total}}] \quad \text{is a minimum,} \end{aligned} \tag{8.3}$$

where $\text{Var}_1[\cdot]$ is the variance of the argument taken *w.r.t.* the information available at the beginning of the first trading period. We now employ the standard technique when optimizing a functional over two, simultaneous constraints; we introduce the parameter

$\Lambda \geq 0$ and rewrite (8.3) as the equivalent problem

$$\min_{\substack{\{\delta H_n\} \\ n=1,\ldots,N}} \left\{ E_1[C_{\text{total}}] + \Lambda \text{Var}_1[C_{\text{total}}] \right\}, \tag{8.4}$$

where Λ is interpreted as the level of the trader's risk-aversion. The variance contribution is positive in (8.4) as we are focusing on liquidation transactions. As $\sum_n \delta H_n < 0$ and since the variance 'risk' effect acts against the expectation effect the overall signs of the two terms must be opposite. If we were calculating the optimal trading strategy for a buy order this variance contribution would instead be negative. Equation (8.4) is the exact form of the problem we will attempt to solve.

8.3 Difficulties with Variance Minimization

Now possessing the exact optimization mean-variance problem, we proceed and attempt to solve (8.4) using a similar method as was used in Chapter 7; first generate the equivalent Bellman equation formulation, solve this equation iteratively backward to find the general optimal strategy, and then re-express this general solution in terms of initial quantities by re-solving it forward.

There is a problem with equation (8.4) that prevents us from carrying-out this procedure. As was mentioned in Section 7.5, the two characteristics necessary for a dynamic optimization problem to have an equivalent Bellman equation are: 1. the principle of optimality must be applicable; and 2. the objective function must be additively separable. While the principle of optimality still applies to the mean-variance problem, the additive separability of the cost-minimizing objective function is now lost due to the presence of the variance term. Variance minimization in stochastic optimization is a notoriously difficult problem and this last statement demonstrates why this is so.

There are actually two separate difficulties with the variance component of our objective function. First, general to variance minimization problems is the fact that the variance operator does not satisfy the smoothing property. Effectively the crucial step in deriving the Bellman equation in the previous chapter was the expansion of $E_n[\cdot]$ into $E_n[E_m[\cdot]]$ between equations (7.10) and (7.11) which was possible only as a result of the expectation operation satisfying the smoothing property. Since $\text{Var}_i[\text{Var}_j[y_k]] \neq \text{Var}_i[y_k]$ for some random variable

y generated at time $k > j$ and for any $j > i$, we will not be able to express our objective function in an additively separable form and therefore not be able to derive a Bellman equation for the problem.

Regardless of the forms of the underlying stochastic process and the specific functional form of the objective function, the difficulty described above will arise whenever a variance term is present within the objective function. There is a second difficulty, however, that is specific to our problem. Expanding the variance term in (8.4) gives

$$\begin{aligned} \mathrm{Var}_1[C_{\text{total}}] &= \mathrm{Var}_1\left[\sum_{n=1}^{N} C_n\right], \\ &= \mathrm{Var}_1\left[\sum_{n=1}^{N} \delta H_n S_n e^{\lambda \delta H_n - (n-1) r \delta t}\right]. \end{aligned} \tag{8.5}$$

If each of the $C_n^{\text{'s}}$ were uncorrelated then we could write

$$\mathrm{Var}_1\left[\sum_{n=1}^{N} \delta H_n S_n e^{\lambda \delta H_n - (n-1) r \delta t}\right] = \sum_{n=1}^{N} \mathrm{Var}_1\left[\delta H_n S_n e^{\lambda H_n - (n-1) r \delta t}\right], \tag{8.6}$$

and the only difficulty would be the non-smoothing issue discussed above. But it is clear that $E_1[S_i S_j] \neq E_1[S_i] E_1[S_j]$ for any (i, j) (except $i = j = 1$) as $E_1[S_i]$ is dependent on all δH_k where $k = 1, \ldots, i-1$.

In the next section we will present a method to overcome the fact that the variance operator does not satisfy the smoothing property, but unfortunately it will not work if covariance terms are present in the objective function. To overcome this obstacle we simply propose to neglect the covariance terms in the objective function. While this may seem ridiculous given that there are $N(N-1)$ of these terms and that their effect is definitely not negligible compared to the variance terms, it *is* reasonable in the following sense. Our original objective was to find the strategy that simultaneously minimizes both the expectation and variance of the cost of the liquidation transaction where the purpose of the variance contribution was to incorporate an effect that penalizes for delaying the sale of an asset due to the increased risk that that delay imparts into the total cost. By switching this time-risk penalty term from $\mathrm{Var}_1\left[\sum_n C_n\right]$ to $\sum_n \mathrm{Var}_1[C_n]$ we are ignoring the covariance contributions to the risk term, but the term itself will still be a measure of the strategies'

risk, just not precisely its variance.

One significant consequence of the above simplification is that, by neglecting the objective function's covariance terms, we are effectively neglecting any serial correlations that exist in the asset price process. As one would expect, there will be consequences for the optimal strategy due to this assumption and these will be discussed in section 8.7 after we derive our main results for the chapter.

As a result of the above discussion we now reformulate the problem we wish to solve; it is now

$$\begin{aligned} &\min_{\substack{\{\delta H_n\} \\ n=1,\ldots,N}} \left\{ \sum_{n=1}^{N} E_1[C_{\text{total}}] + \Lambda \sum_{n=1}^{N} \text{Var}_1[C_n] \right\}, \\ &= \min_{\substack{\{\delta H_n\} \\ n=1,\ldots,N}} \left\{ \sum_{n=1}^{N} \left(E_1[C_n] + \Lambda \text{Var}_1[C_n] \right) \right\}, \end{aligned} \tag{8.7}$$

which from now on we refer to as the Risk-Adjusted optimization problem.

8.4 The Auxiliary Formulation for the Risk-Adjusted (RA) Problem

Due to the presence of the variance term in its objective function, the RA problem is still not solvable by dynamic programming. We will overcome this obstacle by extending a technique developed by Li and Ng [34] to solve the multiperiod mean-variance portfolio selection problem. The basic idea behind the method is to *embed* the RA problem within a closely related, yet more general, stochastic optimization problem that *is* solvable using dynamic programming. With a careful choice of one of the parameters of the more general problem we can then show that its solution is exactly the solution of the RA problem. The remainder of this section will focus on developing this technique and proving the necessary theorems to demonstrate the equivalence of the two problems.

To begin, we define the *auxiliary* problem to the RA problem as

$$\min_{\substack{\{\delta H_n\} \\ n=1,\ldots,N}} \left\{ \sum_{n=1}^{N} \left(\beta_n E_1[C_n] + \Lambda E_1[C_n^2] \right) \right\} \tag{8.8}$$

where β_n is a parameter that is only dependent on n. As the objective function of the auxiliary problem is linear in the expectation operator it will satisfy the smoothing property and can therefore be put into an additively separable form making it solvable using dynamic programming.

We must now show that the solution of the auxiliary problem (8.8) is exactly the solution of the RA problem (8.7) for the appropriate choice of the parameter β_n. To this end we define the set of all trading strategies that solve the RA problem as $\Pi_{RA}(\Lambda)$ and the set of all solutions of the auxiliary problem as $\Pi_A(\beta_n, \Lambda)$. The first step towards this goal is to show that $\Pi_{RA}(\Lambda)$ is a subset of $\Pi_A(\beta_n, \Lambda)$ for the correct choice of β_n; this result is captured in the following theorem which generalises the result in [34] to problems in which the objective function explicitly depends on all n instead of just N.

Theorem 8.4.1 *If* $\pi^* \in \Pi_{RA}(\Lambda)$*, then* $\pi^* \in \Pi_A(1 - 2\Lambda E_1[C_n], \Lambda)$*.*

Proof We proceed by contradiction. Let π^* be a solution of the RA problem, but not of the auxiliary problem with $\beta_n = 1 - 2\Lambda E_1[C_n]$. In this case there must then exist some solution, π, that better minimizes the auxiliary objective function; that is

$$\sum_{n=1}^{N} \left(\beta_n E_1[C_n] + \Lambda E_1[C_n^2]\right)\Big|_{\pi} < \sum_{n=1}^{N} \left(\beta_n E_1[C_n] + \Lambda E_1[C_n^2]\right)\Big|_{\pi^*}. \tag{8.9}$$

Now let

$$\begin{aligned} U_n &= E_1[C_n] + \Lambda \mathrm{Var}_1[C_n], \\ &= E_1[C_n] + \Lambda E_1[C_n^2] - \Lambda E_1[C_n]^2, \end{aligned} \tag{8.10}$$

and Taylor expand $U_n(\pi)$ around the point π^* to get

$$\begin{aligned}
U_n(\pi) &= U_n(\pi^*) \\
&\quad + \left(E_1[C_n(\pi)] - E_1[C_n(\pi^*)]\right) \left.\frac{\partial U_n}{\partial(E_1[C_n])}\right|_{\pi^*} + \left(E_1[C_n^2(\pi)] - E_1[C_n^2(\pi^*)]\right) \left.\frac{\partial U_n}{\partial(E_1[C_n^2])}\right|_{\pi^*} \\
&+ \frac{1}{2}\left(E_1[C_n(\pi)] - E_1[C_n(\pi^*)]\right)^2 \left.\frac{\partial^2 U_n}{\partial(E_1[C_n])^2}\right|_{\pi^*} + \left(E_1[C_n^2(\pi)] - E_1[C_n^2(\pi^*)]\right)^2 \left.\frac{\partial^2 U_n}{\partial(E_1[C_n^2])^2}\right|_{\pi^*} \\
&\quad + \left(E_1[C_n(\pi)] - E_1[C_n(\pi^*)]\right)\left(E_1[C_n^2(\pi)] - E_1[C_n^2(\pi^*)]\right) \left.\frac{\partial^2 U_n}{\partial(E_1[C_n])\partial(E_1[C_n^2])}\right|_{\pi^*} + \cdots .
\end{aligned} \tag{8.11}$$

Differentiating U_n *w.r.t.* the arguments $E_1[C_n]$ and $E_1[C_n^2]$ gives

$$\begin{gathered}
\frac{\partial U_n}{\partial(E_1[C_n])} = 1 - 2\Lambda E_1[C_n], \qquad \frac{\partial U_n}{\partial(E_1[C_n^2])} = \Lambda, \\
\frac{\partial U_n}{\partial(E_1[C_n])^2} = -2\Lambda, \qquad \frac{\partial U_n}{\partial(E_1[C_n^2])^2} = 0, \\
\frac{\partial^2 U_n}{\partial(E_1[C_n])\partial(E_1[C_n^2])} = 0.
\end{gathered}$$

Finally, substituting these partial derivatives back into (8.11) and then summing over all n gives an expression for the total objective function; this is

$$\begin{aligned}
\sum_{n=1}^{N} U_n(\pi) &= \sum_{n=1}^{N} U_n(\pi^*) + \sum_{n=1}^{N} \left(\left(1 - 2\Lambda E_1[C_n(\pi)]\right) E_1[C_n(\pi)] + \Lambda E_1[C_n^2(\pi)]\right) \\
&- \sum_{n=1}^{N} \left(\left(1 - 2\Lambda E_1[C_n(\pi^*)]\right) E_1[C_n(\pi^*)] + \Lambda E_1[C_n^2(\pi^*)]\right) - \Lambda \sum_{n=1}^{N} \left(E_1[C_n(\pi)] - E_1[C_n(\pi^*)]\right)^2 .
\end{aligned} \tag{8.12}$$

From (8.9) the 2^{nd} term on the *r.h.s.* of (8.12) must be less than the 3^{rd} term; as the 4^{th} term is strictly negative this implies $\sum_{n=1}^{N} U_n(\pi) < \sum_{n=1}^{N} U_n(\pi^*)$ which contradicts the original assumption and therefore concludes the proof. ∎

With the result $\pi^* \in \Pi_{RA}(\Lambda) \implies \pi^* \in \Pi_A(1 - 2\Lambda E_1[C_n], \Lambda)$ we now proceed to Theorem (8.4.2) to derive our most important result, the necessary and sufficient condition for a solution of the auxiliary problem to also be a solution of the RA problem. Once again

this is a generalisation of the work in [34] to objective functions that depend on all n rather than N alone.

Theorem 8.4.2 *For a solution, π^*, of the auxiliary problem, a necessary and sufficient condition for π^* to be a solution of the RA problem is $\beta_n(\pi^*) = 1 - 2\Lambda E_1[C_n(\pi^*)]$.*

Proof Any solution, π^*, of the auxiliary problem can be represented as existing in a $N+1$-dimensional vector space parameterized by the values $(\{\beta_n(\pi^*)\}, \Lambda)$. We have already shown in Theorem 8.4.1 that $\Pi_{RA}(\pi^*) \subseteq \Pi_A(\pi^*)$ and so we can generalize the set of solutions for the RA problem by also (somewhat artificially) parameterizing them by $(\{\beta_n(\pi)\}, \Lambda)$. For an arbitrary (but fixed) value of Λ the exact problem can therefore be written

$$\min_{\substack{\{\beta_n\} \\ n=1,\ldots,N}} \left\{ \sum_{n=1}^{N} \left(E_1[C_n(\beta_n(\pi))] - \Lambda E_1[C_n(\beta_n(\pi))^2] + \Lambda E_1[C_n(\beta_n(\pi))]^2 \right) \right\}. \tag{8.13}$$

Equation (8.13) is simply a N-dimensional minimization problem which therefore has the first-order minimization condition

$$\begin{aligned} &\sum_{n=1}^{N} \frac{\partial}{\partial \beta_n(\pi)} \left(E_1[C_n(\beta_n(\pi))] - \Lambda E_1[C_n(\beta_n(\pi))^2] + \Lambda E_1[C_n(\beta_n(\pi))]^2 \right)\Big|_{\beta_n(\pi^*)} = 0, \\ \rightarrow &\sum_{n=1}^{N} \left((1 - 2\Lambda E_1[C_n(\beta_n(\pi^*))]) \frac{\partial E_1[C_n(\beta_n(\pi^*))]}{\partial \beta_n} + \Lambda \frac{\partial E_1[C_n(\beta_n(\pi^*))^2]}{\partial \beta_n} \right) = 0. \end{aligned} \tag{8.14}$$

Now, since $\pi^* \in \Pi_A(\beta_n(\pi^*), \Lambda)$, by definition it solves the auxiliary problem, (8.8). As before, the space $\Pi_A(\{\beta_n(\pi)\}, \Lambda)$ can be completely characterized by the vector $(\{\beta_n(\pi)\}, \Lambda)$; fixing Λ and minimizing the auxiliary objective function *w.r.t.* $\{\beta_n\}$ gives its first-order minimization condition

$$\begin{aligned} &\sum_{n=1}^{N} \frac{\partial}{\partial \beta_n(\pi)} \left(\beta_n(\pi) E_1[C_n(\beta_n(\pi))] + \Lambda E_1[C_n(\beta_n(\pi))^2] \right)\Big|_{\beta_n(\pi^*)}, \\ \rightarrow &\sum_{n=1}^{N} \left(\beta_n(\pi^*) \frac{E_1[C_n(\beta_n(\pi^*))]}{\partial \beta_n(\pi)} + \Lambda \frac{E_1[C_n(\beta_n(\pi^*))^2]}{\partial \beta_n(\pi)} \right) = 0. \end{aligned} \tag{8.15}$$

For the solution, π^*, of the auxiliary problem to be a solution of the RA problem equation

(8.14) must equal equation (8.15) which therefore implies

$$\beta_n(\pi^*) = 1 - 2\Lambda E_1[C_n(\pi^*)], \qquad \forall\ n, \tag{8.16}$$

where the artificial dependence of C_n on β_n has been dropped. ∎

With Theorem 8.4.2 we now have our desired result. To solve the RA problem, (8.7), we only need to solve the much easier auxiliary problem, (8.8). The solution, π^*, of the auxiliary problem will consist of a sequence of optimal trading quantities, $\{\delta H_n^*\}$. With these values we compute each value of $\beta_n(\pi^*)$ and after substituting these back into the auxiliary solution we are left with exactly the solution of the RA problem.

8.5 The Bellman Equation for the Auxiliary Problem

Before solving the auxiliary problem we must derive its equivalent Bellman equation. Toward this goal we follow the same procedure as in the previous chapter. Let J_n be the optimal value function for the auxiliary problem when beginning at the n^{th} trading period; this is written

$$\begin{aligned} J_n &= \min_{\substack{\{\delta H_m\} \\ m=n,\ldots,N}} \left\{ \sum_{m=n}^{N} \left(\beta_m E_n[C_m] + \Lambda E_n[C_m^2] \right) \right\}, \\ &= \min_{\substack{\{\delta H_m\} \\ m=n,\ldots,N}} \left\{ \sum_{m=n}^{N} E_n\left[\beta_m C_m + \Lambda C_m^2 \right] \right\}. \end{aligned} \tag{8.17}$$

Partially expanding the sum gives

$$\begin{aligned} J_n &= \min_{\substack{\{\delta H_m\} \\ m=n,\ldots,N}} \left\{ E_n[\beta_n C_n + \Lambda C_n^2] + \sum_{m=n+1}^{N} E_n\left[\beta_m C_m + \Lambda C_m^2 \right] \right\}, \\ &= \min_{\substack{\{\delta H_m\} \\ m=n,\ldots,N}} \left\{ \beta_n C_n + \Lambda C_n^2 + \sum_{m=n+1}^{N} E_n\left[\beta_m C_m + \Lambda C_m^2 \right] \right\}, \\ &= \min_{\delta H_n} \left\{ \beta_n C_n + \Lambda C_n^2 + \min_{\substack{\{\delta H_m\} \\ m=n+1,\ldots,N}} \sum_{m=n+1}^{N} E_n\left[\beta_m C_m + \Lambda C_m^2 \right] \right\}, \end{aligned}$$

$$= \min_{\delta H_n} \left\{ \beta_n C_n + \Lambda C_n^2 + E_n \left\{ \min_{\substack{\{\delta H_m\} \\ m=n+1,\ldots,N}} \sum_{m=n+1}^{N} E_{n+1} \left[\beta_m C_m + \Lambda C_m^2 \right] \right\} \right\}.$$

For the liquidation transaction the auxiliary problem's Bellman equation formulation is then

$$J_n = \min_{\delta H_n} \left\{ \beta_n C_n + \Lambda C_n^2 + E_n[J_{n+1}] \right\}, \qquad n = 1, \ldots, N, \tag{8.18}$$

$$\text{where} \quad C_n = \delta H_n S_n e^{\lambda \delta H_n - (n-1) r \delta t},$$

$$H_{n+1} = H_n + \delta H_n, \quad H_{N+1} = 0,$$

and H_1 is specified.

8.6 Solving for the Risk-Adjusted Optimal Trading Strategy

8.6.1 Solving the Auxiliary Equation

As we found in Section 8.4, to solve the RA problem we must solve the auxiliary problem and substitute-in the appropriate $\{\beta_n\}$ calculated from this optimal solution. Proceeding as we did with the cost-minimization problem of Chapter 7, we begin at the final trading period and iteratively solve for the optimal trading strategy backward in time.

During the final trading period the final condition, $H_{N+1} = 0$, dictates that the optimal number of assets traded, δH_N^*, in that period is simply

$$\delta H_N^* = -H_N. \tag{8.19}$$

As there will be no assets remaining at the end of the N^{th} trading period there can therefore be none sold during the $N+1^{\text{th}}$ interval giving us the result

$$J_{N+1} = 0. \tag{8.20}$$

For $n = N$ the auxiliary Bellman equation, (8.8) is

$$J_N = \min_{\delta H_N} \left\{ \beta_N C_N + \Lambda C_n^2 + E_N[J_{N+1}] \right\}, \tag{8.21}$$

but after substituting expressions (8.19) and (8.20) into (8.21) it reduces to

$$J_N = -\beta_N H_N S_N e^{-\lambda H_N - (N-1)r\delta t} + \Lambda H_N^2 S_N^2 e^{-2\lambda H_N - 2(N-1)r\delta t}. \tag{8.22}$$

The problem to be solved in the $N-1^{\text{th}}$ trading period is

$$J_{N-1} = \min_{\delta H_{N-1}} \Big\{ \beta_{N-1}\delta H_{N-1} S_{N-1} e^{\lambda \delta H_{N-1} - (N-2)r\delta t} + \Lambda \delta H_{N-1}^2 S_{N-1}^2 e^{2\lambda \delta H_{N-1} - 2(N-2)r\delta t} + E_{N-1}[J_N] \Big\}. \tag{8.23}$$

Substituting (8.22) into the final term of (8.23) gives

$$\begin{aligned} E_{N-1}[J_N] &= -E_{N-1}[\beta_N H_N S_N e^{-\lambda H_N - (N-1)r\delta t} - \Lambda H_N^2 S_N^2 e^{-2\lambda H_N - 2(N-1)r\delta t}], \\ &= -\beta_N H_N E_{N-1}[S_N] e^{-\lambda H_N - (N-1)r\delta t} + \Lambda H_N^2 E_{N-1}[S_N^2] e^{-2\lambda H_N - 2(N-1)r\delta t}, \end{aligned} \tag{8.24}$$

as H_N and $\beta_N = 1 - 2\Lambda E_1[C_N^*]$ are both known at the beginning of the $N-1^{\text{th}}$ period. We now employ the fact that the binomial component of the asset price process can be calibrated to a Geometric Brownian Motion with drift, μ, and volatility, σ, as was discussed in Chapter 1; in this case we have the result

$$E_n[S_{n+1}] = S_n e^{\lambda(1-\alpha)\delta H_n + \mu\delta t}, \tag{8.25}$$

$$E_n[S_{n+1}^2] = S_n^2 e^{2\lambda(1-\alpha)\delta H_n} \left(e^{\mu\delta t} \left(e^{\sigma\sqrt{\delta t}} + e^{-\sigma\sqrt{\delta t}} \right) - 1 \right). \tag{8.26}$$

Substituting (8.25), (8.26), and the continuity relation for $n = N-1$ into (8.24) gives

$$\begin{aligned} E_{N-1}[J_N] &= -\beta_N (H_{N-1} + \delta H_{N-1}) S_{N-1} e^{-\lambda H_{N-1} - \alpha\lambda\delta H_{N-1} + (\mu - (N-1)r)\delta t} \\ &+ \Lambda \left(H_{N-1} + \delta H_{N-1} \right)^2 S_{N-1}^2 \left(e^{\mu\delta t} \left(e^{\sigma\sqrt{\delta t}} + e^{-\sigma\sqrt{\delta t}} \right) - 1 \right) e^{-2\lambda H_{N-1} - 2\alpha\lambda\delta H_{N-1} - 2(N-1)r\delta t}. \end{aligned} \tag{8.27}$$

Combining equations (8.27) and (8.23) leaves us with

$$\begin{aligned} J_{N-1} = \min_{\delta H_{N-1}} \Big\{ & \beta_{N-1}\delta H_{N-1} S_{N-1} e^{\lambda\delta H_{N-1} - (N-2)r\delta t} + \Lambda \delta H_{N-1}^2 S_{N-1}^2 e^{2\lambda\delta H_{N-1} - 2(N-2)r\delta t} \\ & - \beta_N (H_{N-1} + \delta H_{N-1}) S_{N-1} e^{-\lambda H_{N-1} - \alpha\lambda\delta H_{N-1} + (\mu - (N-1)r)\delta t} \\ & + \Lambda \left(H_{N-1} + \delta H_{N-1}\right)^2 S_{N-1}^2 \left(e^{\mu\delta t}\left(e^{\sigma\sqrt{\delta t}} + e^{-\sigma\sqrt{\delta t}}\right) - 1\right) e^{-2\lambda H_{N-1} - 2\alpha\lambda\delta H_{N-1} - 2(N-1)r\delta t} \Big\}. \end{aligned} \tag{8.28}$$

The first-order minimization condition on J_{N-1} is simply

$$\left.\frac{\partial J_{N-1}}{\partial\left(\delta H_{N-1}\right)}\right|_{\delta H_{N-1}^*} = 0, \tag{8.29}$$

where δH_{N-1}^* is the optimal trade quantity for the $N-1^{\text{th}}$ period. After evaluating the derivative in (8.29) and dividing through by $S_{N-1}e^{-(N-2)r\delta t}$ this condition becomes

$$\begin{aligned} & \beta_{N-1}\left(1 + \lambda\delta H_{N-1}^*\right) e^{\lambda\delta H_{N-1}^*} - \beta_N \left(1 - \alpha\lambda\left(H_{N-1} + \delta H_{N-1}^*\right)\right) e^{-\alpha\lambda\delta H_{N-1}^* - \lambda H_{N-1} + (\mu - r)\delta t} \\ & \quad + 2\Lambda S_{N-1} e^{-2\lambda H_{N-1} - (N-2)r\delta t}\Big(\delta H_{N-1}^*\left(1 + \lambda\delta H_{N-1}^*\right) e^{2\lambda\delta H_{N-1}^*} \\ & \quad + c_1 e^{-2r\delta t}\left(H_{N-1} + \delta H_{N-1}^*\right)\left(1 - \alpha\lambda\left(H_{N-1} + \delta H_{N-1}\right)\right) e^{-2\alpha\lambda\delta H_{N-1}^*}\Big) = 0, \end{aligned} \tag{8.30}$$

where $c_1 = e^{\mu\delta t}\left(e^{\sigma\sqrt{\delta t}} + e^{-\sigma\sqrt{\delta t}}\right) - 1$. Because of the exponential terms in (8.30) a closed form of its solution is not possible. Instead we will proceed as in the previous chapter and linearize the exponential terms (except $e^{-(N-2)r\delta t}$ which acts as a discounting factor and will be kept), retain the $O(1)$, $O(\delta t)$, and $O(\lambda)$ terms, and seek an approximate solution to the equation. After this procedure equation (8.30) reduces to

$$\begin{aligned} & -6\Lambda\eta(1-\alpha)\lambda\delta H_{N-1}^{*2} + \Bigg(4\Lambda\eta\left(1 - (1+2\alpha)\lambda H_{N-1} + \left(\mu - r + \frac{1}{2}\sigma^2\right)\delta t\right) \\ & \quad + 2(1+\alpha)\lambda - 4\Lambda\lambda E_1[C_{N-1}^* + \alpha C_N^*]\Bigg)\delta H_{N-1}^* + (1+\alpha)\lambda H_{N-1} - (\mu - r)\delta t \\ & \quad + 2\Lambda\eta H_{N-1}\left(1 - (2+\alpha)\lambda H_{N-1} + 2\left(\mu - r + \frac{1}{2}\sigma^2\right)\delta t\right) \\ & \quad - 2\Lambda E_1[C_N^*]\left(1 - (1+\alpha)\lambda H_{N-1} + (\mu - r)\delta t\right) + 2\Lambda E_1[C_{N-1}^*] = 0, \end{aligned} \tag{8.31}$$

where $\eta = e^{-(N-2)r\delta t}S_{N-1}$, the β's have been written-out explicitly according to equation (8.16), and we have simplified the notation by writing $C_n(\pi^*) = C_n^*$ which will be used throughout the rest of our work.

While equation (8.31) is easily solvable for δH_{N-1}^*, the solution will be large and cumbersome and extracting its qualitative features will be difficult. For a practical application of the model, retaining the full form of (8.31) (as well as possibly some of the higher-order terms in the exponential expansions that have already been dropped) would be straightforward as the calculations would most likely be numerical in nature. But in this work we are primarily concerned with the qualitative aspects of the optimal trading strategy and so further simplification of (8.31) is needed.

As well as the interaction between the dynamic properties of the asset price and the effects that the trading have on this process which was seen in the cost-minimization model, there is now another effect which results from the risk-aversion of the trader. Effectively, the complexity that remains in expression (8.31) results from the fact that Λ is still completely general. To demonstrate this we first imagine the situation $\Lambda \gg 1$ and $\delta t \ll 1$, $\lambda H_1 \ll 1$. In this case the leading-order form of (8.31) is

$$4\Lambda\eta\delta H_{N-1}^* + 2\Lambda\eta H_{N-1} + 2\Lambda E_1[C_N^*] - 2\Lambda E_1[C_{N-1}^*] = 0.$$

Now imagine the opposite situation where again $\delta t \ll 1$ and $\lambda H_1 \ll 1$, but now where Λ is much smaller than either of these two terms. In this case the leading-order form of (8.31) is

$$2(1+\alpha)\lambda\delta H_{N-1}^* + (1+\alpha)\lambda H_{N-1} - (\mu - r)\delta t = 0.$$

In these two cases with very different levels of risk-aversion the leading-order terms within (8.31) are completely different. As Λ changes we therefore expect the form of the solution to change quite drastically which is the cause of the complexity in a general solution.

In order to analyze the RA model it will be necessary to do so for specific ranges of the level of the risk-aversion parameter. As they produce qualitatively very different results and generate very different mathematical challenges, our work on solving the RA problem will therefore focus on the two situations just mentioned; that when the level of risk-aversion is very large and that when it is very small. With the solution techniques developed in these

two examples, the solution of the problem for the remainder of the spectrum of Λ-values should be relatively straightforward.

In the remainder of this chapter we will solve the RA problem in the limit of a very low level of risk-aversion; in the next chapter we will then focus on solving the problem when the level of risk-aversion is very large.

8.6.2 Solving the Auxiliary Problem for Small Risk Aversion

We would like to determine the effect on the optimal strategy found in the previous chapter for the cost-minimization problem when a very small level of risk-aversion is introduced. Toward this goal we begin with the linearized minimization condition, (8.31), and set $\Lambda \ll 1$. For concreteness, let $\mu \approx r \approx \sigma^2$ so that all of $\mu\delta t$, $r\delta t$, and $\sigma^2\delta t$ are $\ll 1$. Also let $\lambda H_1 \ll 1$ and, by the same reasoning as was used in the previous chapter, $\lambda H_1 = O(\delta t)$; if we interpret λH_1 as an inverse measure of the inherent liquidity risk that our trade possesses (*i.e.* the greater either λ or H_1 means the more difficult, and therefore risky, the portfolio will be to liquidate) then this assumption can be interpreted as the trading frequency being 'pegged' to the level of liquidity risk in our portfolio. Finally, we assume $(\lambda H_1)^2 \ll \Lambda \ll \lambda H_1$ (where the first condition is required as we have already neglected the quadratic terms in the expansion of the exponential terms). In terms of the trade's liquidity, the assumption $\Lambda \ll \lambda H_1$ corresponds to the situation where the level of the trader's risk-aversion is very small compared with the overall inherent liquidity risk of the transaction.

Under the above assumptions and only retaining terms up to $O(\Lambda)$, equation (8.31) reduces to

$$(2(1+\alpha)\lambda + 4\Lambda\eta)\,\delta H^*_{N-1} + (1+\alpha)\lambda H_{N-1} - (\mu - r)\delta t + 2\Lambda\left(\eta H_{N-1} - E_1[C^*_N - C^*_{N-1}]\right) \approx 0. \tag{8.32}$$

Finally, we divide (8.32) by $(2(1+\alpha)\lambda + 4\Lambda\eta)$ and linearize this factor to obtain our final answer

$$\delta H^*_{N-1} \approx -\frac{1}{2}H_{N-1} + \frac{(\mu - r)\delta t}{2(1+\alpha)\lambda} - \frac{\Lambda}{(1+\alpha)\lambda}\left(E_1[C^*_{N-1} - C^*_N] + \frac{(\mu - r)\delta t}{(1+\alpha)\lambda} S_{N-1} e^{(-(N-2)r\delta t)}\right). \tag{8.33}$$

To be able to calculate δH^*_{N-2}, δH^*_{N-3}, ... we first need to express J_{N-1} in terms of δH^*_{N-1}.

Writing (8.28) in terms of δH^*_{N-1} we have

$$\begin{aligned} J_{N-1} = \beta_{N-1}\delta H^*_{N-1} S_{N-1} e^{\lambda\delta H^*_{N-1} - (N-2)r\delta t} + \Lambda \delta H^{*2}_{N-1} S^2_{N-1} e^{2\lambda\delta H^*_{N-1} - 2(N-2)r\delta t} \\ - \beta_N (H_{N-1} + \delta H^*_{N-1}) S_{N-1} e^{-\lambda H_{N-1} - \alpha\lambda\delta H^*_{N-1} + (\mu - (N-1)r)\delta t} \\ + \Lambda \left(H_{N-1} + \delta H_{N-1}\right)^2 S^2_{N-1} \left(e^{\mu\delta t}\left(e^{\sigma\sqrt{\delta t}} + e^{-\sigma\sqrt{\delta t}}\right) - 1\right) e^{-2\lambda H_{N-1} - 2\alpha\lambda\delta H^*_{N-1} - 2(N-1)r\delta t}. \end{aligned} \tag{8.34}$$

Now substituting (8.33) into (8.34), expanding the necessary exponential terms, and then retaining only terms up to $O(\Lambda)$ gives

$$\begin{aligned} J_{N-1} \approx -H_{N-1} S_{N-1} e^{-(N-2)r\delta t}\left(1 - \frac{1}{4}(3-\alpha)\lambda H_{N-1} + \frac{1}{2}(\mu + 3r)\delta t + \frac{(\mu - r)^2\delta t^2}{4(1+\alpha)^2 \lambda H_{N-1}}\right) \\ + \frac{1}{2}\Lambda H^2_{N-1} S^2_{N-1} e^{-2(N-2)r\delta t}\left(1 - \frac{2E_1[C^*_{N-1} - C^*_N]}{H_{N-1} S_{N-1} e^{-(N-2)r\delta t}}\right). \end{aligned} \tag{8.35}$$

With (8.33), (8.35), and the continuity relation, $H_{N-1} = H_{N-2} + \delta H_{N-2}$, we can now repeat this entire procedure; for $n = N-2$ this gives

$$\begin{aligned} \delta H^*_{N-2} \approx -\frac{1}{3}H_{N-2} + \frac{(\mu - r)\delta t}{(1+\alpha)\lambda} \\ - \frac{\Lambda}{(1+\alpha)\lambda}\left(E_1[2C^*_{N-2} - C^*_{N-1} - C^*_N] + \frac{2(\mu - r)\delta t}{(1+\alpha)\lambda} S_{N-2} e^{(-(N-3)r\delta t)}\right), \end{aligned} \tag{8.36}$$

and

$$\begin{aligned} J_{N-2} \approx -H_{N-2} S_{N-2} e^{-(N-3)r\delta t}\left(1 - \frac{1}{3}(2-\alpha)\lambda H_{N-2} + (\mu + 2r)\delta t + \frac{3(\mu - r)^2\delta t^2}{4(1+\alpha)^2 \lambda H_{N-2}}\right) \\ + \frac{1}{3}\Lambda H^2_{N-2} S^2_{N-2} e^{-2(N-3)r\delta t}\left(1 - \frac{2E_1[C_{N-2} + C^*_{N-1} - C^*_N]}{H_{N-2} S_{N-2} e^{-(N-3)r\delta t}}\right). \end{aligned} \tag{8.37}$$

Repeating this procedure for $n = N-3,\ N-4,\ \ldots$ reveals the pattern for the form of the

solution in the $N - m^{th}$ period; this is

$$\delta H^*_{N-m} \approx -\frac{1}{m+1}H_{N-m} + \frac{m(\mu-r)\delta t}{2(1+\alpha)\lambda} - \frac{\Lambda}{(1+\alpha)\lambda}\left(2E_1[C^*_{N-m}] - \frac{2}{m}E_1\left[\sum_{i=0}^{m-1} C^*_{N-i}\right] + \frac{m(\mu-r)\delta t}{(1+\alpha)\lambda}S_{N-m}e^{(-(N-(m+1))r\delta t)}\right), \tag{8.38}$$

and

$$J_{N-m} \approx -H_{N-m}S_{N-m}e^{-(N-(m+1))r\delta t}\left(1 - \frac{1}{2(m+1)}((m+2)-m\alpha)\lambda H_{N-m} + \frac{1}{2}(m\mu+(m+2)r)\delta t + \frac{m(m+2)(\mu-r)^2\delta t^2}{2(m+1)(1+\alpha)^2\lambda H_{N-2}}\right) + \frac{1}{m+1}\Lambda\left(H^2_{N-m}S^2_{N-m}e^{-2(N-(m+1))r\delta t} - 2H_{N-1}S_{N-1}e^{-(N-(m+1)r\delta t)}E_1\left[\sum_{i=1}^{m} C^*_{N-i} - C^*_N\right]\right). \tag{8.39}$$

While equation (8.38) *is* the optimal trading strategy for the auxiliary problem of the RA problem, it can be put into a more tractable form. But first we will simplify equation (8.38) with

$$x = \frac{(\mu-r)\delta t}{2(1+\alpha)\lambda}, \tag{8.40}$$

$$y_m = 2E_1\left[C^*_{N-m} - \frac{1}{m}\sum_{i=0}^{m-1} C^*_{N-i}\right]. \tag{8.41}$$

Equation (8.38) then reduces to

$$\delta H^*_{N-m} \approx -\frac{1}{m+1}H_{N-m} + xm - \frac{\Lambda}{(1+\alpha)\lambda}\left(y_m + 2xS_{N-m}e^{-(N-(m+1))r\delta t}m\right) \tag{8.42}$$

As we did in the previous chapter, we begin at the first trading period when $m = N-1$; for this period the optimal trade quantity is

$$\delta H^*_1 \approx -\frac{1}{N}H_1 + x(N-1) - \frac{\Lambda}{(1+\alpha)\lambda}\left(y_{N-1} + 2(N-1)xS_1\right). \tag{8.43}$$

We now use the continuity relation, $H_2 = H_1 + \delta H^*_1$, to get the number of assets remaining

in the portfolio at the beginning of the second period; this is

$$H_2 \approx \frac{N-1}{N} H_1 + x(N-1) - \frac{\Lambda}{(1+\alpha)\lambda} \left(y_{N-1} + 2(N-1)xS_1 \right). \tag{8.44}$$

The optimal trade quantity in the second period is, from equation (8.42),

$$\delta H_2^* \approx -\frac{1}{N-1} H_2 + x(N-2) - \frac{\Lambda}{(1+\alpha)\lambda} \left(y_{N-2} + 2(N-2)xS_2 e^{-r\delta t} \right). \tag{8.45}$$

Substituting (8.44) into (8.45) and simplifying gives

$$\delta H_2^* \approx -\frac{1}{N} H_1 + x(N-3) - \frac{\Lambda}{(1+\alpha)\lambda} \left(y_{N-2} - \frac{1}{N-1} y_{N-1} - 2x \left(S_1 - (N-2)e^{-r\delta t} S_2 \right) \right). \tag{8.46}$$

Repeating this process gives us the general form for δH_n^* for all $n \in \{1, \ldots, N\}$; this is

$$\begin{aligned} \delta H_n^* \approx -\frac{1}{N} H_1 + (N - (2n-1)) \frac{(\mu - r)\delta t}{2(1+\alpha)\lambda} - \frac{\Lambda}{(1+\alpha)\lambda} \Bigg(& y_{N-n} - \sum_{j=1}^{n-1} \frac{1}{N-j} y_{N-j} \\ & - \frac{(\mu - r)\delta t}{(1+\alpha)\lambda} \left((1 - \delta_{n-1}) \sum_{j=1}^{n-1} S_j e^{-(j-1)r\delta t} - (N-n) e^{-(n-1)r\delta t} S_n \right) \Bigg), \end{aligned} \tag{8.47}$$

where

$$\delta_{n-1} = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n \neq 1. \end{cases} \tag{8.48}$$

The presence of the $1 - \delta_{n-1}$ term is a result of the fact that the $\sum S_j$ term is only present for $n > 1$. But from the definition of y_m in (8.41) we can simplify the above expression as

$$\begin{aligned} y_{N-n} &= 2E_1 \left[C_n^* - \frac{1}{N-n} \sum_{i=n+1}^{N} C_i^* \right]. \\ &\equiv Y_n. \end{aligned} \tag{8.49}$$

With this redefinition we finally have

$$\delta H_n^* \approx -\frac{1}{N}H_1 + (N-(2n-1))\frac{(\mu-r)\delta t}{2(1+\alpha)\lambda} - \frac{\Lambda}{(1+\alpha)\lambda}\left(Y_n - \sum_{j=1}^{n-1}\frac{1}{N-j}Y_j \right.$$
$$\left. -\frac{(\mu-r)\delta t}{(1+\alpha)\lambda}\left((1-\delta_{n-1})\sum_{j=1}^{n-1}S_j e^{-(j-1)r\delta t} - (N-n)e^{-(n-1)r\delta t}S_n\right)\right). \quad (8.50)$$

which is the optimal solution of the auxiliary RA problem for a liquidation transaction when the risk-aversion is small.

8.6.3 Finding the Solution of the RA Problem from the Auxiliary Problem

To find the solution to the RA problem we must first evaluate the $E_1[C^*]$'s within (8.50) which are the remnants of the $\beta_n(\pi^*)$'s.

Since δH_n^* depends on Y_j for $j = 1, \ldots, n$ and since Y_j depends on $E_1[C_i^*]$ for $i = j, \ldots, N$, it follows that all δH_n^* depend on every $E_1[C_i^*]$. Crucially then, since $C_n^* = \delta H_n^* S_n e^{\lambda\delta H_n^* - (n-1)r\delta t}$ we have the interesting situation in which δH_n^* depends on the initial expectation of its own value. As we will see in the next chapter, removing this dependence is not a trivial matter, yet doing so is crucial to our model's ability to generate useful results.

In the current case of small Λ we are fortunate that a closed-form solution is possible; the reason for this is that the $E_1[C^*]$ contributions are all multiplied by $\Lambda/((1+\alpha)\lambda)$ and are therefore small. We begin by evaluating $E_1[C_i^*]$; we have

$$\begin{aligned} E_1[C_i^*] &= E_1[\delta H_i^* S_i e^{\lambda\delta H_i^* - (i-1)r\delta t}], \\ &= e^{-(i-1)r\delta t}E_1[S_i\delta H_i^*(1+\lambda\delta H_i^* + \ldots)], \\ &\approx e^{-(i-1)r\delta t}E_1[S_i\delta H_i^*]. \end{aligned} \quad (8.51)$$

We have neglected all but the leading-order term of $e^{\lambda\delta H_i^*}$ since these terms will be multiplied by $\Lambda((1+\alpha)\lambda)$ from equation (8.50) and will therefore make a negligible contribution; we have retained the $e^{-(i-1)r\delta t}$ term, however, as it will factor into the next step. Substituting

(8.50) into (8.51) then gives

$$\begin{aligned}
E_1[C_i^*] &\approx e^{-(i-1)r\delta t} E_1\left[-\frac{1}{N}H_1 S_i + (N-(2i-1))\, S_i \frac{(\mu-r)\delta t}{2(1+\alpha)\lambda}\right], \\
&= e^{-(i-1)r\delta t}\left(-\frac{1}{N}H_1 S_i + (N-(2i-1))\,\frac{(\mu-r)\delta t}{2(1+\alpha)\lambda}\right) E_1[S_i], \\
&\approx S_1 e^{(i-1)(\mu-r)\delta t}\left(-\frac{1}{N}H_1 + (N-(2i-1))\,\frac{(\mu-r)\delta t}{2(1+\alpha)\lambda}\right), \\
&\approx S_1\left(-\frac{1}{N}H_1 + (N-(2i-1))\,\frac{(\mu-r)\delta t}{2(1+\alpha)\lambda}\right), \qquad (8.52)
\end{aligned}$$

where again we have only retained leading-order contributions.

We now use (8.52) in the definition of Y_j (8.49) to obtain

$$\begin{aligned}
Y_j &= 2E_1\left[C_j^* - \frac{1}{N-j}\sum_{k=j+1}^{N} C_k^*\right], \\
&= 2S_1\Bigg(-\frac{H_1}{N} + (N-(2j-1))\frac{(\mu-r)\delta t}{2(1+\alpha)\lambda} \\
&\qquad\qquad -\frac{1}{N-j}\sum_{k=j+1}^{N}\left(-\frac{H_1}{N} + (N-(2k-1))\frac{(\mu-r)\delta t}{2(1+\alpha)\lambda}\right)\Bigg), \\
&= -\frac{4(\mu-r)\delta t}{(1+\alpha)\lambda} S_1\left(j - \frac{1}{N-j}\sum_{k=j+1}^{N} k\right), \\
&= 2\,(N-(j+1))\, S_1 \frac{(\mu-r)\delta t}{(1+\alpha)\lambda}. \qquad (8.53)
\end{aligned}$$

Using (8.53) we can now write

$$Y_n - \sum_{j=1}^{n-1}\frac{1}{N-j}Y_j \approx 2\,(N-(2n-1))\,\frac{(\mu-r)\delta t}{(1+\alpha)\lambda}, \qquad (8.54)$$

and finally combining (8.54) and (8.50) we get

$$\delta H_n^* \approx -\frac{1}{N}H_1 + (N-(2n-1))\,\frac{(\mu-r)\delta t}{2(1+\alpha)\lambda}\left(1-\frac{4\Lambda S_1}{(1+\alpha)\lambda}\right) + \frac{\Lambda(\mu-r)\delta t}{(1+\alpha)^2\lambda^2}((1-\delta_{n-1})\sum_{j=1}^{n-1}S_j e^{-(j-1)r\delta t} - (N-n)e^{-(n-1)r\delta t}S_n), \quad (8.55)$$

which is our final expression for the optimal trading strategy for the Risk-Adjusted problem for $\Lambda \ll \lambda H_1$.

8.7 Discussion

In (8.55) we can see that the leading-order component of the optimal trading strategy when risk-aversion is small is the same as in the cost-minimization problem. The correction to this value is $O(\Lambda/\lambda)$ as we expect and arises from two sources.

One of the sources of the correction enters the solution through the $(N-(2n-1))\frac{(\mu-r)\delta t}{2(1+\alpha)\lambda}$ term. As we mentioned in the previous chapter this term arises from the balance between the expected growth rate of the asset above the risk-free rate and the depression of its value due to our trading. Because we generate more revenue by delaying the sale of an asset when μ is large this leading-order effect causes the fewest assets to be sold at the start of the transaction and the most at the end. As a result of its sign the $O(\Lambda/\lambda)$ correction to this effect moves the optimal strategy opposite to the leading-order behaviour thus slightly increasing the number of assets sold at the start of the transaction and decreasing the number at the end as we would expect as the correction is proportional to our level of risk-aversion.

This *static* correction (why we have chosen this name will become apparent below) has a very interesting interpretation. From our work in Section 8.6.3 it is clear that the static correction results from the Y terms in equation (8.50). From its definition in (8.49), Y_n is our initial expectation of the difference between the cost of the strategy in the n^{th} period and the average cost of all future periods. If we generally expect the revenue streams to decrease from period n to N then (as cost = −revenue) $Y_n < 0$, whereas if we expect the revenue streams to increase then $Y_n > 0$. As the correction results from the term $Y_n - \sum_{j=1}^{n-1}\frac{1}{N-j}Y_j$ it is therefore the difference between the 'current' value of Y and a weighted (more heavily

for periods close to n) average of all previous values of Y. This correction is therefore a running indicator of how our initial expectations about the dynamic behaviour of the revenue streams has changed since the transaction began.

The other correction,

$$\frac{\Lambda(\mu - r)\delta t}{(1+\alpha)^2\lambda^2}\left(\sum_{j=1}^{n-1} S_j e^{-(j-1)r\delta t} - (N-n)e^{-(n-1)r\delta t}S_n\right), \tag{8.56}$$

is interesting as it causes the static nature of the cost-minimization strategy to become dynamic and it will therefore be referred to as the *dynamic* correction. This dynamic behaviour results from the n^{th} term's dependence on the realised value of the asset price, S_n, and means that the exact strategy cannot be determined before the transaction begins. Not only is the optimal strategy dependent on S_n, however, but also on every value of S_i for $i = 1, \ldots, n-1$ and therefore we have the very interesting result that *not only is the optimal strategy of the RA problem dynamic, but it is also fully path-dependent.*

While it is clear from (8.56) that the dynamic correction will be large and negative at the beginning of the transaction and large and positive at the end of the transaction, its precise effect cannot be determined without considering a specific asset price process due to its path-dependence. To demonstrate its impact more concretely we show plots of the dynamic correction over the length of a transaction for several realizations of the asset price process in Figure 8.1.

The plots in Figure 8.1 show that, as expected, the dynamic correction is negative at $n = 1$ and increases monotonically towards $n = N$. It is also clear from the figure that the overall path-dependency within the optimal strategy will be small as the dynamic correction appears quite insensitive to the realized asset price path. To demonstrate the path-dependent component explicitly, we show the difference between the dynamic correction for each of the two asset price paths from Figure 8.1 and the dynamic correction for a time-independent asset price process (*i.e.* $S_n = 100$ for all $n = 1, \ldots, N$) in Figure 8.2. This plot shows that indeed the path-dependent component of the dynamic correction is small and, in general, is negative (positive) for small (large) n when the price process is increasing; this trend is reversed when the process is decreasing.

At first sight, the fact that the optimal solution, (8.55), is path-dependent is not that

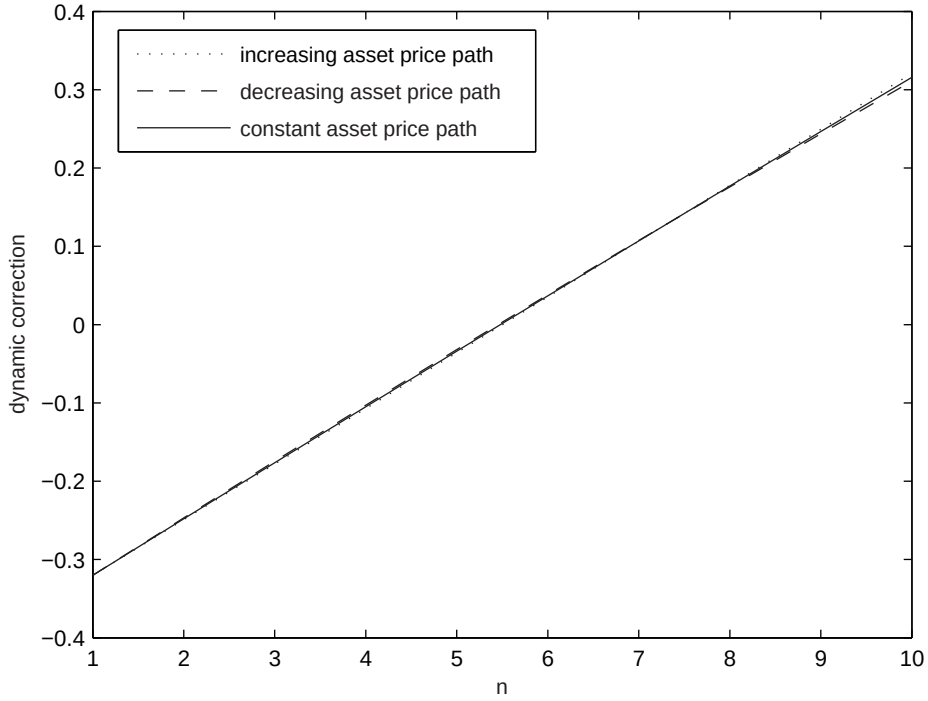


Figure 8.1: The dynamic correction to the leading-order optimal strategy for several asset price paths with small Λ/λ. The dotted line represents the steadily increasing price process, $S_n = S_0 e^{-(n-1)\delta t}$ for all $n = 1, \ldots, 10$, the dashed line represents the steadily decreasing price process, $S_n = S_0 e^{(n-1)\delta t}$ for all $n = 1, \ldots, 10$, and the solid line represents the constant price process, $S_n = S_0$ for all $n = 1, \ldots, 10$. The parameter values used are $\Lambda = 10^{-4}$, $\lambda = 10^{-2}$, $\mu = 0.1$, $r = 0.08$, $\delta t = 0.04$, and $S_0 = 100$.

surprising given that our model for the dynamics of the underlying price possesses a strong trading-induced feedback mechanism. But, as was mentioned in Section 8.3, in reducing the mean-variance problem to the the risk-adjusted problem we have neglected the serial correlations induced by this feedback mechanism. The fact that path-dependency exists in the optimal strategy even though no serial correlations exist in the underlying price process is interesting.

The path-dependency that exists in the small-Λ optimal strategy enters through the presence of the realized values of all past asset prices. If we re-examine the process, (8.42)-(8.47), of re-expressing the optimal strategy in terms of the initial quantity, H_1, we see

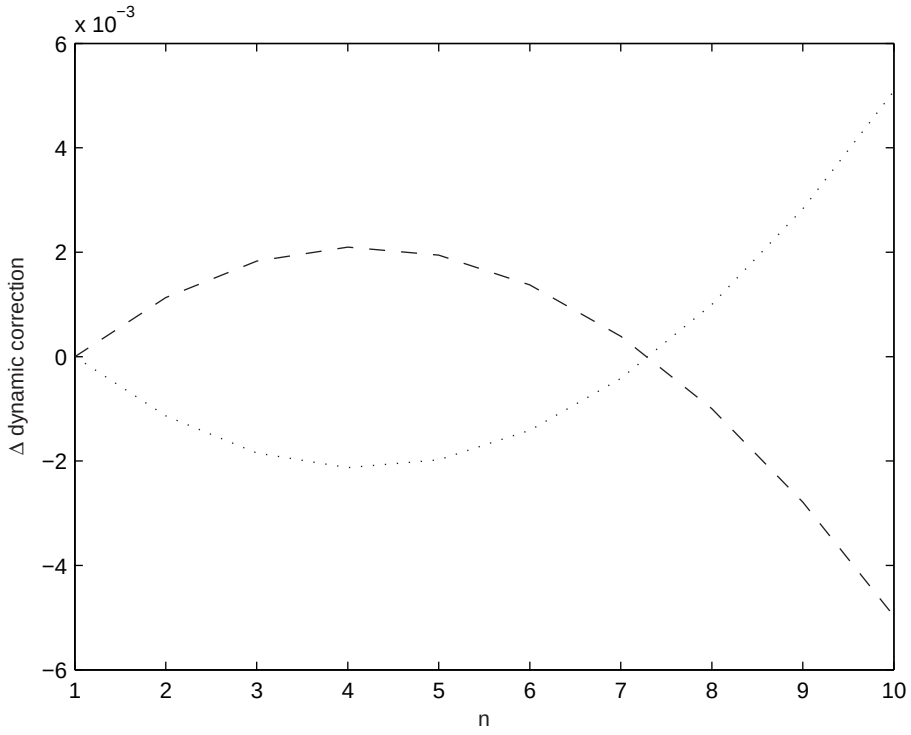


Figure 8.2: The difference between the dynamic correction for the increasing price process and the constant price process (dotted line) and the difference between the dynamic correction for the decreasing price process and the constant price process (dashed line). The parameter values used are $\Lambda = 10^{-4}$, $\lambda = 10^{-2}$, $\mu = 0.1$, $r = 0.08$, $\delta t = 0.04$, and $S_0 = 100$.

that these past prices naturally propagate through the calculation. This demonstrates that serial correlations in the underlying process are not necessary to produce a dynamic, path-dependent optimal strategy; these characteristics arise as a direct byproduct of the dynamic nature of the strategy itself which in turn is a result of the quadratic nature of the objective function.

Chapter 9

Optimal Liquidation with Large Risk Aversion

In the previous chapter we formulated the stochastic optimal control problem for the general Risk-Adjusted (RA) problem. After linearizing the problem it was found that a large and cumbersome quadratic equation needed to be solved for each trading period to determine the optimal strategy. To extract the qualitative features of the solution it was necessary to split the range of Λ into several 'bands' and focus on the problem in each of these regions.

The previous chapter focused on the solution of the RA problem for the smallest levels of risk aversion. In this chapter we will solve for the optimal trading strategy at the opposite end of the Λ spectrum when the level of risk aversion is *very* large. We have chosen this large-Λ example for three reasons. First, being at the opposite end of the Λ spectrum the solution of this problem will allow us to compare and contrast the optimal strategies for the highest and lowest levels of risk aversion. Second, in the large-Λ limit we recover the pure risk-minimization problem; that is, to determine the trading strategy that simply minimizes the time risk associated with the strategy. This risk-minimization problem (closely related to the variance minimization problem) is notoriously difficult and, to our knowledge, has not been examined for the optimal execution problem. Finally, in subsection 8.6.3 we found a closed-form solution to the RA problem from the auxiliary solution by simply substituting the solution into itself and eliminating negligible terms; later in this chapter we will show that this procedure is not possible for general Λ. To generate a solution to the RA problem

in these cases we must instead develop an algorithm that generates the solution numerically. The large-Λ limit is the simplest form of the problem when a closed-form solution to the RA problem is not possible and it will therefore allow us to more clearly demonstrate the numerical algorithm.

Even though the RA problem for large Λ is more simple than in the interior regions of the Λ space, we will see that calculating its solution is much more involved than in the small-Λ case of Chapter 8. To avoid an uninterpretable solution it will therefore be necessary to make a further simplification to the large-Λ problem. *In this chapter we will focus on the problem of optimally liquidating a portfolio in a perfectly liquid market when the only goal is to minimize the transaction's time risk.* This chapter can therefore be thought of as a first step toward finding the finite liquidity optimal strategy for large Λ. As was mentioned in the previous two chapters, one advantage of our model is that more realistic solutions 'simply' involve retaining more terms in the expansion of its nonlinear terms; while this may be complicated and time consuming, with symbolic algebra programs such as MAPLE it should be possible.

The rest of this chapter will be organized as follows: In Section 9.1 we will state the full (*i.e.* finite liquidity) problem for the large-Λ limit and demonstrate the difficulty in its solution; we end the section by presenting the simplified (*i.e.* perfect liquidity) optimization problem that we will solve subsequently. In Section 9.2 we then solve this simplified problem using the auxiliary method developed in Chapter 8 and discuss the results in Section 9.3. In Section 9.4 we then develop the algorithm that will be used to generate the numerical solution of the RA problem from the solution of this auxiliary. In Section 9.5 we present numerical solutions to the large-Λ RA problem and discuss the results.

9.1 A Simplified Risk-Adjusted Problem for Large Λ

We begin with the Bellman equation for the auxiliary to the RA problem for general Λ developed in the previous chapter; this is

$$\begin{aligned} J_n &= \min_{\delta H_n} \left\{ \beta_n^* \delta H_n S_n e^{\lambda \delta H_n - (n-1) r \delta t} + \Lambda \delta H_n^2 S_n^2 e^{2\lambda \delta H_n - 2(n-1) r \delta t} + E_n[J_{n+1}] \right\}, \\ \beta_n^* &= 1 - 2\Lambda E_1[C_n^*], \qquad C_n^* = \delta H_n^* S_n e^{\lambda \delta H_n - (n-1) r \delta t}, \\ H_{n+1} &= H_n + \delta H_n, \qquad H_{N+1} = 0, \end{aligned} \tag{9.1}$$

where H_1 is specified, δH_n^* is the optimal value of δH_n for the n^{th} trading period, and the quadratic risk term contributes positively to the objective function since the problem is specific for a liquidation transaction.

To demonstrate the difficulty in solving (9.1) even for a linearized impact function we will focus on its solution for the $N-1^{\text{th}}$ period. For a liquidation transaction we have the final condition $H_{N+1} = 0$ which further implies that $J_{N+1} = 0$. To satisfy these conditions we must have

$$\delta H_N^* = -H_N, \tag{9.2}$$

which gives

$$J_N = -\beta_N^* H_N S_N e^{\lambda H_N - (N-1) r \delta t} + \Lambda H_N^2 S_N^2 e^{2\lambda H_N - 2(N-1) r \delta t}. \tag{9.3}$$

For $n = N-1$ and with (9.3) substituted into the Bellman equation of (9.1), then after taking expectations, linearizing the exponential terms (except the discounting factor), and differentiating the resulting expression *w.r.t.* δH_{N-1} we obtain the first-order minimization condition for J_{N-1} which is given in (8.31).

For $\Lambda \gg (\{r, \mu\}\delta t)$ and $\Lambda \gg \lambda H_1$ (*i.e.* the large-Λ limit), all non-Λ terms can be

eliminated from (8.31) and we are left with

$$
\begin{aligned}
&-6\eta(1-\alpha)\lambda\delta H_{N-1}^{*2} \\
&\quad+\left(4\eta\left(1-(1+2\alpha)\lambda H_{N-1}+\left(\mu-r+\frac{1}{2}\sigma^2\right)\delta t\right)-4\lambda E_1[C_{N-1}^*+\alpha C_N^*]\right)\delta H_{N-1}^* \\
&\qquad+2\eta H_{N-1}\left(1-(2+\alpha)\lambda H_{N-1}+2\left(\mu-r+\frac{1}{2}\sigma^2\right)\delta t\right) \\
&\qquad-2E_1[C_N^*]\left(1-(1+\alpha)\lambda H_{N-1}+(\mu-r)\delta t\right)+2E_1[C_{N-1}^*]=0. \quad (9.4)
\end{aligned}
$$

Up to the first correction, the solution of (9.4) is

$$
\begin{aligned}
\delta H_{N-1}^* \approx &-\frac{1}{2}H_{N-1}+\frac{1}{2\eta}E_1[C_N^*-C_{N-1}^*] \\
&-\left(\frac{1}{2\eta}E_1[C_N^*-C_{N-1}^*]-\frac{1}{2}H_{N-1}\right)\left(\left(\mu-r+\frac{1}{2}\sigma^2\right)\delta t-4\lambda E_1[C_{N-1}^*+\alpha C_N]\right. \\
&\left.-(1+2\alpha)\lambda H_{N-1}\right)+\frac{3}{2}(1-\alpha)\lambda\left(\frac{1}{2\eta}E_1[C_N^*-C_{N-1}^*]-\frac{1}{2}H_{N-1}\right)^2 \\
&-\frac{1}{2}H_{N-1}\left(2\left(\mu-r+\frac{1}{2}\sigma^2\right)\delta t-(2+\alpha)\lambda H_{N-1}\right) \\
&-\frac{1}{2\eta}E_1[C_N^*]\left((1+\alpha)\lambda H_{N-1}-(\mu-r)\delta t\right). \quad (9.5)
\end{aligned}
$$

As we found in the previous two chapters, solving for the optimal strategy, $\{\delta H_{N-m}^*\}$, involves iteratively solving the Bellman equation backward in time. For a static optimal strategy, as was the case in the cost-minimization setting of Chapter 7, the number of terms in the expression for each δH_{N-m}^* is a constant. When the solution is dynamic, however, as was the case in Chapter 8, the number of terms in the expression for δH_{N-m}^* increases linearly with m. As (9.5) is S-dependent (through η) the optimal solution of the large-Λ problem will also be dynamic, but with the extra complication that the quadratic term in (9.4) now makes a contribution. The result of this contribution in (9.5) is the presence of the $E_1[C_N^*]^2$, $E_1[C_{N-1}^{*2}]$, and $E_1[C_N^*]E_1[C_{N-1}^*]$ terms. As m increases the number of terms in the solution will increase as m^2; since the solution for the $N-1^{\text{th}}$ period is already quite complicated, the process of finding the solution for general m will be *very* involved.

While the δt terms contribute to the complicated nature of (9.5), it is the λ contri-

butions that cause the quadratic increase in the number of terms in the solution. As a first step toward solving this complicated large-Λ problem, we will neglect these liquidity contributions in this chapter and simply solve the problem of finding the optimal strategy in a perfectly liquid market when time-risk minimization is the only concern. In this much simplified framework we have two main goals: To develop the method for solving the large-Λ problem (which we will see is distinctly different to the small-Λ problem of Chapter 8), and to demonstrate that this method gives a solution that agrees with our intuition. With this method and the perfectly liquid solution we hope to be able to examine the finite liquidity case in a future study.

When $\Lambda \gg 1$ and $\lambda = 0$ the optimization problem (9.1) for the liquidation transaction reduces to

$$\begin{aligned} J_n = \min_{\delta H_n} \left\{ \Lambda \delta H_n^2 S_n^2 e^{-2(n-1)r\delta t} - 2\Lambda E_1[C_n^*]\delta H_n S_n e^{-(n-1)r\delta t} + E_n[J_{n+1}] \right\}, \\ C_n^* = \delta H_n^* S_n e^{-(n-1)r\delta t}, \\ H_{n+1} = H_n + \delta H_n, \qquad H_{N+1} = 0, \end{aligned} \tag{9.6}$$

where

$$S_{n+1} = \begin{cases} u \cdot S_n & \text{probability } p, \\ d \cdot S_n & \text{probability } 1-p, \end{cases} \tag{9.7}$$

and, as usual, u, d, and p are calibrated to a GBM with drift μ and volatility σ. Solving (9.6)-(9.7) is the goal of this chapter.

9.2 Solution of the Auxiliary for the Perfectly Liquid RA Problem

If we set $\lambda = 0$ then (9.5) reduces to

$$\begin{aligned} \delta H_{N-1}^* = &-\frac{1}{2} H_{N-1} + \frac{1}{2\eta} E_1[C_N^* - C_{N-1}^*] \\ &- \frac{1}{2}\left(\mu - r + \frac{1}{2}\sigma^2\right)\delta t \left(H_{N-1} + \frac{1}{\eta} E_1[C_N^* - C_{N-1}^*]\right) + \frac{1}{2\eta} E_1[C_N^*](\mu - r)\delta t. \end{aligned} \tag{9.8}$$

The method used in the previous chapters to determine the optimal trade quantity for the general $N-m^{\text{th}}$ period was to calculate δH^*_{N-m} explicitly for $m = 1, 2, \ldots$ and from these solutions determine the pattern of the factor in front of each term. Because the terms $\mu - r + 1/2\sigma^2$ and $\mu - r$ occur independently and because of the two terms involving $E_1[C^*_N - C^*_{N-1}]$ in (9.8), if we attempt this method for the current problem we will find that the solutions δH^*_{N-2}, δH^*_{N-3}, ... quickly become unmanageable . We must therefore be more careful in proceeding with this case and toward this end we note that the optimal value function for the $n+1^{\text{th}}$ period can be written in general (see, for example, equation (9.3)) as

$$J_{n+1} = \Lambda \frac{A_{n+1}}{B_{n+1}} S_{n+1} H_{n+1} e^{-nr\delta t} + \Lambda \frac{1}{B_{n+1}} S^2_{n+1} H^2_{n+1} e^{-2nr\delta t} - \Lambda F_{n+1}, \tag{9.9}$$

for arbitrary functions A_{n+1}, B_{n+1}, and F_{n+1} that are predictable *w.r.t.* the information known at the beginning of period n. Our goal now is to find A, B, and F for the large-Λ problem.

Substituting (9.9) into (9.6) gives the expression for the Bellman equation at the n^{th} period as

$$\begin{aligned} J_n &= \min_{\delta H_n} \Big\{ -2\Lambda E_1[C^*_n]\delta H_n S_n e^{-(n-1)r\delta t} + \Lambda \delta H^2_n S^2_n e^{-2(n-1)r\delta t} \\ &\qquad + \Lambda \frac{A_{n+1}}{B_{n+1}} E_n[S_{n+1}] H_{n+1} e^{-nr\delta t} + \Lambda \frac{1}{B_{n+1}} E_n[S^2_{n+1}] H^2_{n+1} e^{-2nr\delta t} - \Lambda F_{n+1} \Big\}, \\ &= \min_{\delta H_n} \Big\{ -2\Lambda a_{n-1} E_1[C^*_n]\delta H_n S_n + \Lambda a^2_{n-1} \delta H^2_n S^2_n \\ &\qquad + \Lambda c_2 a_n \frac{A_{n+1}}{B_{n+1}} S_n H_{n+1} + \Lambda c_1 a^2_n \frac{1}{B_{n+1}} S^2_n H^2_{n+1} - \Lambda F_{n+1} \Big\}, \end{aligned} \tag{9.10}$$

where

$$c_1 = e^{\mu\delta t}\left(e^{\sigma\sqrt{\delta t}} + e^{-\sigma\sqrt{\delta t}}\right) - 1, \tag{9.11}$$

$$c_2 = e^{\mu\delta t}, \tag{9.12}$$

$$a_n = e^{-nr\delta t}. \tag{9.13}$$

Applying the continuity relation, $H_{n+1} = H_n + \delta H_n$, to (9.10) gives

$$J_n = \min_{\delta H_n}\Big\{-2\Lambda a_{n-1}E_1[C_n^*]\delta H_n S_n + \Lambda a_{n-1}^2 \delta H_n^2 S_n^2 \\ +\Lambda c_2 a_n \frac{A_{n+1}}{B_{n+1}} S_n \left(H_n + \delta H_n\right) + \Lambda c_1 a_n^2 \frac{1}{B_{n+1}} S_n^2 \left(H_n + \delta H_n\right)^2 - \Lambda F_{n+1}\Big\}. \quad (9.14)$$

Differentiating (9.14) and solving $\frac{\partial J_n}{\partial(\delta H_n)}\Big|_{\delta H_n^*} = 0$ gives the optimal trade quantity, δH_n^*, for the n^{th} period; this is

$$\delta H_n^* = -\frac{c_1 a_n^2}{c_1 a_n^2 + a_{n-1}^2 B_{n+1}} H_n + \frac{2a_{n-1}E_1[C_n^*]B_{n+1} - c_2 a_n A_{n+1}}{2S_n(c_1 a_n^2 + a_{n-1}^2 B_{n+1})}. \quad (9.15)$$

Finally we substitute (9.15) back into (9.14); after some algebra we obtain the relation

$$J_n = \frac{\Lambda}{c_1 a_n^2 + a_{n-1}^2 B_{n+1}}\Bigg(\left(2c_1 a_n^2 a_{n-1} E_1[C_n^*] + c_2 a_n a_{n-1}^2 A_{n+1}\right) S_n H_n \\ +c_1 a_n^2 a_{n-1}^2 S_n^2 H_n^2 - \frac{1}{4}\left(2a_{n-1}E_1[C_n^*]B_{n+1} - c_2 a_n A_{n+1}\right)^2\Bigg) - \Lambda F_{n+1}. \quad (9.16)$$

From our assumed form of J in (9.9) we can also write J_n as

$$J_n = \Lambda \frac{A_n}{B_n} S_n H_n a_{n-1} + \Lambda \frac{1}{B_n} S_n^2 H_n^2 a_{n-1}^2 - \Lambda F_n. \quad (9.17)$$

By comparing terms in expressions (9.16) and (9.17) we have

$$A_n = 2c_1 a_n^2 E_1[C_n^*] + c_2 a_n a_{n-1} A_{n+1}, \quad (9.18)$$

$$B_n = 1 + \frac{a_{n-1}^2}{c_1 a_n^2} B_{n+1}, \quad (9.19)$$

$$F_n = F_{n+1} + \frac{\left(2a_{n-1}E_1[C_n^*]B_{n+1} - c_2 a_n A_{n+1}\right)^2}{4\left(c_1 a_n^2 + a_{n-1}^2 B_{n+1}\right)}. \quad (9.20)$$

Expressions (9.18), (9.19), and (9.20) are recurrence relations for the three functions A_n, B_n, and F_n. Our goal now is to solve these equations for general n; with this result and equation (9.15) we will then have the desired optimal strategy.

Before we can proceed, however, we must calibrate the recurrence relations ; we do this by calculating the explicit values of δH_n^* and J_n for $n = N - 1$. Using the usual method we

readily find

$$\delta H^*_{N-1} = -\frac{c_1 a^2_{N-1}}{a^2_{N-2} + c_1 a^2_{N-1}} H_{N-1} + \frac{E_1[a_{N-2} C^*_{N-1} - c_2 a_{N-1} C^*_N]}{S_{N-1}\left(a^2_{N-2} + c_1 a^2_{N-1}\right)}, \tag{9.21}$$

and

$$J_{N-1} = \frac{\Lambda}{a^2_{N-2} + c_1 a^2_{N-1}} \Big(\left(2c_1 a^2_{N-1} a_{N-2} E_1[C^*_{N-1}] + 2c_2 a_{N-1} a^2_{N-2} E_1[C^*_N]\right) S_{N-1} H_{N-1} \\ + c_1 a^2_{N-1} a^2_{N-2} S^2_{N-1} S^2_{N-1} H^2_{N-1} - \left(a_{N-2} E_1[C^*_{N-1}] - c_2 a_{N-1} E_1[C^*_N]\right)^2 \Big). \tag{9.22}$$

Comparing (9.21) and (9.21) with (9.15) and (9.16) we get

$$A_N = 2E_1[C^*_N], \tag{9.23}$$

$$B_N = 1, \tag{9.24}$$

$$F_N = 0. \tag{9.25}$$

To derive the general expression for A_n we begin with (9.18); at $n = N-1$ we have

$$A_{N-1} = 2c_1 a^2_{N-1} E_1[C^*_{N-1}] + c_2 a_{N-1} a_{N-2} A_N. \tag{9.26}$$

We now substitute (9.23) into (9.26) and use the definitions, (9.11)-(9.13), to give

$$A_{N-1} = 2e^{-2(N-1)r\delta t}\left(e^{\mu\delta t}\left(e^{\sigma\sqrt{\delta t}} + e^{-\sigma\sqrt{\delta t}}\right) - 1\right) E_1[C^*_{N-1}] + 2e^{(-2Nr+3r+\mu)\delta t} E_1[C^*_N]. \tag{9.27}$$

To derive a tractable result we must simplify the $e^{\mu\delta t}\left(e^{\sigma\sqrt{\delta t}} + e^{-\sigma\sqrt{\delta t}}\right) - 1$ term in the above expression; toward this goal we use the assumption $\delta t \ll 1$. Expanding the exponential terms in (9.27), neglecting all $o(\delta t)$ terms, and re-expressing the $O(\delta t)$ expansions in exponential form (*i.e.* $1 + x\delta t = e^{x\delta t} + o(\delta t)$) for the sake of conciseness; (9.27) then reduces to

$$A_{N-1} = 2e^{-2((N-1)r - z)\delta t}\left(E_1[C^*_{N-1}] + e^{(r+\mu-2z)\delta t} E_1[C^*_N]\right) + o(\delta t), \tag{9.28}$$

where

$$z = \mu + \frac{1}{2}\sigma^2. \tag{9.29}$$

At the next step when $n = N - 2$ equation (9.18) is

$$A_{N-2} = 2c_1 a_{N-2}^2 E_1[C_{N-2}^*] + c_2 a_{N-2} a_{N-3} A_{N-1}, \tag{9.30}$$

which, after substituting in the expressions for c_1, c_2, a_n, and (9.28) for A_{N-1}, becomes

$$\begin{aligned} A_{N-2} = 2e^{(-2Nr+4r)\delta t}\left(e^{\mu\delta t}\left(e^{\sigma\sqrt{\delta t}} + e^{-\sigma\sqrt{\delta t}}\right) - 1\right) E_1[C_{N-2}^*] \\ + 2e^{(-4Nr+7r+2z+\mu)\delta t} E_1[C_{N-1}^*] + 2e^{(-4Nr+8r+2\mu)\delta t} E_1[C_N^*]. \end{aligned} \tag{9.31}$$

Again, we now approximate (9.31) to $O(\delta t)$ by expanding exponentials, neglecting $o(\delta t)$ terms, and then re-expressing in terms of exponentials; our final expression for A_{N-2} is then

$$\begin{aligned} A_{N-2} = 2e^{-2((N-2)r-z)\delta t}\Big(E_1[C_{N-2}^*] \\ + e^{(-(2N-3)r+\mu)\delta t} E_1[C_{N-1}^*] + e^{(-2((N-2)r-\mu))\delta t} E_1[C_N^*]\Big) + o(\delta t). \end{aligned} \tag{9.32}$$

Repeating this procedure gives the general result

$$\begin{aligned} A_{N-m} = 2e^{-m((2N-(m+2))r-\mu)\delta t} E_1[C_N^*] \\ + 2e^{-2((N-m)r-z)\delta t} \sum_{i=1}^{m} e^{-(m-i)((2N-(m+i))r-\mu)\delta t} E_1[C_{N-i}^*] + o(\delta t). \end{aligned} \tag{9.33}$$

Equation (9.33) can be put in a more compact form if we re-express the $E_1[C_N^*]$ exponential factor in (9.33) as

$$e^{-m((2N-(m+2))r-\mu)\delta t} = e^{-2((N-m)r-z)\delta t} e^{-(2mNr-m^2r-m\mu)\delta t} + 2(Nr - z)\delta t + o(\delta t). \tag{9.34}$$

The $E_1[C_N^*]$ term can now be incorporated into the summation term in (9.33) and the general expression for A_{N-m} reduces to

$$A_{N-m} = 2e^{-2((N-m)r-z)\delta t} \sum_{i=0}^{m} e^{-(m-i)((2N-(m+i))r-\mu)\delta t} E_1[C_{N-i}^*] + 4(Nr-z)E_1[C_N^*]\delta t + o(\delta t), \tag{9.35}$$

which is valid for $m = 1, \ldots, N-1$.

To find the general expression for B we follow the same steps as for A above; beginning with (9.19) and the final condition (9.24) we have

$$\begin{aligned} B_{N-1} &= 1 + \frac{a_{N-2}^2}{c_1 a_{N-1}^2}, \\ &= 1 + \frac{e^{-2(N-2)r\delta t}}{\left(e^{\mu\delta t}\left(e^{\sigma\sqrt{\delta t}} + e^{-\sigma\sqrt{\delta t}}\right) - 1\right) e^{-2(N-1)r\delta t}}, \\ &= 1 + e^{2(r-z)\delta t} + o(\delta t), \\ &= 2 + 2(r-z)\delta t + o(\delta t), \\ &= 2e^{(r-z)\delta t} + o(\delta t). \end{aligned} \tag{9.36}$$

At the next interval the equation for B is

$$B_{N-2} = 1 + \frac{a_{N-3}^2}{c_1 a_{N-2}^2} B_{N-1}, \tag{9.37}$$

which, after using the expression for B_{N-1}, (9.36), and following the same steps as for B_{N-1} reduces to

$$B_{N-2} = 3e^{2(r-z)\delta t} + o(\delta t). \tag{9.38}$$

Repeating gives the general result for B; this is

$$B_{N-m} = (m+1)e^{m(r-z)\delta t} + o(\delta t), \tag{9.39}$$

for $m = 1, \ldots, N-1$.

Finally, since $F_N = 0$ we can immediately write

$$F_{N-m} = \frac{1}{4} \sum_{i=N-m+1}^{N-1} \frac{\left(2e^{-(i-1)r\delta t} E_1[C_i^*] B_{i+1} - e^{(\mu - ir)\delta t} A_{i+1}\right)^2}{e^{(\mu-(i-1)r)\delta t} + e^{-2(i-1)r\delta t} B_{i+1}}, \tag{9.40}$$
$$m = 0, 1, \ldots, N-1,$$

but because of the presence of the $E_1[\cdot]$ operator and the quadratic nature of each term in this series a more compact form for the expression is not possible.

We now substitute equations (9.33) and (9.39) into (9.15); after some algebra we obtain

$$\delta H^*_{N-m} = -\frac{1}{m+1}e^{m(z-r)\delta t}H_{N-m} + \frac{m}{(m+1)S_{N-m}}e^{-((N-m)r-z)\delta t}E_1\Bigg[C^*_{N-m} - \frac{1}{m}e^{((m-1)z-mr-\mu)\delta t}\sum_{i=0}^{m}(e^{-(m-i)((2N-(m+i))r-\mu)\delta t}-1)C^*_{N-i}\Bigg] - \frac{4}{S_{N-m}(m+1)}(Nr-z)E_1[C^*_N]\delta t + o(\delta t). \quad (9.41)$$

Defining

$$R_{N-m} = \frac{m}{(m+1)S_{N-m}}e^{-((N-m)r-z)\delta t}E_1\Bigg[C^*_{N-m} - \frac{1}{m}e^{((m-1)z-mr-\mu)\delta t}\sum_{i=0}^{m}(e^{-(m-i)((2N-(m+i))r-\mu)\delta t}-1)C^*_{N-i}\Bigg] - \frac{4}{S_{N-m}(m+1)}(Nr-z)E_1[C^*_N]\delta t + o(\delta t), \quad (9.42)$$

the expression for the initial optimal traded quantity (at $m = N-1$) can now be written

$$\delta H^*_1 = -\frac{1}{N}e^{(N-1)(z-r)\delta t}H_1 + R_1. \quad (9.43)$$

Expanding (9.43) gives

$$\delta H^*_1 = -\frac{1}{N}H_1 - \frac{(N-1)}{N}(z-r)\delta t H_1 + R_1. \quad (9.44)$$

Applying the continuity relation, $H_2 = H_1 + \delta H^*_1$, allows us to write

$$\begin{aligned} H_2 &= \frac{N-1}{N}H_1 - \frac{(N-1)}{N}(z-r)\delta t H_1 + R_1, \\ &\approx \frac{N-1}{N}e^{-(z-r)\delta t}H_1 + R_1. \end{aligned} \quad (9.45)$$

At the second trading period when $m = N-2$ we have

$$\delta H^*_2 = -\frac{1}{N-1}e^{(N-2)(z-r)\delta t}H_2 + R_2. \quad (9.46)$$

After substituting (9.45) into (9.46) and simplifying we get

$$\delta H_2^* = -\frac{1}{N}e^{(N-3)(z-r)\delta t}H_1 - \frac{1}{N-1}e^{(N-2)(z-r)\delta t}R_1 + R_2. \tag{9.47}$$

Repeating this procedure gives the general result

$$\delta H_n^* = -\frac{1}{N}e^{(N-(2n-1))\left(\mu-r+\frac{1}{2}\sigma^2\right)\delta t}H_1 + R_n - \sum_{i=1}^{n-1}\frac{1}{N-i}e^{(N-2n+i+1)\left(\mu-r+\frac{1}{2}\sigma^2\right)\delta t}R_i, \tag{9.48}$$

where $n = 2, \ldots, N-1,$. With R_n defined in (9.42), writing $z = \mu + \frac{1}{2}\sigma^2$ explicitly and including the boundary terms, we now have

$$\left.\begin{aligned} \delta H_1^* &= -\frac{1}{N}e^{(N-1)\left(\mu-r+\frac{1}{2}\sigma^2\right)\delta t}H_1 + R_1, \\ \delta H_n^* &= -\frac{1}{N}e^{(N-(2n-1))\left(\mu-r+\frac{1}{2}\sigma^2\right)\delta t}H_1 + R_n - \sum_{i=1}^{n-1}\frac{1}{N-i}e^{(N-2n+i+1)\left(\mu-r+\frac{1}{2}\sigma^2\right)\delta t}R_i, \\ \delta H_N^* &= -H_N, \end{aligned}\right\} \tag{9.49}$$

which is our final result for the optimal trading strategy for large Λ in a perfectly liquid market.

Furthermore, substituting (9.33), (9.39), and (9.40) into (9.16) gives our final result for the optimal value function; this is

$$\begin{aligned} J_{N-m} = \frac{2\Lambda}{m+1}S_{N-m}H_{N-m}e^{-(N-1)r\delta t+m\left(\mu+\frac{1}{2}\sigma^2\right)\delta t}\sum_{i=N-m}^{N}e^{(N-m-i)\left(\mu+\sigma^2-r\right)\delta t}E_1[C_i^*] \\ + \frac{\Lambda}{m+1}e^{m\left(\mu-r+\frac{1}{2}\sigma^2\right)\delta t}S_{N-m}^2H_{N-m}^2 \\ - \frac{1}{4}\sum_{i=N-m+1}^{N-1}\frac{\left(2e^{-(i-1)r\delta t}E_1[C_i^*]B_{i+1} - e^{(\mu-ir)\delta t}A_{i+1}\right)^2}{e^{(\mu-(i-1)r)\delta t} + e^{-2(i-1)r\delta t}B_{i+1}}, \end{aligned} \tag{9.50}$$

where A_{N-i} and B_{N-i} are defined in (9.36) and (9.33), respectively, and we have made the $O(\delta t)$ approximation $c_1 \approx e^{(2\mu+\sigma^2)\delta t}$.

9.3 Discussion of Analytical Results

The first notable feature of the optimal strategy is the fact that the first term in (9.49) is exactly the n^{th} component of the naive strategy corrected by the variance-modified cost-of-carry effect, $\mu - r$. When $\mu - r + \frac{1}{2}\sigma^2 = 0$ this term is exactly the naive strategy and says that the transaction should be split into equal parcels and sold off constantly over its entire duration; when $\mu - r + \frac{1}{2}\sigma^2 > 0$, as we expect in reality, this linear trade profile becomes convex. In this case the transaction should be weighted more heavily to the beginning of the interval which is consistent with the act of minimizing its variance.

The second contribution to δH_n^* is

$$R_n - \sum_{i=0}^{n-1} \frac{1}{N-i} e^{(N-(2(n-1)-(i-1)))\left(\mu - r + \frac{1}{2}\sigma^2\right)\delta t} R_i. \tag{9.51}$$

This term is very similar to one that was found in the previous chapter. R_n is approximately the difference in initial expectations between the cost of the optimal strategy during the n^{th} period and the average cost of all future periods (except that in this case the present period's cost is weighted far more heavily) and so (9.51) can be interpreted as an indicator of how our initial expectations about the individual costs of the transaction have changed since the transaction began.

Making the substitution $n = N - m$, the definition of R_{N-m}, (9.42), can be rewritten

$$\begin{aligned} R_n = \frac{N-n}{(N-n+1)S_n} e^{-(nr-\left(\mu+\frac{1}{2}\sigma^2\right))\delta t} E_1 \Bigg[C_n^* \\ - \frac{1}{N-n} e^{((N-n-1)\left(\mu+\frac{1}{2}\sigma^2\right)-(N-n)r-\mu)\delta t} \sum_{i=n}^{N} (e^{-(i-n)((i+n)r-\mu)\delta t} - 1) C_i^* \Bigg] \\ - \frac{4}{S_n(N-n+1)} \left(Nr - \left(\mu + \frac{1}{2}\sigma^2 \right) \right) E_1[C_N^*]\delta t + o(\delta t). \end{aligned}$$

Examining this definition of R_n we see that it depends on the realized value of the asset price, S_n. As δH_n^* depends on all R_i for $i \leq n$ we find, as we did for the small-Λ case, that not only is the optimal strategy dynamic, but also path-dependent. The fact that we have assumed a perfectly liquid market and therefore, by definition, have neglected any possible trading-induced serial correlation in the asset price process is not inconsistent with this

finding. As we discussed at the end of Chapter 8, path-dependency in the optimal strategy results not only from serial correlations in the underlying price process, but also inherently from the quadratic nature of the objective function. It is this inherent effect that has caused the path-dependency in our large-Λ results.

9.4 Calculating the Solution of the RA Problem

In Chapter 8 we found that the optimal strategy, (8.50), in the small-Λ region contained a term dependent on the expected transaction cost during each period. Since the expected cost was in turn a function of the optimal trade quantity, this term caused the strategy to be implicit. While the implicit nature of the solution could have been a potential problem for generating a closed-form, explicit solution to the RA problem, we were able to exploit the fact that the implicit term was a small correction; after the solution was substituted into the correction we were able to truncate the expression leaving only the leading-order explicit term.

In the large-Λ region we again find a similar implicit term in the optimal strategy, (9.49). If we examine (9.49) closely, though, it is clear that this implicit term is part of the leading-order solution and cannot be considered small. The method used in the small-Λ region of substitution and truncation will not work in this case as each substitution will only succeed in generating more leading-order terms within the solution and this process will not terminate.

Instead of a closed-form solution we will seek a numerical solution to the RA problem. Because of the form of (9.51), solving for $\{\delta H_n^*\}$ is most easily achieved using an iterative method.

Let $k = 1, 2, \ldots$ be an iteration variable and $\delta H_n^{*(k)}$ be the k^{th} iteration of the n^{th} period optimal trade quantity. Beginning with an initial guess, $\{\delta H_n^{*(1)}\}$, the $k+1^{\text{th}}$ iteration of the n^{th} period component to the strategy is then given by the equation

$$\delta H_n^{*(k+1)} = -\frac{1}{N} e^{(N-(2n-1))\left(\mu - r + \frac{1}{2}\sigma^2\right)\delta t} H_1 + R_n^{(k)} - \sum_{i=1}^{n-1} \frac{1}{N-i} e^{(N-(2(n-1)-(i-1)))\left(\mu - r + \frac{1}{2}\sigma^2\right)\delta t} R_i^{(k)} \quad (9.52)$$

where

$$R_i^{(k)} = \frac{N-i}{(N-i+1)S_n} e^{\left(\mu - ir + \frac{1}{2}\sigma^2\right)\delta t} E_1\Bigg[C_i^{*(k)} - \frac{1}{N-i} e^{((N-i-1)\left(\mu+\frac{1}{2}\sigma^2\right) - (N-i)r - \mu)\delta t} \sum_{j=i}^{N} (e^{-(j-i)((j+i)r-\mu)\delta t} - 1) C_j^{*(k)}\Bigg] - \frac{4}{S_i(N-i+1)}\left(Nr - \left(\mu + \frac{1}{2}\sigma^2\right)\right) E_1[C_N^{*(k)}]\delta t, \quad (9.53)$$

and

$$C_i^{*(k)} = \delta H_i^{*(k)} S_i e^{-(i-1)r\delta t}. \quad (9.54)$$

Now the $\{\delta H_n^{*(k)}\}$ are known quantities and so we can easily calculate

$$\begin{aligned} E_1[C_i^{*(k)}] &= E_1\left[\delta H_i^{*(k)} S_i e^{-(i-1)r\delta t}\right], \\ &= \delta H_i^{*(k)} S_1 e^{(i-1)(\mu-r)\delta t}, \end{aligned} \quad (9.55)$$

and (9.53) can therefore be rewritten

$$R_i^{(k)} = \frac{(N-i)S_1}{(N-i+1)S_n} e^{\left(\mu - ir + \frac{1}{2}\sigma^2\right)\delta t} E_1\Bigg[\delta H_i^{*(k)} e^{(i-1)(\mu-r)\delta t} - \frac{1}{N-i} e^{((N-i-1)\left(\mu+\frac{1}{2}\sigma^2\right) - (N-i)r - \mu)\delta t} \sum_{j=i}^{N} (e^{-(j-i)((j+i)r-\mu)\delta t} - 1) e^{(j-1)(\mu-r)\delta t} \delta H_j^{*(k)}\Bigg] - \frac{4S_1}{S_i(N-i+1)}\left(Nr - \left(\mu + \frac{1}{2}\sigma^2\right)\right) \delta H_N^{*(k)}\delta t, \quad (9.56)$$

Using an initial guess, $\{\delta H_n^{*(1)}\}$, we then iterate the solution using (9.52) and (9.56) until we meet the convergence criterion

$$\sum_{n=1}^{N} \left|\delta H_n^{*(k+1)} - \delta H_n^{*(k)}\right| < \epsilon, \quad (9.57)$$

for some arbitrary constant ϵ.

9.5 Numerical Results and Further Discussion

We have calculated the solution to the large-Λ RA problem, (9.49), using the iterative algorithm, (9.52), with the convergence criterion, (9.57). In all cases the algorithm was run with $H_1 = 100000$, $N = 10$, and $\epsilon = 0.1$; this choice for ϵ was made as any smaller value did not increase the precision of the $\delta H_n^{'s}$ to the nearest unit of the asset. The calculations were performed using various initial guesses for the solution, including the naive strategy, the "sell everything in the first period" strategy, and the "sell everything in the final period" strategy; in all cases the solution was independent of this guess.

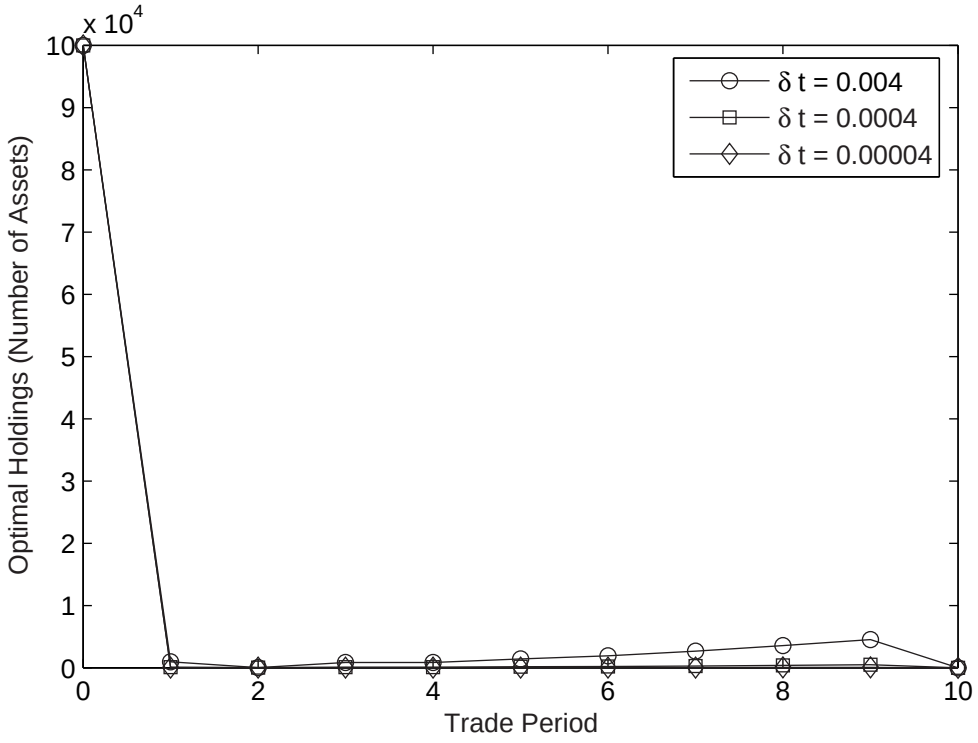


Figure 9.1: Optimal holdings for a trader with a large level of risk aversion in a perfectly liquid market. In this example $H_1 = 100000$, $N = 10$, $\mu = 0.08$, $r = 0.05$ and $\sigma = 0.3$. Plots are shown for $\delta t = 0.004$ ($\circ$), $\delta t = 0.0004$ ($\square$), and $\delta t = 0.00004$ ($\diamond$).

In Figure 9.1 we show the optimal holdings, $H_n^* = H_1 + \sum_{i=1}^{n-1} \delta H_i^*$, for several values of δt. By holding N constant and varying δt we are effectively examining the situation of how the optimal strategy varies as we execute the same number of trades over a varying horizon.

Specifically, for $\delta t = 0.004$ the curve represents a daily trade program over 2 weeks (10 trading days), for $\delta t = 0.0004$ the transaction approximately represents an hourly program over 1 day, and for $\delta t = 0.00004$ the transaction approximately represents 10 trades spanning 1 hour.

It is quite clear from these results that the strategy that minimizes the variance of the transaction cost in a perfectly liquid market is to liquidate the portfolio as quickly as possible. While this trade profile is only approximate in our results, we can see that $H_n^* \to 0$ for all $0 < n \leq N$ as $\delta t \to 0$ and these imperfections therefore appear to be a result of the $O(\delta t)$ approximations made in the derivation. There is a common pattern in these imperfections, however, in which H_n^* grows constantly after the second trade until the final period when any remaining contents of the portfolio are sold off. Why this growth pattern in the portfolio occurs is not exactly clear, but may be a result of the fact that the objective function for the large-Λ problem still contains the $E_1[C_n^*]$ terms of the auxiliary formulation. Since these terms represent the initial expected cost during the n^{th} period, in minimizing the objective function, it may be that the cost-of-carry effect of the cost-minimization problem enters the solution.

As we saw in Section 9.3, the analytic form of the auxiliary solution indicated that the optimal strategy is dynamic and path-dependent. Solutions of our numerical algorithm were also performed for three different asset price paths: These are

1. $S_n = e^{(n-1)\mu\delta t} S_1$, for all $n = 1, \ldots, N$ (the steadily increasing process),
2. $S_n = e^{-(n-1)\mu\delta t} S_1$, for all $n = 1, \ldots, N$ (the steadily decreasing process),
3. $S_n = e^{\frac{1}{2}((-1)^n+1)\mu\delta t} S_1$ for all $n = 1, \ldots, N$ (the oscillating process).

In all cases the effect on the solution was minimal and was on the order of the δt effect mentioned above which makes perfect sense given the definition of u and d. From a practical standpoint, therefore, even though the optimal strategy is technically dynamic, since all assets are sold in the first period when the asset price is predictable the strategy in reality is static.

On its own, our result in this chapter is not tremendously ground-breaking as simple intuition leads to the same result. By definition, variance is a non-negative quantity; the variance of the transaction's cost therefore cannot be negative and at minimum will be zero.

Furthermore, since the initial asset price is known when the transaction begins the variance of the cost of the first trade will be zero. If any number of assets are not sold during the first period their subsequent sale will contribute positive variance to the transaction as a whole; it is therefore obvious that the unique variance-minimizing strategy consists of a single sell order for the entire portfolio in the first period.

While the form of the variance-minimizing strategy in a perfectly liquid market may not have been a complete unknown, the work in this chapter has served three main purposes. First, we simply wanted to show that the auxiliary method works for this problem by generating the solution we expect from our intuition. Second, we wanted to demonstrate the numerical method necessary for generating the solution to the auxiliary when the $E_1[C_n^*]$ contributions are part of the leading-order solution.

Finally, as was mentioned at the beginning of the chapter, this work is intended as a first step toward finding a more general optimal solution for the large-Λ RA problem in a finitely liquid market. Given the small size of the market depth parameter, λ, it is hoped that the results of this chapter can be used as the leading-order component of an expansion solution in a finitely liquid market. Calculating this finite liquidity solution will probably involve two steps. As we have thus far only minimized the time risk of the transaction cost, the first step will be to reincorporate the expected cost aspect into the minimization problem exploiting the fact that it is small compared to the Λ contributions; that is, to assume Λ is still large, but not so large that the 1 within $\beta_n = 1 - 2\Lambda E_1[C_n^*]$ can be neglected. As this expected-cost contribution will introduce cost-of-carry effects, we should expect the optimal strategy found in this chapter to smooth out and the transaction to become more heavily weighted toward the end of the trading horizon since delaying the sale of an asset will increase the revenue generated from the transaction. The effect of this 'spreading', however, will be small as we will still be in the large-Λ region.

The second step to determining a solution of the finite-liquidity RA problem for large Λ will be to re-incorporate liquidity effects and assume $\lambda H_1 \ll 1$. Using the solution to the first step described in the last paragraph as the leading-order term it is hoped that the finite liquidity solution would then be expressible as an expansion in powers of λ and we could then solve for the $O(\lambda)$ correction. As we found in Chapter 7, minimizing liquidity costs in the absence of cost-of-carry effects would dictate that the transaction should be

spread evenly over its duration; in this large-Λ case this finite liquidity correction would therefore most likely tend to spread the trade out more, but the effect would be very small as $\lambda H_1 \ll 1 \ll \Lambda$.

Chapter 10

Conclusions and Areas for Future Study

In financial markets there exist costs due to trading an asset. In addition to the straightforward cost arising from the asset's bid-ask spread, there is a more indirect cost due to the feedback effect that a trader's trade will have on the asset price. Not only will a trade possibly need to tap multiple layers of the asset's order book thereby reducing (inflating) the price obtained (paid) per asset for the trade, but this trade might also result in a permanent shift in the asset price; if the position then needs to be reversed a net cost will result from the round-trip transaction. Both these market impact effects and the bid-ask spread effect can have serious consequences for anyone needing to execute a dynamic trading strategy in the asset. Two of the most important situations where dynamic trading strategies appear are in the pricing and hedging of derivatives and in the execution of large portfolio transactions. This thesis has therefore focussed on the problems of derivative pricing and the execution of portfolio transactions in the presence of liquidity effects.

The fundamental object from which all of our work has been developed is the Bakstein-Howison liquidity asset price model. This model is unique as it possesses effects due to both temporary and permanent market impact *and* the bid-ask spread. In addition to being easily calibratible, the BH model is attractive since it is relatively simple, possessing only three parameters in addition to those of the Black-Scholes model, and portable across different problems.

In Chapter 2 we derived the no-arbitrage price for a plain vanilla derivative on a finitely liquid asset. Using a hedging portfolio construction we found the trading strategy that minimized the risk of the portfolio was actually the delta-hedging strategy for general values of the liquidity parameters. For this general case, however, the risk was only minimized and not eliminated completely and expectations were required to produce a deterministic derivative price. Furthermore, it was shown that the PDE for the expected price in this general case had several forms depending on how the rehedging frequency scales with the bid-ask spread; as was mentioned in Chapter 2, examining the behaviour of the derivative price for each of these scalings is still an open area for research.

An interesting situation occurs when the bid-ask spread can be considered negligible. In this case the Delta hedging strategy creates a perfect hedge and the derivative price is found to obey a true continuous-time PDE similar in structure to the Black-Scholes equation, but with the addition of two nonlinear terms. One of these terms is proportional to Γ^3 and, as it contains the parameter α, results from the permanent shift in the asset price due to the hedging strategy. While the Γ^3 term can be positive or negative depending on the specific payoff structure of the contract, the other nonlinear term is proportional to Γ^2 and is therefore strictly positive. The fact that this second term is proportional to Γ^2 makes it behave as if it were the result of a transaction cost effect, this was shown to have an interesting consequence. Because it is strictly positive this transaction cost term always acts to increase hedging costs and, as a result, affects the buyer and seller of the contract in opposite ways; even though the bid-ask spread in the underlying was ignored, one has been generated in the price of the derivative by the $\gamma = 0$ equation through the asset's finite market depth.

The majority of Chapter 2 was just a summary of results from [4]; the original derivative pricing work in this thesis has focussed on two applications of and one extension to this standard BH liquidity pricing framework. In Chapter 3 we performed an asymptotic analysis of the ask price of a European call option in the $\gamma = 0$ model and found that there exists a boundary layer near expiry and centered around the strike of size $O(\bar{\lambda}^2)$ and $O(\bar{\lambda})$, respectively. In the outer region we showed that the leading-order component of the option's value is simply the Black-Scholes value with the dominant liquidity correction being due to the transaction cost (*i.e.* Γ^2) effect. Near strike and as expiry is approached,

the magnitude of the liquidity terms increases until they become equal in size to the linear Black-Scholes terms on the edge of the boundary layer. As we move into the boundary layer the liquidity terms dominate as the option's Γ becomes very localized around strike. For very small times to expiry we have shown that the Γ^3 term drives the leading-order behaviour of the option's value as the price slippage (or permanent market impact) effect dominates hedging costs. Furthermore, in this very restricted region of the boundary layer it was found that, to leading order, the option's Γ displayed compact support thus implying the contract needs no rehedging outside of a very narrow band around the strike.

As an application of the liquidity pricing framework, in Chapter 4 we have constructed and priced an American-type forward contract in the $\gamma = 0$ model focussing on the dependence of the liquidity premium and the BH induced spread on the liquidity parameters. It was found that this spread increased with decreasing market depth and increasing price slippage and that it was asymmetric; the liquidity premium being greater for the short position than for the long. Because the call and put forms of the contract provide guaranteed supply and demand, respectively, for the underlying at any time during its life it is thought that the American forwards could be effective hedging instruments against liquidity risks in an illiquid market. We have found that the price of these contracts is much smaller than the equivalent American option due to the possible downside risk to the holder of the forward. Furthermore, because of the very small Γ of the American forwards (even near the free boundary), the liquidity premia of the contracts is also very small; not only are the contracts an effective liquidity hedging tool, but they are also very inexpensive to the holder.

Even in a very shallow market with significant price slippage we have seen that the values of the short positions of the American forwards are quite small; while this result may not seem surprising, it is considering the very large loss that the writer could possibly incur. The holder of a liquidity hedge possesses the contract in order to provide a supply or demand of the asset when the liquidity of the market drops significantly which would most likely result from a large boom or crash in the asset price. In deriving the model for the asset price dynamics we assumed that (in the limit of continuous time) the diffusion component of the asset price process is generated by a Geometric Brownian Motion, but this construction does not reasonably allow for large jumps in the asset price. In valuing

the contract we have therefore calculated the liquidity premium for a hedge against the likelihood of the asset making a large change given Gaussian increments and it is therefore not surprising that the premium is small. As a next step it would be useful to investigate the valuation of the contract in an illiquid market where jumps are incorporated into the price process.

While transaction costs are a concern when pricing derivatives, also is the fact that there is the risk that these costs may change. In Chapters 5 and 6 we extended the $\gamma = 0$ constant-liquidity derivative pricing framework developed in Chapter 2 to account for liquidity risk (specifically market depth risk) in the underlying. The market depth process was modelled exogenously as some function of an Ornstein-Uhlenbeck process and the no-arbitrage price of a plain vanilla contingent on the asset was found to satisfy a highly nonlinear two spatial dimension PDE. To simplify this liquidity risk pricing framework we utilized the empirical fact that an asset's liquidity mean reverts with a period much shorter than the typical life of a derivative written on it. Using this fact we were able to adapt the multiscale asymptotic analysis developed by Fouque, Papanicolaou, and Sircar to reduce our nonlinear PDE to a series of nonhomogeneous Black-Scholes equations. In all cases it was found, supporting the results of Chapter 3, that the leading-order component of the solution is exactly the Black-Scholes value of the contract. Furthermore, the first correction is always the typical Γ^2 transaction cost effect found in Chapter 3, but with the long-run distributional average of the market depth process. The dominant correction resulting from the liquidity risk in the underlying was found to enter typically as the second correction and was driven by two effects; one transaction cost-like effect proportional to Γ^2 and a new effect dependent on $\partial\Gamma/\partial S$. Depending on the size of the market depth process relative to its rate of mean reversion, the size of the liquidity risk correction varies relative to the size of the first liquidity correction with the liquidity risk correction becoming less important as the market depth process fluctuates at greater frequencies which is understandable as it behaves more like its average in this case.

While the reduction of the nonlinear PDE into a series of linear nonhomogeneous PDEs greatly simplifies solving for the pricing equation, probably the most useful aspect of the asymptotic analysis is the fact that the liquidity parameters (including those due to the liquidity risk and the market price of market depth risk function) are reduced to only

three independent universal group parameters which are independent of the specific market depth model assumed and the contract being priced. Furthermore, we have extended the calibration procedure developed by Fouque, Papanicolaou, and Sircar to our stochastic liquidity example through the use of further regular perturbation methods. Defining the new concept of implied market depth we have developed a relatively simple method of calibrating the model using market options data. While the precise method for this calibration has been outlined in our work, it has not been carried out; determining the structure of the implied market depth surface as well as numerical estimates for the universal group parameters are areas for further study.

In Chapter 4 we examined the American forward as a possible hedging instrument against an asset's liquidity risk, but in that work the contract was priced using the standard BH model; as market depth is constant in the standard model our work in Chapter 4 effectively priced the hedging instrument in a market in which there is no liquidity risk. As a simple application of our stochastic liquidity work and to price the American forward in a more realistic setting we ended Chapter 6 by pricing the American forward in a market with stochastic liquidity. This work involved firstly extending the prior asymptotic analysis to account for the early exercise aspect of the American contract and it was found that the same series of nonhomogeneous Black-Scholes equations hold for the series expansion solution, but now only in the hold region of the contract with the appropriate free-boundary conditions. As expected, liquidity risk in the underlying increases the premium on the contract above that in the standard BH model, but interestingly the size of this premium has been found to be different for the call and put versions of the contract; this result has been explained in terms of the presence of the $\partial\Gamma/\partial S$ contribution in the liquidity risk correction. This term describes the skew of the contract's value which is mostly positive for the call-forward, but negative for the put-forward. An increase in the asset price will generally result in increased rehedging activity for the call-forward whereas the opposite will occur with the put-forward.

In Chapters 7, 8, and 9 we studied the problem of optimally liquidating a large portfolio of a single asset. The dynamics of the asset price have been modelled using the Bakstein-Howison model. Although the model possesses a bid-ask spread effect and a nonlinear form for both the temporary and permanent impact functions, we have used a linearized form of

the impact function and focused on the case where the bid-ask spread is negligible as this work has mainly been intended as a first step towards the general problem. It has been seen, however, that in most situations, extending these simplifications (as well as those involving buy orders and multiple assets) is a relatively straightforward task. We therefore feel that this formulation has the potential to form the basis of a practical model, although most likely in numerical form as the increase in algebraic complexity associated with these extensions is significant.

In Chapter 7 we examined how to liquidate a portfolio when cost-minimization is the only optimality criterion. Within this relatively simple framework we explicitly developed the stochastic optimal control problem and showed how to derive the equivalent Bellman equation from it. In order to compare our results with those in the literature we focused on the leading-order solution and found the optimal strategy consisted of two parts. First was the naive strategy component found by both Bertsimas and Lo [6] and Almgren and Chriss [2]. Not seen in either of these studies (as interest rates were neglected), however, was a second term due to a cost-of-carry effect which, in general, delays the sale of an asset to profit from its expected growth rate above that of a risk-free asset. With realistic values of the model parameters it was demonstrated that this second component could very easily dominate the first resulting in buy orders to begin the liquidation transaction.

In Chapters 8 and 9 we then examined portfolio liquidation to minimize a combination of the transaction's expected cost and its associated time risk. With the inclusion of a variance contribution into the program it was shown that the problem was unsolvable using the technique of dynamic programming. To overcome this difficulty it was necessary to neglect the covariance terms within the objective function and therefore focus on solving a 'risk-adjusted' optimal strategy as opposed to the exact mean-variance problem. To solve this risk-adjusted problem we then presented a method that involved embedding the objective function within a closely related, yet more general problem and then showed that the two problems are equivalent with the appropriate choice of the embedding parameter, β.

For general values of the trader's level of risk aversion, Λ, we showed that solving for the optimal strategy is enormously involved. We therefore split the range of Λ and solved the problem first for very small levels of the trader's risk aversion. In this small-Λ limit we

were able to find a closed-form explicit solution for the optimal strategy which consisted of two parts; a leading-order component that was precisely the optimal strategy of the cost-minimization problem and first-order corrections proportional to Λ. The effect of the correction terms acted to weight the strategy more heavily towards the beginning of the transaction as would be expected from a time-risk contribution. It was found that not only did this correction depend on the realized value of the asset price during each trading period, thus making the strategy dynamic, but also on all asset prices prior to that period thereby producing a path-dependent optimal strategy. We then showed with numerical calculations, however, that the path-dependent component of the solution is only a small correction to the size of the $O(\Lambda)$ correction itself and will therefore have little effect on the leading-order behaviour of the strategy. Due to a lack of serial correlations between asset prices in our simplified model, we reasoned that this path-dependency was a direct result of the quadratic nature of the objective function.

In chapter 9 we examined the large-Λ problem and found it to be much more involved than for small Λ; to obtain tractable results it was necessary to further simplify the problem and focus only on the perfect liquidity case. Even in this greatly simplified set-up we found that a closed-form, explicit optimal strategy was no longer possible. We developed an iterative algorithm to solve the implicit solution and found the strategy that minimizes the transaction cost's risk is simply that in which all assets in the portfolio are liquidated in the first period as is to be expected.

As we have mentioned several times already, the work presented here is only intended to be a first step toward solving for an optimal strategy in a more realistic setting. The next obvious steps are to include nonlinear terms of the impact function and solve for the optimal strategy when the risk-minimization effect is the same order as the cost-minimization effect. But the results of our work suggest some other extensions to the framework. Throughout our work we have treated the transaction horizon, T, as a constant; one such interesting extension would be to instead allow T to vary. Because of the form of the optimal strategy when Λ is large it is unlikely this extension have any effect in this case, but in other cases when the transaction is weighted more heavily near T it would be interesting to determine what the effect would be of a variable horizon.

As we saw in Chapter 7 it is very possible for the optimal liquidation strategy to be

non-monotonic (*i.e.* to consist of both individual buy and sell orders). In practice the trader may not be allowed to buy assets while liquidating the portfolio and we should explore the implications of a no-buy constraint on the optimal strategy. We have also neglected the bid-ask spread of the asset in our work and its effect should be included in the model. We note that if the bid-ask spread is included into the constrained optimization mentioned above then the trade program will be monotonic and the spread's effect will be trivial as it will contribute a fixed cost to each individual trade within the transaction as a whole. Finally, we have only considered a constant market impact function in our work; as the depth of a market's asset is typically time-dependent (and actually stochastic) it would be interesting to examine the problem of the optimal strategy with a time-dependent market impact function.

On a final note, in our work on both derivative pricing and portfolio transactions we have seen that the effect due to the finite market depth is often dominated by other effects; in the case of pricing derivatives, the market impact correction is *very* small compared to the Black-Scholes value for market values of the market depth parameter and when executing portfolio transactions the linearizing effect on the optimal strategy of the market depth effect can easily be dominated by the cost-of-carry effect. A very obvious next step for all topics we have examined in this thesis is to include the effects of the asset's bid-ask spread. From the market calibrated values of γ quoted in subsection 1.5.1 we can see that the bid-ask spread could have a much greater impact on transaction costs and its effect combined with the market depth effect should be investigated.

Another situation in which the liquidity effects will play a greater role is for an asset (such as a small cap stock) with a much less developed market than those for which the BH model has been calibrated for. With such an undeveloped market for the asset the layers within its order book may be very sparse and it is not immediately clear whether the BH model can accurately describe its impact function; investigating whether or not the BH model can accurately describe the price dynamics for such an illiquid asset is an interesting area for future study. If the model *can* be used in this case we of course expect larger calibrated values for both λ and γ and so the liquidity costs will impact more heavily.

Bibliography

[1] Viral V. Acharya and Lasse Heje Pedersen. Asset pricing with liquidity risk. Accepted for publication in *Journal of Financial Economics*, July 2004.

[2] R. Almgren and N. Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3(2):5–39, 2000.

[3] David Bakstein and Sam Howison. Using options on greeks as liquidity protection. University of Oxford, 2001.

[4] David Bakstein and Sam Howison. A risk-neutral parmetric liquidity model for derivatives. University of Oxford, 2002.

[5] G. Barles and H. M. Soner. Option pricing with transaction costs and a nonlinear black-scholes equation. *Finance and Stochastics*, 2:369–397, 1998.

[6] D. Bertsimas and A. W. Lo. Optimal control of execution costs. *Journal of Financial Markets*, 1:1–50, 1998.

[7] Tomas Björk. *Arbitrage Theory in Continuous Time*. Oxford University Press, 1998.

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

[9] Phelim P. Boyle and Ton Vorst. Option replication in discrete time with transaction costs. *The Journal of Finance*, 47(1), 1992.

[10] Umut Cetin, Robert A. Jarrow, and Philip Protter. Liquidity risk and arbitrage pricing theory. *Finance and Stochastics*, 8:1–31, 2004.

[11] Umut Cetin, Robert A. Jarrow, Philip Protter, and M. Warachka. Pricing options in an extended black scholes economy with illiquidity: Theory and empirical evidence. April 2005.

[12] Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam. Commonality in liquidity. *Journal of Financial Economics*, 56:3–28, 2000.

[13] Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam. Market liquidity and trading activity. *The Journal of Finance*, 56(2):501–530, April 2001.

[14] Tarun Chordia, Asani Sarkar, and Avanidhar Subrahmanyam. An empirical analysis of stock and bond market liquidity. *The Review of Financial Studies*, 18(1), 2005.

[15] J. Cox, S. Ross, and M. Rubenstein. Option pricing: A simplified approach. *Journal of Financial Economics*, 7:229–263, 1979.

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

[17] A. K. Dixit. *Optimization in Economic Theory.* Oxford University Press, second edition, 1990.

[18] Jean-Pierre Fouque, George Papanicolaou, and K. Ronnie Sircar. Asymptotics of a two-scale stochastic volatility model. In *Equations aux dérivées partielles et applications, in honour of Jacques-Louis Lions*, pages 517–525. Gauthier-Villars, 1998.

[19] Jean-Pierre Fouque, George Papanicolaou, and K. Ronnie Sircar. Fiancial modeling in a fast mean-reverting stochastic volatility environment. *Asia-Pacific Financial Markets*, 6(1):37–49, 1999.

[20] Jean-Pierre Fouque, George Papanicolaou, and K. Ronnie Sircar. Mean-reverting stochastic volatility. *International Journal of Theoretical and Applied Finance*, 3(1):101–142, 2000.

[21] Jean-Pierre Fouque, George Papanicolaou, and K. Ronnie Sircar. Stochastic volatility: Calibrating random volatility. *Risk*, pages 89–82, February 2000.

[22] Jean-Pierre Fouque, George Papanicolaou, and K. Ronnie Sircar. *Derivatives in Financial Markets with Stochastic Volatility*. Cambridge University Press, 2001.

[23] Jean-Pierre Fouque, George Papanicolaou, and K. Ronnie Sircar. From the implied volatility skew to a robust correction to black-scholes american option prices. *International Journal of Theoretical and Applied Finance*, 4(4):651–675, 2001.

[24] Rüdiger Frey. Perfect option replication for a large trader. *Finance and Stochastics*, 2:115–142, 1998.

[25] Rüdiger Frey and Alexander Stremme. Market volatility and feedback effects from dynamic hedging. *Mathematical Finance*, 7(4), October 1997.

[26] S. D. Hodges and A. Neuberger. Optimal replication of contingent claims under transaction costs. *Review of Futures Markets*, 8:222–239, 1989.

[27] T. Hoggard, A. E. Whalley, and P. Wilmott. Hedging option portfolios in the presence of transaction costs. *Advances in Futures and Options Research*, 7:21–35, 1994.

[28] Robert Jarrow and Andrew Rudd. *Option Pricing*. Dow Jones-Irwin, 1983.

[29] Robert A. Jarrow. Market manipulation, bubbles, corners, and short squeezes. *Journal of Financial and Quantitative Analysis*, 27(3):311–336, 1992.

[30] Robert A. Jarrow. Derivative security markets, market manipulation, and option pricing theory. *Journal of Financial and Quantitative Analysis*, 29(2):241–261, 1994.

[31] Robert A. Jarrow and Philip Protter. Liquidity risk and option pricing theory. February 2005.

[32] J. R. King and C. P. Please. Diffusion of dopant in crystalline silicon: An asymptotic analysis. *IMA Journal of Applied Mathematics*, 37:185–197, 1986.

[33] Hayne E. Leland. Option pricing and replication with transactions costs. *The Journal of Finance*, 40(5):1283–1301, 1985.

[34] D. Li and W. Ng. Optimal dynamic portfolio selection: Multiperiod mean-variance formulation. *Mathematical Finance*, 10(3):387–406, July 2000.

[35] R. C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3:373–413, 1971.

[36] John Ockendon, Sam Howison, Andrew Lacey, and Alexander Movchan. *Applied Partial Differential Equations*. Oxford University Press, 1999.

[37] Lubos Pastor and Robert F. Stambaugh. Liquidity risk and expected stock returns. *Journal of Political Economy*, 111(3):642–685, 2003.

[38] P. A. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics*, 51(3):239–246, August 1969.

[39] Philipp J. Schönbucher and Paul Wilmott. The feekback effect of hedging in illiquid markets. *SIAM Journal on Applied Mathematics*, 61:232–272, 2000.

[40] Myron Scholes. Liquidity options. *Risk*, Nov. 1999.

[41] K. Ronnie Sircar and George Papanicolaou. General black-scholes models accounting for increased market volatility from hedging strategies. *Applied Mathematical Finance*, 5:45–82, 1998.

[42] Nassim Taleb. *Dynamic Hedging: Managing Vanilla and Exotic Options*. John Wiley and Sons, Inc., 1997.

[43] A. E. Whalley. *Option Pricing with Transaction Costs*. PhD thesis, University of Oxford, 1998.

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