Euro. Jnl of Applied Mathematics (1999), *vol.* 10, *pp.* 117–139. Printed in the United Kingdom

Optimal hedging of options with small but arbitrary transaction cost structure

A. E. WHALLEY[1] and P. WILMOTT[2]

[1] *Warwick Business School, University of Warwick, Coventry, CV4 7AL UK*
[2] *Mathematical Institute, 24–29 St Giles, Oxford OX1 3LB UK and Department of Mathematics, Imperial College, London, UK*

(*Received 12 December 1996; in revised form 3 November 1998*)

In this paper we consider the problem of hedging options in the presence of cost in trading the underlying asset. This work is an asymptotic analysis of a stochastic control problem, as in Hodges & Neuberger [1] and Davis, Panas & Zariphopoulou [2]. We derive a simple expression for the 'hedging bandwidth' around the Black–Scholes delta; this is the region in which it is optimal not to rehedge. The effect of the costs on the value of the option, and on the width of this hedging band is of a significantly greater order of magnitude than the costs themselves. When costs are proportional to volume traded, rehedging should be done to the edge of this band; when there are fixed costs present, trading should be done to an optimal point in the interior of the no-transaction region.

1 Introduction

The seminal work in the theory of option pricing by Black and Scholes [3] obtained the value of an option on an asset, V, as the solution to a linear partial differential equation with independent variables the underlying asset price S and time t:

$$\mathscr{L}_{BS}(V) \equiv V_t + \tfrac{1}{2}\sigma^2 S^2 V_{SS} - rV + rSV_S = 0 \qquad (1.1)$$

subject to appropriate boundary and final conditions. In this equation, r is the interest rate, measuring the continuous compounding of interest on cash in the bank, and σ is the volatility of the asset, a measure of the standard deviation of the asset price returns. This model assumes, as we do here, that the underlying asset follows the random walk given by the stochastic differential equation

$$\mathrm{d}S = \mu S \mathrm{d}t + \sigma S \mathrm{d}X.$$

One of the simplest options is a European call option; this gives its holder the right but not the obligation to *buy* the asset at a given *exercise price*, E, at a given *maturity date*, T. For such an option, the final condition at $t = T$ is

$$V(S,T) = \max(S-E, 0)$$

for the buyer and

$$V(S,T) = -\max(S-E, 0) = \min(E-S, 0)$$

for the seller or *writer*. This has an explicit closed form solution for the buyer:

$$V(S,t) = S\Phi(d_1) - Ee^{-r(T-t)}\Phi(d_2), \tag{1.2}$$

where

$$d_1 = \frac{\ln(S/E) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{(T-t)}}$$

$$d_2 = \frac{\ln(S/E) + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{(T-t)}} = d_1 - \sigma\sqrt{(T-t)}$$

and $\Phi(d)$ is the cumulative density function for the standardized normal distribution.

The derivation of the Black–Scholes equation involves the construction of a weighted portfolio of the option and a number, $-\Delta$, of the underlying asset. Δ is selected so that at any instant the risk associated with the portfolio is minimized (under their assumptions the portfolio is actually risk-free), and as a result Δ is itself a function of S and t – in fact $\Delta = V_S(S,t)$, the derivative of the option price with respect to the asset price, and hence it changes continuously with time. This implies that the number of assets in the portfolio must be adjusted *continuously*, a process which is called *delta-hedging*. However transacting continuously in order to delta-hedge is not only unrealistic in itself, since in practice trades in the underlying asset can only occur at discrete points in time, but also gives rise to problems when we attempt to incorporate the costs of transacting in the underlying asset. If we maintain the strategy of trading continuously, the costs associated with this trading could be unbounded and swamp the option value. Hence a new valuation strategy is needed to incorporate costs and to provide for trading at discrete points in time. It turns out that even though the transaction costs are themselves small (and indeed we utilize this fact in this paper), their effect is of a greater order of magnitude. We find that the effect on the option price is of the order of the square root of the costs, and the effect on the band within which the option delta is allowed to move before a transaction is made is of the order of the quarter power of the costs. Hence whilst these effects are still small, they are significantly less small than the costs themselves.

There are two main approaches to the modelling of hedging strategies and option values in the literature: local in time and global in time. The former was started by Leland [4] and extended by Boyle & Vorst [5], Hoggard, Whalley & Wilmott [6], Avellaneda & Paras [7], Toft [8], Whalley & Wilmott [9] and Henrotte [10]. The first five of these assume hedging takes place at given discrete time intervals (Boyle & Vorst [5] is actually a binomial model) and the last two assume flexible but prescribed trading rules. These involve a band around the *ideal* value of Δ, within which the number of assets *actually* held in the portfolio is allowed to vary. In all of these the hedging strategy is, however, given exogenously, i.e. the hedging strategy is chosen *a priori*, as distinct from the current model where the strategy is determined as part of the model solution, and is chosen *optimally*; such prescribed hedging strategies are often used in practice. They typically result in partial differential equations for the value of the option, V hereafter, which are similar to the Black–Scholes equation (1.1) above, but with an extra term representing the effect of the transaction costs. An example of such an equation is

$$\mathscr{L}_{BS}(V) \equiv V_t + \tfrac{1}{2}\sigma^2 S^2 V_{SS} - rV + rSV_S = \mathscr{K}(S,t,V_{SS}), \tag{1.3}$$

https://doi.org/10.1017/S095679259900368X Published online by Cambridge University Press

where $\mathscr{K}$ is in general a nonlinear functional of the *Gamma*, $\Gamma = V_{SS}$, the second derivative of the option with respect to the asset price, and as a result these equations are nonlinear. However, since they only involve derivatives with respect to the asset price S and time t, they are invariably quick to compute.

The global-in-time models have been proposed by Hodges & Neuberger [1] and Davis, Panas & Zariphopoulou [2]. Such models achieve an element of 'optimality', since they are based on the approach of utility maximization. These models result in both an (endogenous) optimal hedging strategy, which turns out to involve a band around Δ as described above, and an option value. One of the disadvantages of such an approach is, however, speed of computation. The models are slow to compute since they usually result in three- or four-dimensional free boundary problems.

In this paper, we perform an asymptotic analysis of a utility maximizing model which is a generalization of the Davis, Panas & Zariphopoulou [2] model, but assuming that transaction costs are small. We derive simple formulae and equations for the optimal hedging strategy and option price respectively. The particular case of this model with proportional transaction costs only (described in Whalley & Wilmott [11]) was tested against others in the Monte Carlo simulations of Mohamed [12] and found to be the most successful strategy that was tested. In this paper we generalize this approach to consider an *arbitrary* transaction cost structure. We give, as examples, simple explicit results for the optimal 'hedging bandwidth' and optimal 'rebalance points' in the cases of fixed transaction costs and fixed plus proportional. The arbitrary cost structure case is more complex than the proportional cost case since we must now find not only the optimal hedging bandwidth (outside of which the number of assets held is not allowed to go), but also the optimal rebalance point. When buying or selling becomes optimal it is to this optmal rebalance point that we must rehedge.

In § 2 we very briefly describe the stochastic control problem of Davis, Panas & Zariphopoulou [2] with an arbitrary transaction cost structure. The interested reader is advised to read that paper carefully in conjunction with this. In § 3 we consider the asymptotic limit of small transaction costs; this results in an inhomogeneous diffusion equation for the price of an option and formulae for both the width of the hedging bandwidth and the optimal rebalance point. In § 4 we consider some special cases of particular, practical interest. In § 5 we discuss certain issues in the hedging of portfolios of options. § 6 concludes.

2 The utility maximizing model

In this section we describe the utility maximizing model derived by Hodges & Neuberger (HN) and Davis, Panas & Zariphopoulou (DPZ) and formulate it in a convenient form for our subsequent calculations. In this context, the 'utility' of wealth is used as a measure of the value of money to an individual. It is a popular economic concept, easily illustrated by the example of the pauper and the prince. Decreasing marginal (i.e. the first derivative of) utility, which is often assumed to apply in practice, implies the 'value' of one pound to the pauper is considerably higher than the value to the prince even though the purchasing power is the same. We use the notation $U(W)$ for the utility of an amount of wealth W.

Given an initial wealth (in cash) of $B_{wo}(t)$, the investor can invest in a portfolio of the risky share and riskless bond with the objective of maximising the expected utility of their wealth at a given future time T. This is equivalent to choosing, at any time τ between now and T, the quantity of shares $y(S, \tau)$ to be held in the portfolio. Later we shall consider portfolios including an option position: we use the sub or superscript wo to denote quantities 'without' the option obligation, and shall use the sub or superscript w for quantities 'with' the option obligation. J^{wo} is the maximized expected utility (of final wealth) without the option position:

$$J^{wo}(t, S, B_{wo}, y_{wo} = 0) = \max_{y_{wo}(\tau), t \leqslant \tau \leqslant T} \{E\left[U\left(W_{wo}(T)\right)\right]\}. \tag{2.1}$$

Here final wealth is defined to be wealth in cash after transaction costs

$$W_{wo}(T) = y_{wo}(S, T)S + B_{wo}(T) - k(S, y_{wo}(S, T)), \tag{2.2}$$

where $y_{wo}(S, t)$ represents the quantity of risky assets held at time t when the asset price is S, $B_{wo}(t)$ is the amount of risk-free bonds held at time t (which receive a return of r, the risk-free interest rate), and $k(S, u)$ is the transaction costs associated with transacting the quantity u of the risky asset S. It is assumed that transactions in risk-free bonds do not incur transaction costs; alternatively, k can be regarded as representing both sets of costs. In Equation (2.2) the cost is simply associated with liquidation of the final asset position.

Then define $\hat{B}_{wo}$ as the *minimum* initial wealth which delivers a non-negative maximum expected utility of final wealth:

$$\hat{B}_{wo} = \inf\{B : J^{wo}(t, S, B, 0) \geqslant 0\}.$$

For concave increasing utility functions with $U(0) = 0$, this implies an investor with zero initial wealth is indifferent beween paying (since $\hat{B}_{wo} \leqslant 0$) the amount $|\hat{B}_{wo}|$ and transacting as determined by the utility maximizing strategy, and doing nothing. It is the amount they are willing to pay to enter the market. More importantly, however, it serves as a reference for comparison once we introduce options.

If we consider the same situation: maximizing the expected utility of wealth at time T, but where we have in addition an option holding or liability which expires at T, and compare the minimum current wealth which gives non-negative expected utility of final wealth *including* the option position, say $\hat{B}_w$, with $\hat{B}_{wo}$, this will give a value for the option position. So the amount the investor is willing to pay to enter the market when they do not have the option position equals the amount they are willing to pay (or receive) to enter the market when they have an option position *plus* the value of the option position to them, $\hat{B}_{wo} = \hat{B}_w + V$, or

$$V = \hat{B}_{wo} - \hat{B}_w.$$

Furthermore, the difference between the portfolio strategies derived under the utility maximization with and without the option position will be the hedging strategy required to achieve the maximum expected utility of final wealth and hence the option valuation.

Now we perform the maximization in (2.1), with the terminal wealth (at time T) altered by a portfolio of options with payoff $V(S, T) = V_T(S)$ at maturity[1] T. We assume that

[1] Portfolios of options with a number of maturities are dealt with in a similar manner solving backwards in time from the furthest maturity date and adding in payoffs for earlier maturities as appropriate.

the option position will be held until expiry, the position will not be closed before then. Final wealth net of transaction costs is then given by

$$W_w(T) = y_w(S,T)S + B_w(T) - k(S, y_w(S,T)) + V_T(S). \tag{2.3}$$

For the writer of a European call option, for example, we have

$$W_w(T) = \begin{cases} y_w(S,T)S + B_w(T) - k(S, y_w(S,T)), & S \leqslant E \\ y_w(S,T)S + B_w(T) - k(S, y_w(S,T)) + E - S, & S > E \end{cases} \tag{2.4}$$

Note we are assuming here that the option is settled in cash. For options with delivery of the asset on exercise the analysis below remains the same; the final conditions merely alter, but this only affects the option price at a higher order of magnitude than we shall consider here.

Letting J^w be the maximized expected utility including the option position,

$$J^w(t, S, B_w, y_w = 0) = \max_{y_w(\tau), t\leqslant\tau\leqslant T} \{E\,[U\,(W_w(T))]\}, \tag{2.5}$$

$\hat{B}_w$ is then defined by

$$\hat{B}_w = \inf\{B : J^w(t, S, B, 0) \geqslant 0\};$$

it is the minimum initial wealth which delivers non-negative maximum expected utility of final wealth with the option obligation. It turns out that the problems for J^{wo} and J^w are identical except for the conditions at maturity. For the general analysis we shall therefore drop sub- and superscripts of *wo* and *w* for the two problems and consider the problems separately only when necessary.

For a general utility function $U(x)$ and for the particular case of proportional transaction costs, $k(S, y) = k_3 S|y|$, HN and DPZ show that this problem is described by

$$\max \Bigg\{ J_y - S(1+k_3)J_B, \tag{2.6}$$

$$-\left(J_y - S(1-k_3)J_B\right), \tag{2.7}$$

$$J_t + rBJ_B + \mu S J_S + \frac{\sigma^2 S^2}{2} J_{SS} \Bigg\} = 0, \tag{2.8}$$

where B is the amount of risk-free bonds held in the portfolio, and

$$\mathrm{d}B = rB\mathrm{d}t - S\mathrm{d}y - k(S, \mathrm{d}y)$$

(J_y, J_B, J_t and J_S, J_{SS} represent derivatives of J with respect to y, B, t and S respectively).

One of the most popular choices for utility function is the negative exponential utility function:[2]

$$U(x) = 1 - \exp(-\gamma x). \tag{2.9}$$

For the present problem with this choice of utility function, HN and DPZ show that the monetary amount invested in the risky asset is independent of total wealth, and hence

[2] This utility function is popular for its simplicity and its property of constant 'absolute risk aversion'. Absolute risk aversion is given by U''/U' and is a measure of the reluctance of an investor to accept a 'fair bet'.

https://doi.org/10.1017/S095679259900368X Published online by Cambridge University Press

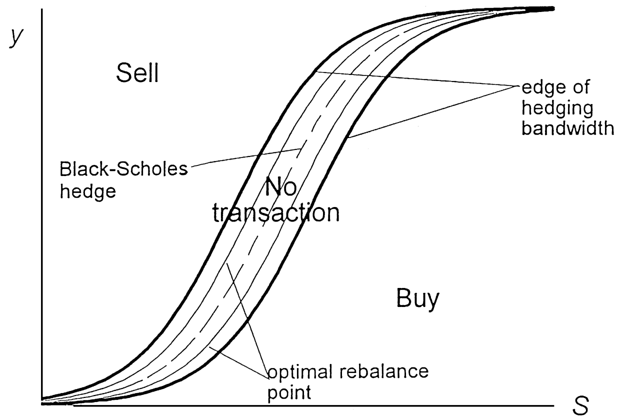

FIGURE 1. A schematic diagram of (S, y) space showing the buy, sell and no-transaction regions, and the optimal rebalance lines.

that the dependence on B can be eliminated and the problem can be reduced to a three-dimensional free boundary problem. The structure of the solution then takes the form (2.11) – see below. This reduction of a dimension is not possible for an arbitrary utility function, and so in this work we concentrate on this special though important case. In the case of exponential utility, for the *writer* of a European call option facing proportional transaction costs, HN and DPZ show that (S, y) space divides up into just three regions, within each of which one of the terms (2.6), (2.7) and (2.8) equals zero; these are shown schematically in Figure 1 (ignore, for the moment, the dashed lines in this figure). In regions where (2.6) equals zero, the strategy is to buy underlying assets (shares); where (2.7) vanishes, shares should be sold according to the strategy; and regions where (2.8) vanishes are regions where no transactions occur. The number of shares to be bought or sold will be found as part of the solution.

The middle line in Figure 1 is the curve along which the investor must move *in the absence of transaction costs*; this curve is denoted by

$$y = y^*(S, t).$$

The writer of the option must always maintain her portfolio in the region of the (S, y) space bounded by the two outer curves. While inside this region she does not transact. So when (2.8) holds, y does not change.

The boundaries where the buy and no-transaction (sell and no-transaction respectively) regions meet, which mark the edges of the no-transaction region, are of particular interest since they represent the points in (S, y) space at which rebalancing occurs. Should a movement of the asset price take the writer to the edge of this no-transaction region, she must trade so as to stay inside. If she hits the top boundary, $y^+(S, t)$, she must sell shares; if she hits the bottom boundary, $y^-(S, t)$, she must buy shares. The number of shares bought or sold is chosen to move the position to the rebalancing position. The points to which rebalancing occurs must also be determined as part of the solution; these will be inside or on the boundaries of the no-transaction region.

The portfolio adjustment strategy in the presence of transaction costs and the hedging strategy for an option in the presence of transaction costs is derived from these two sets of points.

In the proportional cost case studied by HN and DPZ, the optimal rebalance points are just the boundaries of the no-transaction region; on hitting either of the boundaries, trading occurs so as to move along that boundary. This feature of transacting the minimum possible in order to stay within the no-transaction region is due to the particular, proportional nature of the transaction costs. In the case of arbitrary transaction costs it is perfectly possible that the (S, y) space could be divided into more than three regions; there could be more than one buy region, say. However, for typical transaction cost structures this is unlikely to occur. Although our analysis will be very general, we shall have at the back of our minds a costs structure of the form fixed plus proportional. In this case, and because of the fixed cost term in particular, we find that the (S, y) space is divided up as shown in Figure 1. There are still the buy, sell and no-transaction regions but now on rehedging the portfolio we must buy or sell shares to move to the 'optimal rebalance point', shown as the dashed lines in this figure. The positions of the hedging bandwidth and the optimal rebalance points must all be found.

For the holder of a European call option or indeed a portfolio of held and written options the diagram can be transformed in the obvious way; the structure remains the same. Asymptotic expressions for y^* and the position of the upper and lower boundaries for the proportional cost case described above were found by Whalley & Wilmott [11]. In this paper we consider a more general form of transaction costs incorporating a fixed component in addition to costs proportional to the value (or number of assets) traded and show how different types of costs have differing effects on option valuation and, more strikingly, on the hedging strategy followed.

Our objectives are to solve for the value of the option, V, and, equally if not more importantly from the point of view of the practicalities of trading, find the edges between the no-transaction and buy and sell regions and the optimal rebalance points. In full generality this involves the solution of a four-dimensional free boundary problem (for J), from which the option value and trading strategy are derived by the process described above, the latter as the free boundaries between the no-transaction and buy and sell regions. This is a computationally expensive and time-consuming exercise.

We also use the negative exponential utility function

$$U(x) = 1 - \exp(-\gamma x), \tag{2.10}$$

and show how the problem can be rewritten in an intuitively appealing form which gives the option value as the difference between the values of the wealth invested in the risky asset with and without the option position. Like HN and DPZ, we find that the particular choice of the negative exponential utility function results in the reduction of the problem to a single partial differential equation valid in the no-transaction region, with explicit formulae for the solutions in the buy and sell regions, respectively. The reduction of dimension which occurs in this special case does not happen with a general utility function. We shall leave the analysis of the arbitrary case for the future; we are sure that a similar analysis to that in the present paper is possible.

We rewrite J, the maximum expected utility of final wealth, in the form

$$J(t,S,B,y) = 1 - \exp\left(-\frac{\gamma}{\delta(t,T)}(B + W(t,S,y))\right). \tag{2.11}$$

Here γ is the measure of absolute risk aversion in the utility of final wealth $U(x)$,

$$\gamma = -\frac{U''(x)}{U'(x)},$$

and is constant, and $\delta(t,T) = e^{-r(T-t)}$ is a discount factor which converts wealth at maturity of the option to current wealth. $B = B(t)$ is the current amount held in risk-free bonds, and $W(t,S,y)$ is the expected value at time t of the utility-maximized wealth held in the risky asset at time T. The system of equations for J, (2.6), (2.7) and (2.8) transform into the following equations satisfied by W:

$$\min\left\{ W_y - \frac{\gamma}{\delta(t,T)}\left(S + \frac{\partial k}{\partial y}\right), \right. \tag{2.12}$$

$$-\left(W_y - \frac{\gamma}{\delta(t,T)}\left(S - \frac{\partial k}{\partial y}\right)\right), \tag{2.13}$$

$$\left. W_t - rW + \mu S W_S + \frac{\sigma^2 S^2}{2}\left(W_{SS} - \frac{\gamma}{\delta(t,T)}(W_S)^2\right) \right\} = 0. \tag{2.14}$$

(2.12) and (2.13) can be solved to give the system as

$$W_t - rW + \mu S W_S + \frac{\sigma^2 S^2}{2}\left(W_{SS} - \frac{\gamma}{\delta(t,T)}(W_S)^2\right) \geqslant 0. \tag{2.15}$$

subject to the constraints

$$W(t,S,y+u) \geqslant W(t,S,y) + \frac{\gamma}{\delta(t,T)}(uS + k(S,u)), \tag{2.16}$$

$$W(t,S,y-u) \geqslant W(t,S,y) - \frac{\gamma}{\delta(t,T)}(uS - k(S,u)) \tag{2.17}$$

where $u > 0$, or more succinctly ((2.15) is unaffected)

$$W(t,S,y+u) \geqslant W(t,S,y) + \frac{\gamma}{\delta(t,T)}(uS + (k_1 + (k_2 + k_3 S)|u|)). \tag{2.18}$$

These must be solved subject to 'value-matching' and 'smooth-pasting' conditions at the boundaries between (2.16), (2.17) and (2.15), which have still to be found, and final conditions at T: $W(T,S,y(S,T))$ represents the wealth invested in the risky asset at T, including any positions held for the hedging of options, net of transaction costs incurred in converting the risky assets into risk-free bonds. For the portfolio without the option position this is simply

$$W_{wo}(T) = y_{wo}(S,T)S - k(S, y_{wo}(S,T)), \tag{2.19}$$

and for the portfolio including, say, a written European call option position (with cash delivery),

$$\begin{aligned} W_w(T) &= y_w(S,T)S - k(S, y_w(S,T)) + \min(E - S, 0) \\ &= \begin{cases} y_w(S,T)S - k(S, y_w(S,T)), & S \leqslant E \\ y_w(S,T)S - k(S, y_w(S,T)) + E - S, & S > E \end{cases} \end{aligned} \tag{2.20}$$

https://doi.org/10.1017/S095679259900368X Published online by Cambridge University Press

Recall that the option value V is defined to be the difference between the minimum initial cash amounts which delivered non-negative maximum expected final utility at T for portfolios with and without the option position, respectively,

$$V = \hat{B}_{wo} - \hat{B}_w.$$

Using the negative exponential utility function formulation (2.11), we see that (since $U(x)$ is a monotone increasing function), zero maximum expected utility of final wealth implies

$$\hat{B}_w = -W_w(t, S, 0), \quad \hat{B}_{wo} = -W_{wo}(t, S, 0),$$

where W_w (W_{wo} respectively) is the current value of the expected maximized utility of the holdings of the risky asset at T including (excluding respectively) the effects of the option liability. Evaluation at $y = 0$ represents the fact that these values are *net* of all transaction costs. This means W_w (W_{wo}, respectively) is the solution to (2.16), (2.17) and (2.15) with conditions (2.20) and (2.19), respectively.

So the option value is given by

$$V = W_w(t, S, 0) - W_{wo}(t, S, 0). \tag{2.21}$$

The hedging strategy for the option is given by $y_w(S, t) - y_{wo}(S, t)$, the difference between the (endogenously determined) values of the amount of the underlying asset held in each of the two portfolios y_w and y_{wo} (with and without the option liability) respectively.

3 Asymptotic analysis for small levels of transaction costs

In this problem the transaction costs associated with trading in the underlying asset are, in practice, small. We shall therefore find it convenient to introduce the parameter ϵ as a measure of the size of the transaction costs. Thus at each rehedge, there will be an associated cost K which is of $O(\epsilon)$. We shall then take asymptotic expansions of the W functions in powers of ϵ. From these expansions we find that the hedging bandwidth, i.e. the width of the band about the number of assets which would be held in the absence of transaction costs, is $O(\epsilon^{\frac{1}{4}})$, and that the effect of the costs on the option value is $O(\epsilon^{\frac{1}{2}})$. We shall see later in this section how the expansion in fractional powers of ϵ arises naturally from the structure of the problem.

We therefore translate the y coordinate according to

$$y = y^*(S, t) + \epsilon^{\frac{1}{4}} Y. \tag{3.1}$$

The rescaled variable Y is a dimensionless quantity which will be $O(1)$. It is a measure of the difference between the number of shares *actually* held in the portfolio and the *ideal* number we would hold in the absence of transaction costs, y^*. This rescaling ensures that our analysis is concentrated in the area of interest, the $O(\epsilon^{\frac{1}{4}})$ region around the value of the hedge ratio in the absence of costs. Our independent variables are now S, t and Y. (We shall find an explicit expression for y^* as a function of S and t.) We also rescale the cost function by writing

$$k(S, \epsilon^{\frac{1}{4}} U) = \epsilon K(S, U).$$

This scaling acknowledges the fact that the hedging bandwidth is $O(\epsilon^{\frac{1}{4}})$ and the cost of a trade is $O(\epsilon)$.

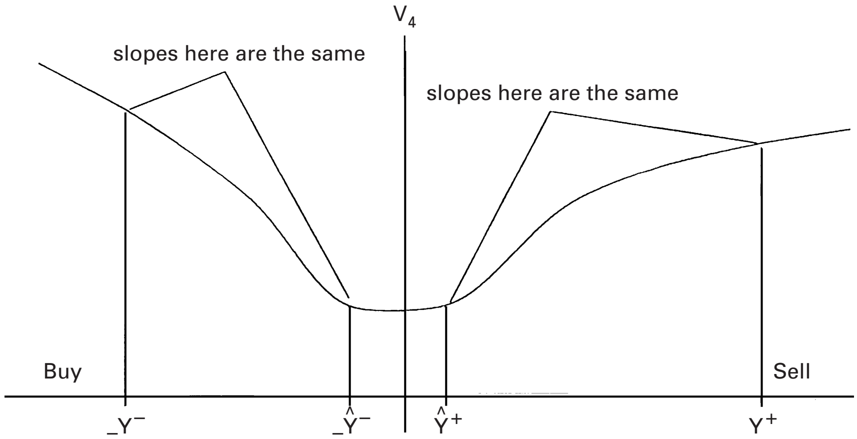

FIGURE 2. A schematic diagram of V_4 against Y.

Fractional asymptotic scalings have occured in a number of other transaction or adjustment cost problems. The scaling of $\epsilon^{\frac{1}{4}}$ was found and exploited in a similar manner in the portfolio analysis problem of Atkinson & Wilmott [13]. Shreve [14] has results which suggest an asymptotic power of one third for the width of the no-transaction interval for an optimal investment and consumption model with transaction costs under a different utility function, and notes that Fleming *et al.* [15] have also obtained this scale. Dixit [16] also obtains asymptotic scales of one quarter for the width of a no-transaction region for fixed costs of adjustment but with quadratic penalty costs of deviation from an ideal value of the state variable, and an asymptotic scale of one third when the penalty costs are instead equal to the modulus of the deviation from the ideal. The exact fractional power for a no-transaction region in the presence of adjustment costs varies therefore, depending on the form of the adjustment costs and other aspects of the particular problem, but in all these cases first order costs produce fractional order (i.e. larger) widths of the bands within which no adjustment occurs.

Having rescaled our variables we can refer to Figure 2, which is a more detailed schematic diagram of the buy, sell and no-transaction regions. The horizontal axis in this figure is the Y-axis. We shall explain the vertical axis in a moment. This figure shows the boundaries of the no-transaction region as $Y = -Y^{-}(S,t)$ and $Y = Y^{+}(S,t)$. Inside this region are the optimal rebalance points $Y = -\hat{Y}^{-}(S,t)$ and $Y = \hat{Y}^{+}(S,t)$: when the asset price random walk takes us to $Y^{+}(S,t)$ we must sell shares to move to $\hat{Y}^{+}(S,t)$. It is worth mentioning that if there is no fixed element to the transaction costs then the optimal rebalance point and the edge of the no-transaction region coincide. That is $Y^{+} = \hat{Y}^{+}$. In the case including fixed transaction costs, $\hat{Y}^{-}$ and $\hat{Y}^{+}$ are strictly inside the no-transaction region, and so there is a jump in the number of shares held.

If we are totally within the buy region, say, then we must transact to the optimal rebalance point. Thus, in the buy region $Y < -Y^{-}(S,t)$

$$W(S,Y,t) = W(S,-\hat{Y}^{-},t) + \epsilon^{\frac{1}{4}}(Y + \hat{Y}^{-})S - \epsilon K(S, Y + \hat{Y}^{-}). \tag{3.2}$$

Similarly, in the sell region $Y > Y^{+}(S,t)$ we have

$$W(S,Y,t) = W(S,\hat{Y}^{+},t) + \epsilon^{\frac{1}{4}}(Y - \hat{Y}^{+})S - \epsilon K(S, Y - \hat{Y}^{+}). \tag{3.3}$$

In the no-transaction cost region we have equality in equation (2.15), i.e.

$$W_t - rW + \mu S W_S + \frac{\sigma^2 S^2}{2}\left(W_{SS} - \frac{\gamma}{\delta(t,T)}(W_S)^2\right) = 0. \tag{3.4}$$

We impose value-matching and smooth pasting conditions, which we discuss in more detail in § 3.4. These conditions correspond to continuity of the option price and optimality of the hedging strategy. When there are fixed costs, these are

$$W(S, Y^+, t) = W(S, \hat{Y}^+, t) + \epsilon^{\frac{1}{4}}(Y^+ - \hat{Y}^+)S - \epsilon K(S, Y^+ - \hat{Y}^+), \tag{3.5}$$

$$\left.\frac{\partial W}{\partial Y}\right|_{(S,Y^+,t)} = \epsilon^{\frac{1}{4}} S - \epsilon \left.\frac{\partial K}{\partial U}\right|_{(S,U=Y^+-\hat{Y}^+)}, \tag{3.6}$$

and

$$\left.\frac{\partial W}{\partial \hat{Y}^+}\right|_{(S,\hat{Y}^+,t)} = \epsilon^{\frac{1}{4}} S - \epsilon \left.\frac{\partial K}{\partial U}\right|_{(S,U=Y^+-\hat{Y}^+)} \tag{3.7}$$

(similar conditions hold at the buy boundary). In the absence of fixed costs, $Y^+ = \hat{Y}^+$ and the boundary conditions become

$$\left.\frac{\partial W}{\partial Y}\right|_{(S,Y^+,t)} = \epsilon^{\frac{1}{4}} S - \epsilon \left.\frac{\partial K}{\partial U}\right|_{(S,U=0)}, \tag{3.8}$$

and

$$\left.\frac{\partial^2 W}{\partial Y^2}\right|_{(S,Y^+,t)} = 0. \tag{3.9}$$

We can write the solution in the no-transaction region as

$$\begin{aligned} W = (y^* + \epsilon^{\frac{1}{4}} Y)S + V_0(S,t) + \epsilon^{\frac{1}{4}} V_1(S,t) + \epsilon^{\frac{1}{2}} V_2(S,t) \\ + \epsilon^{\frac{3}{4}} V_3(S,t) + \epsilon V_4(S,Y,t) + \epsilon^{\frac{5}{4}} V_5(S,Y,t) + \cdots . \end{aligned} \tag{3.10}$$

We shall justify the omission of Y from the V_0, V_1, V_2 and V_3 terms in this expansion later. We must now find the functions V_i such that equation (3.4) and all relevant boundary and smoothness conditions are satisfied. We shall see, in performing this analysis, that the choice of a series expansion in powers of $\epsilon^{1/4}$ is inevitable.

The $\epsilon^{\frac{1}{4}}$ term in (3.10) proportional to Y is forced on us as it represents the cash (without costs) from a sale of $O(\epsilon^{\frac{1}{4}})$ shares on rehedging. We shall find that the $O(\epsilon^{\frac{1}{4}})$ term V_1 is identically zero, but we leave it in for the sake of completeness. We also find that the leading order term which depends explicitly on Y is the $O(\epsilon)$ term, $V_4(S, Y, t)$. The dependence of V_4 on Y is required by the continuity conditions at the edges of the no-transaction regions to match the transaction cost terms in (3.2) and (3.3). So V_4 represents the cost of a single transaction to stay within the hedging band.

Now observe that since the width of the no-transaction region is $O(\epsilon^{\frac{1}{4}})$ the expected time between rehedges is $O(\epsilon^{\frac{1}{2}})$. Therefore during the lifetime of the option there will be $O(\epsilon^{-1/2})$ rehedges. This gives a *total* cost – individual cost multiplied by number of hedges – of $O(\epsilon^{\frac{1}{2}})$. Thus, the $O(\epsilon^{\frac{1}{2}})$ term above, the V_2 term, will contain the leading correction to the Black–Scholes equation; this is what we are interested in. So the V_4 term represents the cost of the individual trade and V_2 represents the accumulated transaction costs of all trades. As we shall see, these two functions are intimately related. We can now say that the vertical axis in Figure 2 represents the Y-dependence of the unknown function V_4.

https://doi.org/10.1017/S095679259900368X Published online by Cambridge University Press

Since the derivatives in (3.4) are with respect to t and S *keeping y fixed*, then

$$\frac{\partial}{\partial y} \to \epsilon^{-\frac{1}{4}} \frac{\partial}{\partial Y},$$

$$\frac{\partial}{\partial S} \to \frac{\partial}{\partial S} - \epsilon^{-\frac{1}{4}} y_S^* \frac{\partial}{\partial Y},$$

$$\frac{\partial}{\partial t} \to \frac{\partial}{\partial t} - \epsilon^{-\frac{1}{4}} y_t^* \frac{\partial}{\partial Y}.$$

Thus we readily find from (3.10) and (3.1) that

$$\begin{aligned}
\frac{\partial W}{\partial t} \equiv W_t &= y_t^* S + V_{0_t}(S,t) + \epsilon^{\frac{1}{4}} V_{1_t}(S,t) + \epsilon^{\frac{1}{2}} V_{2_t}(S,t) + \epsilon^{\frac{3}{4}} V_{3_t}(S,t) \\
&\quad + \epsilon V_{4_t}(S,Y,t) - y_t^* \left(S - \epsilon^{\frac{3}{4}} V_{4_Y} + \cdots\right) + \cdots, \\
\frac{\partial W}{\partial S} \equiv W_S &= y_S^* S + y^* + \epsilon^{\frac{1}{4}} Y + V_{0_S}(S,t) + \epsilon^{\frac{1}{4}} V_{1_S}(S,t) + \epsilon^{\frac{1}{2}} V_{2_S}(S,t) \\
&\quad + \epsilon^{\frac{3}{4}} V_{3_S}(S,t) + \epsilon V_{4_S}(S,Y,t) - y_S^* \left(S + \epsilon^{\frac{3}{4}} V_{4_Y} + \cdots\right) + \cdots, \\
\frac{\partial^2 W}{\partial S^2} \equiv W_{SS} &= y_S^* + V_{0_{SS}}(S,t) + \epsilon^{\frac{1}{4}} V_{1_{SS}}(S,t) + \epsilon^{\frac{1}{2}} V_{2_{SS}}(S,t) + \epsilon^{\frac{3}{4}} V_{3_{SS}}(S,t) \\
&\quad - y_{SS}^* \epsilon^{\frac{3}{4}} V_{4_Y} - y_S^* \epsilon^{\frac{3}{4}} V_{4_{YS}} \\
&\quad - y_S^* \left(1 + \epsilon^{\frac{3}{4}} V_{4_Y} - \epsilon^{\frac{1}{2}} y_S^* V_{4_{YY}} - \epsilon^{\frac{3}{4}} y_S^* V_{5_{YY}} + \cdots\right) + \cdots.
\end{aligned}$$

It will be observed that each of the above can be slightly simplified. We have retained them in this form to help the reader perform their own calculations.

The advantage of asymptotic analysis will now become clear when we perform the next step, to substitute these expressions into (3.4) and equate powers of $\epsilon^{\frac{1}{4}}$.

3.1 The $O(\epsilon^{-\frac{1}{2}})$ and $O(\epsilon^{-\frac{1}{4}})$ equations

The $O(\epsilon^{-\frac{1}{2}})$ terms lead to an ordinary differential equation for V_0 with independent variable Y

$$V_{0_{YY}} = \frac{\gamma}{\delta(t,T)} V_{0_Y}^2,$$

This is solved with boundary conditions the leading order terms in the value-matching and smooth pasting conditions which imply $V_{0_Y} = 0$. Similar ordinary differential equations together with the relevant terms in the value-matching and smooth pasting conditions are found for V_1, V_2 and V_3 (we omit the details). This implies $V_{0_Y} = V_{1_Y} = V_{2_Y} = V_{3_Y} = 0$ for all Y, and we have represented this explicitly in (3.10).

3.2 The $O(1)$ equation

The $O(1)$ terms are

$$W_t = V_{0_t}$$
$$W_S = V_{0_S} + y^*$$
$$W_{SS} = V_{0_{SS}},$$

https://doi.org/10.1017/S095679259900368X Published online by Cambridge University Press

and so the $O(1)$ equation is

$$V_{0_t} - rSy^* - rV_0 + \mu S\left(V_{0_S} + y^*\right) + \frac{\sigma^2 S^2}{2}\left(V_{0_{SS}} - \frac{\gamma}{\delta(t,T)}\left(V_{0_S} + y^*\right)^2\right) = 0. \quad (3.11)$$

This is the equation in the absence of any costs and we will see later how it results in the Black–Scholes equation for the option value to leading order.

3.3 The $O(\epsilon^{\frac{1}{4}})$ equation

We can take this procedure to the next order, equating powers of $\epsilon^{\frac{1}{4}}$. We find that

$$V_{1_t} - rSY - rV_1 + \mu S\left(Y + V_{1_S}\right) + \frac{\sigma^2 S^2}{2}\left(-\frac{2\gamma}{\delta(t,T)}\left(Y + V_{1_S}\right)\left(V_{0_S} + y^*\right) + V_{1_{SS}}\right) = 0.$$

This equation contains a term proportional to Y and one independent of Y. Since all the other terms in the equation are independent of Y, these terms must separately be zero. From the first of these we find that

$$y^*(S,t) = -V_{0_S} + \frac{\delta(t,T)(\mu - r)}{\gamma S \sigma^2}, \quad (3.12)$$

as given by Davis *et al.* Thus, if we can find V_0 (for each of the problems with and without the option liability) then we have found the leading order expressions for y^* in each case, and hence the difference, which gives the hedging strategy.

Equation (3.12) determines the hedging strategy in the absence of transaction costs, y^*, in terms of the leading order 'option value' V_0. If we substitute this back into (3.11) we find that V_0 satisfies

$$V_{0_t} + \frac{\sigma^2 S^2}{2} V_{0_{SS}} + rSV_{0_S} - rV_0 = -\frac{\delta(t,T)(\mu - r)^2}{2\gamma\sigma^2}. \quad (3.13)$$

The particular solution of this with zero final data is, remembering that $\delta(t,T) = e^{-r(T-t)}$,

$$\frac{\delta(t,T)(\mu - r)^2(T - t)}{2\gamma\sigma^2}. \quad (3.14)$$

The general solution is thus any solution satisfying the Black–Scholes equation plus the particular solution (3.14).

We then retrace our steps to get from V_0 to V, the option price, using (3.10) and the boundary conditions (2.19) and (2.20) respectively. We find that the leading order final data in the portfolio without the option liability, W_{wo}, is $V_0(S,T) = 0$, whereas in the portfolio with the call option liability, W_w, it has the usual payoff functional form $V_0(S,T) = -\max(S-E,0)$. So from the linearity of (3.13) we see that, to leading order (or in the absence of any costs) the option value is simply the Black–Scholes value. Similarly the extra number of shares required in the portfolio with the additional option liability is, to leading order, the Black–Scholes delta value.

We now consider the terms independent of Y, which give an equation for V_1:

$$V_{1_t} - rV_1 + \frac{\sigma^2 S^2}{2} V_{1_{SS}} + V_{1_S}\left(\mu S - \frac{\gamma\sigma^2 S^2}{\delta(t,T)}\left(y^* + V_{0_S}\right)\right) = 0.$$

If we substitute for $V_{0_S} + y^*$ using (3.12), we find that V_1 satisfies the Black–Scholes

https://doi.org/10.1017/S095679259900368X Published online by Cambridge University Press

equation. The final condition for this equation for both W_w and W_{wo} is $V_1(S,T) = 0$. (This is found by expanding the final conditions in powers of $\epsilon^{\frac{1}{4}}$ and considering the terms of $O(\epsilon^{\frac{1}{4}})$.)

Thus V_1 is identically zero for all S and $t < T$, and so the leading order correction to the Black–Scholes value occurs at the $O(\epsilon^{\frac{1}{2}})$ level.

3.4 The $O(\epsilon^{\frac{1}{2}})$ equation

We now take the analysis to higher order to find the correction to the Black–Scholes value due to transaction costs. If we examine the $O(\epsilon^{\frac{1}{2}})$ terms in (3.4), we find that

$$V_{2_t} - rV_2 + \mu S V_{2_S} + \frac{\sigma^2 S^2}{2}\left(V_{2_{SS}} + y_S^{*^2} V_{4_{YY}} - \frac{\gamma}{\delta(t,T)}\left(Y^2 + 2V_{2_S}\left(y^* + V_{0_S}\right)\right)\right) = 0. \quad (3.15)$$

This is an *ordinary* differential equation for V_4 with dependent variable Y. We note in passing that the equivalent ordinary differential equations for V_0 - V_3 as a function of Y are homogeneous. The application of the relevant boundary conditions leads us to conclude that V_0, V_1, V_2 and V_3 are independent of Y.

Equation (3.15) is easily integrated to give

$$V_4 = \frac{\gamma Y^4}{12\delta(t,T)y_S^{*^2}} - \frac{Y^2 \mathscr{L}_{BS}(V_2)}{\sigma^2 S^2 y_S^{*^2}} + cY \quad (3.16)$$

where without loss of generality we set $V_4(S,0,t) = 0$, and

$$\mathscr{L}_{BS}(V_2) = V_{2_t} + rSV_{2_S} + \frac{\sigma^2 S^2}{2} V_{2_{SS}} - rV_2.$$

The conditions of continuity and smoothness that we shall require, and describe shortly, will be sufficient to find all of $Y^+(S,t)$, $Y^-(S,t)$, $\hat{Y}^+(S,t)$ and $\hat{Y}^-(S,t)$, as well as $\mathscr{L}_{BS}(V_2)$ and c.

3.5 Boundary conditions

The six unknown functions are listed at the end of the previous section. To determine the correct boundary conditions we must refer to Figure 2 in some detail. Suppose we begin in the no-transaction region, that is $-Y^- \leqslant Y \leqslant Y^+$. Should the random walk of the asset take us to the edge of this region, for example to $Y = Y^+$, the edge of the sell region, then we must trade to some interior point. We trade from point B to point A; both A and B are to be determined. Such a trade costs $K(S, Y^+ - \hat{Y}^+)$ in scaled variables. This cost is offset exactly by the added value of the better-hedged position. Thus, the value functions at A and B must differ by this cost (this is the 'value-matching condition') i.e.

$$V_4(S, \hat{Y}^+, t) - V_4(S, Y^+, t) = K(S, Y^+ - \hat{Y}^+). \quad (3.17)$$

When the costs include a fixed element, there is a discrete jump in the number of shares held, (Y), and this results in a jump in the value function. We shall consider the case of purely proportional costs shortly.

Furthermore, the points A and B (or Y^+ and $\hat{Y}^+$) are optimal. Thus (these are the 'smooth pasting' conditions),

$$\frac{\partial V_4}{\partial \hat{Y}^+}(S, \hat{Y}^+, t) = \frac{\partial K}{\partial \hat{Y}^+}(S, Y^+ - \hat{Y}^+) = -\frac{\partial K}{\partial U}(S, U = Y^+ - \hat{Y}^+) \tag{3.18}$$

and

$$-\frac{\partial V_4}{\partial Y^+}(S, Y^+, t) = \frac{\partial K}{\partial Y^+}(S, Y^+ - \hat{Y}^+) = \frac{\partial K}{\partial U}(S, U = Y^+ - \hat{Y}^+), \tag{3.19}$$

or equivalently,

$$\frac{\partial V_4}{\partial Y}(S, \hat{Y}^+, t) = \frac{\partial V_4}{\partial Y}(S, Y^+, t) = -\frac{\partial K}{\partial U}(S, U = Y^+ - \hat{Y}^+). \tag{3.20}$$

Similar equations hold at the buy boundary and rebalance point; simply change signs in the obvious manner.

The above equations hold when Y^+ and $\hat{Y}^+$ are distinct. They coincide when transaction costs have no fixed component; this was the case discussed in Whalley & Wilmott [11], HN and DPZ. In this case, control of the number of shares held is instantaneous and no jump occurs ($Y^+ = \hat{Y}^+$), so the value-matching condition becomes

$$\frac{\partial V_4}{\partial Y}(S, Y^+, t) = -\frac{\partial K}{\partial U}(S, U = Y^+ - \hat{Y}^+), \tag{3.21}$$

which replaces (3.17). The optimality condition at $Y^+ = \hat{Y}^+$ also changes: $\partial^2 V_4/\partial Y^2$ is required to be continuous there (across the edge of the no-transaction region). This gives us the alternative condition to (3.20) (one less condition is required since $Y^+ = \hat{Y}^+$)

$$\frac{\partial^2 V_4}{\partial Y^2}(S, Y^+, t) = -\frac{\partial^2 K}{\partial U^2}(S, U = Y^+ - \hat{Y}^+)(= 0). \tag{3.22}$$

We use (3.21) and (3.22) to derive the equations in § 4.1. For the remainder of this section we shall consider only the case with $Y^+ \neq \hat{Y}^+$. For more detailed discussion of the $Y^+ = \hat{Y}^+$ case, see Whalley & Wilmott [11].

We have here only given a heuristic explanation of the boundary conditions. Dumas [17] and Dixit [18] derive them more rigorously.

We could use equation (3.16) to derive the six equations for the six unknowns. However, these equations are unwieldy and writing them down explicitly does not add any insight into our problem. Instead, we shall for the rest of this paper concentrate on the special case in which buying and selling is 'symmetric' in the sense that the cost of a purchase or sale only depends on the magnitude of the trade. Thus the function $K(S, U)$ is an even function of U. In this case, we can write $Y^+ = Y^-$ and $\hat{Y}^+ = \hat{Y}^-$, with $c = 0$.

We find that the equations for the edge of the hedging bandwidth and the optimal rebalance point are

$$\frac{\gamma}{3\delta(t,T) y_S^{*2}} Y^+ \hat{Y}^+ (Y^+ + \hat{Y}^+) = \frac{\partial K}{\partial U}(S, Y^+ - \hat{Y}^+) \tag{3.23}$$

and

$$\frac{\gamma}{12\delta(t,T) y_S^{*2}} (Y^+ - \hat{Y}^+)(Y^+ + \hat{Y}^+)^3 = K(S, Y^+ - \hat{Y}^+). \tag{3.24}$$

with the equation for V_2 now being

$$\mathscr{L}_{BS}(V_2) = \frac{\gamma \sigma^2 S^2}{6\delta(t,T)} (Y^{+^2} + Y^+ \hat{Y}^+ + \hat{Y}^{+^2}). \tag{3.25}$$

https://doi.org/10.1017/S095679259900368X Published online by Cambridge University Press

To implement this model, we first solve (3.23) and (3.24) to find Y^+ and $\hat{Y}^+$ as functions of S, t and the Black–Scholes Gamma, $V_{0_{SS}}$ (via y_S^*). This gives us our hedging strategy for each portfolio seperately and hence for the option, i.e. hedging bandwidth and rebalance points, right up to expiry of the option. To find the option *value* we put Y^+ and $\hat{Y}^+$ into (3.25). This equation is a Black–Scholes-type diffusion equation for V_2 with a non-zero right-hand side that is a function of the costs and the Black–Scholes gamma.

4 Special cases

4.1 Proportional costs only

For our first special case let us consider proportional costs only. Costs proportional to value traded were first examined by Whalley & Wilmott [11]. We generalize this to costs proportional to the number (quantity) of assets traded ($k_2|U|$) plus those proportional to the value traded ($k_3S|U|$), since the results are computationally similar. For example, commission costs of trading are often of the form $k_2|U|$. and the bid-ask spread represents a cost of the form $k_3S|U|$. So $K(S,U) = (k_2 + k_3S)|U|$. From equations (3.21) and (3.22) we find that

$$V_{2_t} + rSV_{2_S} + \frac{\sigma^2S^2}{2}V_{2_{SS}} - rV_2 = \frac{\sigma^2S^2}{2}\left(\frac{9\gamma(k_2+k_3S)^2}{4\delta(t,T)}\right)^{\frac{1}{3}}\left(\left|V_{0_{SS}} + \frac{\delta(t,T)(\mu-r)}{\gamma S^2\sigma^2}\right|\right)^{\frac{4}{3}} \quad (4.1)$$

with

$$Y^+ = \left(\frac{3(k_2+k_3S)\delta(t,T)}{2\gamma}\left(V_{0_{SS}} + \frac{\delta(t,T)(\mu-r)}{\gamma S^2\sigma^2}\right)^2\right)^{\frac{1}{3}}.$$

and

$$\hat{Y}^+ = Y^+.$$

This last equation tells us that the optimal rebalance point coincides with the edge of the hedging bandwidth. In other words, when we rehedge we only do the minimum amount of trading to stay inside the bandwidth. The equation for Y^+ gives the semi-width of the hedging bandwidth.

Equation (4.1) is to be solved subject to the final condition $V_2(S,T) = 0$.

It is now important to distinguish between the two problems for W_w and W_{wo}. The V_2 component of W_{wo} satisfies (4.1) with $V_{0_{SS}} = 0$ i.e.

$$V_{2_t} + rSV_{2_S} + \frac{\sigma^2S^2}{2}V_{2_{SS}} - rV_2 = \frac{1}{2}\left(\frac{3(k_2+k_3S)}{2\sigma S}\right)^{\frac{2}{3}}\frac{\delta(t,T)(\mu-r)^{\frac{4}{3}}}{\gamma}. \quad (4.2)$$

Using $V^{BS}(S,t)$ to denote the Black–Scholes option value we see that the V_2 component of W_w satisfies (4.1) with $V_{0_{SS}}$ being the Black–Scholes value for the gamma, i.e. V_{SS}^{BS}. Denoting the solution of this by V_2^w and the solution of (4.2) by V_2^{wo}, the option value correct to $O(\epsilon^{\frac{1}{2}})$ is given by

$$V(S,t) = V^{BS}(S,t) + \epsilon^{\frac{1}{2}}\left(V_2^w(S,t) - V_2^{wo}(S,t)\right) + \cdots,$$

These equations are easily solved by finite-difference methods (see Wilmott, Dewynne & Howison [19]). This is clearly much faster to do that to solve the original three-dimensional problem.

https://doi.org/10.1017/S095679259900368X Published online by Cambridge University Press

We can also derive the hedging strategy for an option by comparing the boundaries of the no-transaction regions with and without the option liability. Thus if we write y_w^* and y_{wo}^* for the centres, and $\pm Y_w^+$ and $\pm Y_{wo}^+$ for the edges of the no-transaction region respectively with $(\cdot_w)$ and without $(\cdot_{wo})$ the option position and set

$$D^+ = y_w^* + Y_w^+ - (y_{wo}^* + Y_{wo}^+) \tag{4.3}$$

$$D^- = y_w^* - Y_w^+ - (y_{wo}^* - Y_{wo}^+) \tag{4.4}$$

and

$$-\Delta = y_w^* - y_{wo}^* \tag{4.5}$$

then we see to leading order that the centre of the no-transaction region, to which the region collapses in the absence of transaction costs ($\epsilon \to 0$), is

$$\Delta = V_{0_S} \tag{4.6}$$

which is exactly the Black–Scholes delta.

The bandwidth, $D^- \leqslant -\Delta \leqslant D^+$, within which the actual number of assets is allowed to vary when transaction costs are present, is given by

$$-V_{0_S} - \left(\frac{3(k_2+k_3S)\delta(t,T)}{2\gamma}\right)^{\frac{1}{3}}\left(\left|V_{0_{SS}} + \frac{\delta(t,T)(\mu-r)}{\gamma S^2\sigma^2}\right|^{\frac{2}{3}} - \left(\frac{\delta(t,T)(\mu-r)}{\gamma S^2\sigma^2}\right)^{\frac{2}{3}}\right)$$
$$\leqslant -\Delta \leqslant$$
$$-V_{0_S} + \left(\frac{3(k_2+k_3S)\delta(t,T)}{2\gamma}\right)^{\frac{1}{3}}\left(\left|V_{0_{SS}} + \frac{\delta(t,T)(\mu-r)}{\gamma S^2\sigma^2}\right|^{\frac{2}{3}} - \left(\frac{\delta(t,T)(\mu-r)}{\gamma S^2\sigma^2}\right)^{\frac{2}{3}}\right)$$

When $V_{0_{SS}} \gg \Lambda = (\delta(t,T)(\mu-r))/(\gamma S^2\sigma^2)$, we can approximate this by

$$-V_{0_S} - \left(\tfrac{3(k_2+k_3S)\delta(t,T)}{2\gamma}\right)^{\frac{1}{3}}|V_{0_{SS}}|^{\frac{2}{3}}$$
$$\leqslant -\Delta \leqslant$$
$$-V_{0_S} + \left(\tfrac{3(k_2+k_3S)\delta(t,T)}{2\gamma}\right)^{\frac{1}{3}}|V_{0_{SS}}|^{\frac{2}{3}}.$$

This is valid when $V_{0_{SS}}$ is large, *e.g.* for times close to expiry and asset prices close to the exercise price, for large values of the risk aversion parameter, γ, and for small values of the expected growth rate of the asset above the risk-free rate, all of which decrease Λ.

In general we see that the bandwidth increases with the option's gamma: that is, the bandwidth is relatively large wherever the gamma is large, in particular close to the strike price near expiry.

Mohamed [12] has performed a Monte Carlo simulation of this hedging strategy in the special case $\mu = r$ and found that it was the best strategy he tested (he also tested fixed-period hedging and hedging based on fixed changes in delta).

https://doi.org/10.1017/S095679259900368X Published online by Cambridge University Press

4.2 Fixed costs only

When costs only have a fixed component we have $K(S,U) = k_1$. In this case we apply equations (3.17) and (3.20) and find that

$$V_{2_t} + rSV_{2_S} + \frac{\sigma^2 S^2}{2} V_{2_{SS}} - rV_2 = \sigma^2 S^2 \left(\frac{k_1 \gamma}{3\delta(t,T)}\right)^{\frac{1}{2}} \left| V_{0_{SS}} + \frac{\delta(t,T)(\mu - r)}{\gamma S^2 \sigma^2}\right| \tag{4.7}$$

with

$$V_2(S,T) = 0,$$

$$Y^+ = \left(\frac{12 k_1 \delta(t,T)}{\gamma}\left(V_{0_{SS}} + \frac{\delta(t,T)(\mu - r)}{\gamma S^2 \sigma^2}\right)^2\right)^{\frac{1}{4}},$$

and

$$\hat{Y}^+ = 0.$$

The last equation tells us that rehedging should be to the centre of the band: since costs are fixed one might as well hedge to maximize the expected time to the next hedge.

We now distinguish between the two problems for W_w and W_{wo}. The V_2 component of W_{wo}, V_2^{wo}, satisfies (4.7) with $V_{0_{SS}} = 0$, i.e.

$$\mathscr{L}_{BS}(V_2) = V_{2_t} + rSV_{2_S} + \frac{\sigma^2 S^2}{2} V_{2_{SS}} - rV_2 = \frac{k_1^{\frac{1}{2}}(\mu - r)}{(3\gamma)^{\frac{1}{2}}} e^{-r(T-t)/2},$$

and with zero final condition.

The V_2 component of W_w, V_2^w, satisfies (4.7) with $V_{0_{SS}}$ being the Black–Scholes value for the gamma, i.e. V_{SS}^{BS}. The option value correct to $O(\epsilon^{\frac{1}{2}})$ is again given by

$$V(S,t) = V^{BS}(S,t) + \epsilon^{\frac{1}{2}}\left(V_2^w(S,t) - V_2^{wo}(S,t)\right) + \cdots, \tag{4.8}$$

One point to note about the above is the appearance of the modulus of the gamma in equation (4.7). A term like this was first found by Hoggard, Whalley & Wilmott [6] in their extension of the Leland (fixed hedging period) model.

Again, we can extract the hedging strategy for the option position by comparison of the edges of the bandwidths in the portfolios with and without the options. In this case, we have

$$|D^\pm + V_{0_S}| = \left(\frac{12\delta(t,T)k_1}{\gamma}\right)^{\frac{1}{4}}\left(\left|V_{0_{SS}} + \frac{\delta(t,T)(\mu - r)}{\gamma S^2\sigma^2}\right|^{\frac{1}{2}} - \left(\frac{\delta(t,T)(\mu - r)}{\gamma S^2 \sigma^2}\right)^{\frac{1}{2}}\right),$$

or, when $V_{0_{SS}} \gg \Lambda$,

$$|D^\pm + V_{0_S}| = \left(\frac{12\delta(t,T)k_1}{\gamma}\right)^{\frac{1}{4}} |V_{0_{SS}}|^{\frac{1}{2}}.$$

The bandwidth again increases with the modulus of the gamma, and for large Γ is proportional to a fractional power of it.

https://doi.org/10.1017/S095679259900368X Published online by Cambridge University Press

4.3 Proportional plus fixed cost

When $K(S,U) = k_1 + (k_2 + k_3 S)|U|$ we must solve

$$Y^+\hat{Y}^+(Y^+ + \hat{Y}^+) = \frac{3(k_2 + k_3 S)\delta(t,T)y_S^{*2}}{\gamma} \tag{4.9}$$

$$(Y^+ - \hat{Y}^+)^3(Y^+ + \hat{Y}^+) = \frac{12k_1\delta(t,T)y_S^{*2}}{\gamma} \tag{4.10}$$

for Y^+ and $\hat{Y}^+$. Then $\mathscr{L}(V_2)$ is given by equation (3.25). Equations (4.9) and (4.10) must be solved numerically.

Even without solving the full problem numerically, we can however gain some useful insights into the effect of transaction costs on option hedging strategies and valuation. We shall consider two of these in the next section.

5 Hedging portfolios

5.1 Economies of scale

In this section we consider the effects of scale on the pricing and hedging of a portfolio of options in the presence of fixed and proportional transaction costs.

Consider hedging first a single option, where the hedging has a certain fixed cost k_1 associated with it and also proportional costs $k_2|N|$ and $k_3 S|N|$, where N is the quantity of the underlying asset traded. For a single option we would expect the k_1 fixed term to have the greatest impact (otherwise the problem reduces to the proportional case studied in §4.1 for all sizes of derivative portfolios). In this case, the problem with general costs $k_1 + k_2|N| + k_3 S|N|$ behaves (to leading order) as in the analysis in §4.2 for purely fixed costs: hedging is to the optimal quantity of assets, at the centre of the hedging band which has width proportional to $k_1^{\frac{1}{4}}|V_{SS}|^{\frac{1}{2}}$, and the option value is given by (4.8).

As the number of options in the portfolio increases, the relative effect of the fixed costs of transactions, k_1, decreases whilst that of the proportional costs, $k_2|N| + k_3 S|N|$ increase as $|N|$ increases. For an extremely large portfolio of options, we would expect the proportional terms to have the dominant effect, so the bandwidth becomes proportional to $(k_2+k_3S)^{\frac{1}{3}}|V_{SS}|^{\frac{2}{3}}$ and rehedging is of the minimum possible amount so as to stay within the hedging band. For intermediate sizes of option portfolios, where the relative sizes of the transaction cost terms are comparable, the optimal rebalance points move gradually from the centre of the band to the edges.

We can formalize this by writing

$$V = nZ,$$

where n represents the quantity of options held in the portfolio. This implies that

$$y_w^* = nZ_{0_S} - \frac{(\mu - r)\delta(t,T)}{\gamma\sigma^2 S}$$

$$\frac{\partial y_w^*}{\partial S} = nZ_{0_{SS}} + \frac{(\mu - r)\delta(t,T)}{\gamma\sigma^2 S^2}$$

and for $n \gg 1$, that $(y_w^*)_S = O(n)$.

https://doi.org/10.1017/S095679259900368X Published online by Cambridge University Press

Considering the general case of fixed (k_1) and proportional (k_2, k_3) costs from § 4.3 we substitute to obtain

$$\zeta\eta(\zeta+\eta) = \frac{3(k_2+k_3S)\delta(t,T)}{\gamma}\frac{\left(nZ_{0_{SS}}+\xi\right)^2}{n^3} = O(n^{-1}), \tag{5.1}$$

$$(\zeta-\eta)^3(\zeta+\eta) = \frac{12k_1S\delta(t,T)}{\gamma}\frac{\left(nZ_{0_{SS}}+\xi\right)^2}{n^4} = O(n^{-2}), \tag{5.2}$$

where we have used the notation $Y_w^+ = n\zeta$, $\hat{Y}_w^+ = n\eta$, and $\xi = (\mu - r)\delta(t,T)/(\gamma\sigma^2S^2)$.

As $n \to \infty$ the contribution from ξ and from $Y_{wo}^+, \hat{Y}_{wo}^+$ can be neglected and there are two alternatives: either

$$O(\eta) \ll O(\zeta) \quad \text{and} \quad O(\zeta-\eta) = O(\zeta),$$

or

$$O(\eta) = O(\zeta) \quad \text{and} \quad O(\zeta-\eta) \ll O(\zeta).$$

In the former case, we see from (5.2) that $O(\zeta) = O(n^{\frac{1}{2}})$, and so from (5.1), $O(\eta) = O(1) > O(\zeta)$, which is a contradiction. Hence we must have

$$O(\eta) = O(\zeta) = O(n^{-\frac{1}{3}})$$

and

$$O(\zeta-\eta) = O(n^{-\frac{5}{9}}).$$

We see that the relative size of both the width of the hedging band per option, ζ, and the distance of the optimal rebalance point from the centre of the band per option, η, decrease (as $n^{-\frac{1}{3}}$) as the number of options in the portfolio, n, increases. However, the relative distance between the edge of the hedging band and the optimal rebalance point also decreases (as $n^{-\frac{5}{9}}$, so faster than the relative decrease in the bandwidth), and the ratio of the distance of the optimal rebalance point from the hedging bandwidth to the width of the hedging bandwidth is $O(n^{-\frac{2}{9}})$ i.e.

$$\frac{\zeta-\eta}{\zeta} = O(n^{-\frac{2}{9}}).$$

Thus the optimal rebalance point moves relatively closer and closer to the edge of the hedging bandwidth as the size of the portfolio increases, demonstrating the obvious economies of scale.

5.2 The marginal value of an option

Finally, let us look at the marginal value of an option in a portfolio. We shall assume that costs are only proportional to the value traded.

Suppose that we have a portfolio of options (expiring on the same date for simplicity) with value $P(S,t)$. According to the above analysis this function is given by the solution of a certain partial differential equation. Now, suppose that we have the opportunity to add another option into our portfolio. Let us call the value of this new portfolio $\bar{P}(S,t) = P(S,t) + V(S,t)$. We shall find the equation satisfied by the marginal value of the new option $V(S,t)$.

The function $\bar{P}$ satisfies the same equation as P. The only difference between them is that they have different payoffs at the expiry date: $\bar{P}$ contains the extra obligation of the additional option. If we assume that the value V of our extra option is small compared with the value of our original portfolio P, i.e.

$$|V| \ll |P|,$$

then we can readily find the equation satisfied by V by expanding terms such as

$$\left| P_{SS} + \frac{\delta(t,T)(\mu - r)}{\gamma S^2 \sigma^2} + V_{SS} \right|^{\frac{4}{3}}$$

using the binomial expansion.

We find that the option value V is given by

$$V(S,t) = V^{BS}(S,t) + \epsilon^{\frac{1}{2}} V_2(S,t) + \cdots.$$

Here V_2 satisfies

$$\mathscr{L}(V_2) = \frac{2\delta(t,T)}{3\gamma} \left(\frac{3\gamma^2 S^4 \sigma^3}{2\delta(t,T)^2} \right)^{\frac{2}{3}} \operatorname{sgn}\left(P_{SS} + \frac{\delta(t,T)(\mu - r)}{\gamma S^2 \sigma^2} \right) \left| P_{SS} + \frac{\delta(t,T)(\mu - r)}{\gamma S^2 \sigma^2} \right|^{\frac{1}{3}} V_{SS}^{BS},$$

where V^{BS} is now the Black–Scholes value of the additional option.

Whether this gives an option value that is greater or less than the Black–Scholes depends upon the sign of the product of the Black–Scholes Gamma of the option and

$$P_{SS} + \frac{\delta(t,T)(\mu - r)}{\gamma S^2 \sigma^2}.$$

This can be explained quite simply by example. Assume for the sake of simplicity that $\mu = r$. Suppose that we hold a portfolio with a net negative gamma. If we add to this portfolio an option also having a negative Black–Scholes gamma then we are making our position with regard to transaction costs worse. However, if we add an option with a positive gamma we are improving our position by reducing our transaction costs. Thus we see that it is very important to examine the marginal effect of an option on the portfolio.

A similar effect is observed in the behaviour of the hedging bandwidth. Obviously the centre of the no-transaction region is at the Black–Scholes delta, but the increase in the size of the semi-width of this band is given by

$$\epsilon^{\frac{1}{4}} \frac{2}{3} \left(\frac{3S\delta(t,T)}{2\gamma} \right)^{\frac{2}{3}} \operatorname{sgn}\left(P_{SS} + \frac{\delta(t,T)(\mu - r)}{\gamma S^2 \sigma^2} \right) \left| P_{SS} + \frac{\delta(t,T)(\mu - r)}{\gamma S^2 \sigma^2} \right|^{-\frac{1}{3}} V_{SS}^{BS}.$$

Thus the bandwidth can be greater or smaller with the addition of the new claim, or possibly greater for some values of S and t but smaller for other values.

6 Conclusion

In this paper we have derived a generalization of the global-in-time transaction cost model of Davis, Panas & Zariphopoulou. This optimizes the hedging strategy under a negative exponential utility function. We adopt an arbitrary transaction cost structure, and have simplified the Davis, Panas & Zariphopoulou three-dimensional free boundary problem using asymptotic analysis. This results in a single parabolic partial differential equation for the option value.

The optimal ‘hedging bandwidth’ around the Black–Scholes delta, and optimal ‘rebalance points’, to which the portfolio should be adjusted, are given explicitly as the solution of simple equations. In the case of purely fixed costs, transactions take place to the centre of the bandwidth; however for costs proportional to the size of the transaction, the optimal rebalance point moves away from the centre of the band, and, as the value of the portfolio increases, the proportional costs increase in significance relative to the fixed costs, moving the rebalance points closer to the edges of the bandwidth.

The nonlinearity of the option valuation equation also has interesting consequences for the marginal value of an option. The value of such an extra option is strongly affected by the portfolio to which it is being added, so a short option can have a lower fair value to its writer than the Black–Scholes price if it is offsetting the overall portfolio position! Explicit equations are given for these effects.

For a very general transaction cost regime, the solutions for the bandwidth, rebalance points, and option value still need to be found numerically, but the asymptotically derived equations in this paper are significantly less demanding and computationally time-consuming to solve than the three-dimensional free boundary problem from which they were derived.

Acknowledgement

P. W. would like to thank the Royal Society for their support.

References

[1] Hodges, S. D. & Neuberger, A. (1989) Optimal replication of contingent claims under transaction costs. *Rev. Futures Markets* **8**, 222–239.

[2] Davis, M. H. A., Panas, V. G. & Zariphopoulou, T. (1993) European option pricing with transaction costs. *SIAM J. Control & Optimisation* **31**, 470–493.

[3] Black, F. & Scholes, M. (1973) The pricing of options and corporate liabilities. *J. Political Economy* **81**, 637–54.

[4] Leland, H. E. (1985) Option pricing and replication with transaction costs. *J. Finance* **40**, 1283–301.

[5] Boyle, P. P. & Vorst, T. (1992) Option replication in discrete time with transaction costs. *J. Finance* **47**, 271–293.

[6] Hoggard, T., Whalley, A. E. & Wilmott, P. (1994) Hedging option portfolios in the presence of transaction costs. *Advances in Futures and Options Res.* **7**, 21–35.

[7] Avellaneda, M. & Paras, A. (1995) Optimal hedging portfolios for derivative securities in the presence of large transaction costs. *Appl. Math. Finance* **2**.

[8] Toft, K. B. (1996) On the mean variance tradeoff in option replication. *J. Financial & Quantitative Anal.* **31**(2).

[9] Whalley, A. E. & Wilmott, P. (1993) A comparison of hedging strategies. *ECMI Proc.* 427–434.

[10] Henrotte, P. (1993) Transaction Costs and Duplication Strategies. *Working paper*, Stanford University et Groupe HEC, Jouy-en-Josas.

[11] Whalley, A. E. & Wilmott, P. (1997) An asymptotic analysis of the Davis, Panas & Zariphopoulou model for option pricing with transaction costs. *Math. Finance* **7**, 307–324.

[12] MOHAMED, B. (1994) Simulations of transaction costs and optimal rehedging. *Appl. Math. Finance* **1**, 49–62.

[13] ATKINSON, C. & WILMOTT, P. (1995) Portfolio management with transaction costs: an asymptotic analysis. *Math. Finance* **5**.

[14] SHREVE, S. E. & SONER, H. M. (1994) Optimal investment and consumption with transaction costs. *Ann. Appl. Probability* **4**, 680–692.

[15] FLEMING, W. H., GROSSMAN,S. G., VILA, J.-L. & ZARIPHOPOULOU, T. (1990) Optimal portfolio rebalancing with transaction costs. *Working paper*, Division of Applied Mathematics, Brown University.

[16] DIXIT, A. (1991) Analytical Approximations in Models of Hysteresis. *Rev. Economic Stud.* **58**, 141–151.

[17] DUMAS, B. (1991) Super contact and related optimality conditions. *J. Economic Dynamics & Control* **15**, 675–685.

[18] DIXIT, A. (1991) A simplified treatment of the theory of optimal regulation of Brownian motion. *J. Economic Dynamics & Control* **15**, 657–673.

[19] WILMOTT, P., DEWYNNE, J. N. & HOWISON, S. D. (1993) *Option Pricing: Mathematical models and computation*, Oxford Financial Press.