Asymptotic Analysis for Optimal Investment and Consumption with Transaction Costs

Karel Janeček
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, PA 15213
kjanecek@andrew.cmu.edu

Steven E. Shreve[1]
Department of Mathematical Sciences and
Center for Computational Finance
Carnegie Mellon University
Pittsburgh, PA 15213
shreve@cmu.edu

July 7, 2003

Abstract

We consider an agent who invests in a stock and a money market and consumes in order to maximize the utility of consumption over an infinite planning horizon in the presence of a proportional transaction cost $\lambda > 0$. The utility function is of the form $U(c) = c^{1-p}/(1-p)$ for $p > 0$, $p \neq 1$. We provide a heuristic and a rigorous derivation of the asymptotic expansion of the value function in powers of $\lambda^{1/3}$, and we also obtain asymptotic results on the boundary of the "no-trade" region.

Short title: Asymptotic Transaction Costs
JEL classification: G13
Mathematics Subject Classification (1991): 90A09, 60H30, 60G44

[1] Work supported by the National Science Foundation under grants DMS-0103814 and DMS-0139911.

1 Introduction

Consider an agent seeking to optimally invest and consume in the presence of proportional transaction costs. The agent can invest in a stock, modelled as a geometric Brownian motion, and in a money market with constant rate of interest. She may also consume and get utility $U(c) = c^{1-p}/(1-p)$, where $p > 0$, $p \neq 1$. In addition, the agent must pay a proportional transaction cost $\lambda > 0$ for transferring capital between the stock and money market. All consumption is done from the money market. The agent wishes to maximize the expected discounted integral over $[0, \infty)$ of the utility of consumption.

When the transaction cost λ is zero, the agent's optimal policy is to keep a constant proportion of wealth, which we call the *Merton proportion* and denote θ_p, invested in the stock; see Merton [40]. When $\lambda > 0$, the optimal policy is to maintain a position inside a wedge called the *No Trade* (NT) region. Trading occurs when the position hits the boundaries of the NT region and no trading occurs in the interior of NT. If the agent's position is initially outside NT, she should immediately sell or buy stock in order to move to the boundary of NT. Except when the left boundary of NT is the positive y-axis, the set of trading times has zero Lebesgue measure, and the total amount of capital traded can be characterized by a possible initial jump plus local time on the boundary of NT. When the left boundary of NT is the positive y-axis, the agent will stay on the y-axis once she arrives there. This problem was formulated by Magill & Constantinides [39], solved under restrictive conditions by Davis & Norman [14], and thoroughly analyzed by Shreve & Soner [43]. An explicit formula for the solution is not known.

The Hamilton-Jacobi-Bellman (HJB) equation for this problem is a partial differential equation in two variables. For the power utility functions considered here, this can be reduced to an equation in one variable. Numerical results are provided by [1], [46], [47]. A useful and perhaps more informative approach for obtaining explicit results, the approach of this paper, is to develop a power series expansion for the value function and the boundaries of the NT region in powers of λ. For example, Constantinides [8] numerically computed the effect of transaction costs on the value function for our problem, and observed that transaction costs have a "first-order effect on assets' demand" and a "second-order effect on equilibrium asset return." This effect has been made precise by formal power series expansions in a variety of models. Atkinson & Wilmott [2] accomplished this for a model with fixed transaction costs studied by Morton & Pliska [41]. Whalley & Wilmott [49] applied this approach to utility-based option pricing, described below. Korn [33] developed it for maximization of exponential utility in the presence of fixed and proportional transaction costs. Korn & Laue [34] carried this analysis farther, including the case of two stocks and a money market. In this paper, we treat only the case of proportional transaction costs, first obtaining a formal power series expansion in the spirit of the previous papers, but then providing a rigorous justification for the leading terms in the expansion. In particular, we observe that the width of

the NT region is large, of order $\lambda^{1/3}$, whereas the effect of transaction cost on the value function is smaller, of order $\lambda^{2/3}$.

The asymptotic expansion of this paper is valid for any Merton proportion θ_p, except for the case in which all wealth is invested in stock ($\theta_p = 1$). In particular, we refute the conjecture in [43] that the Merton line (the set of stock/money market positions which are in the Merton proportion) is outside NT for $\theta_p > 1$. This is the case for *sufficiently high* transaction costs. However, for small transaction costs the Merton line is inside NT.

In the presence of transaction costs, contingent claim pricing by replication, or more generally by super-replication, has received considerable attention but often leads to a trivial result: the cheapest strategy is buy-and-hold. See [6], [13], [17], [19], [28], [29], [31], [32], [37], [44].

One alternative to the conservative super-replication method for contingent claim prices, pioneered by Leland [36], is to strike a balance between transaction costs and "hedge slippage," and this leads to a modified Black-Scholes equation; see, e.g., [3], [4], [5], [7], [18], [21], [23], [45]. Another method, proposed by Hodges and Neuberger [22], is to price an option so that a utility maximizer is indifferent between having a certain initial capital for investment or holding the option but having initial capital reduced by the price of the option. This produces both a price and a hedge, the latter being the difference in the optimal trading strategies in the problem without the option and the problem with the option. This utility-based option pricing is examined in [9], [10], [15]. A formal asymptotic analysis of such an approach appears in [49]. Once again, the methodology developed in this paper suggests how to make this analysis rigorous. We note that in some of these papers the utility function is $U(c) = c^{1-p}/(1-p)$ with p restricted to be in $(0,1)$. We include $p > 1$ in our analysis because $p \in (0,1)$ leads to intolerably risky behavior. See Samuelson [42] for the argument in words of one syllable that this is the case even for logarithmic utility ($p = 1$).

The transaction cost problem with multiple assets was studied by [1], [6], [27], [28]. For a jump diffusion model, see [20]. Transaction cost problems have dual formulations which can shed light on their solutions; see [11], [12], [16] [38]. Other papers which study super-martingales and conditions for no-arbitrage in these models are [24], [25], [26], [35], [48].

In Section 2 we set out the model. Section 3 provides a heuristic expansion of the value function in powers of $\lambda^{1/3}$. The key results of Section 3 are proved in Section 4, using viscosity sub- and supersolutions.

2 Model set-up and known results

The model is similar to Shreve & Soner [43]. An agent is given an initial position of x dollars in money market and y dollars in stock. The stock price is given by $dS_t = \alpha S_t \, \mathrm{d}t + \sigma S_t \, \mathrm{d}W_t$, where α and σ are positive constants and $\{W_t, t \geq 0\}$ is a standard Brownian motion on a filtered probability space $\big(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\big)$. We assume a constant interest rate

$r < \alpha$. The agent chooses a policy of three adapted processes C, L, and M. The consumption process C is nonnegative and integrable on each finite interval. The processes L and M are nondecreasing and right-continuous with left limits, and $L_{0-} = M_{0-} = 0$. L_t represents the cumulative value of stock purchased up to time t, and M_t is the cumulative value of stock sold.

Let X_t denote the wealth invested in the money market and Y_t the wealth invested in stock, with $X_{0-} = x$, $Y_{0-} = y$. The agent's position evolves as

$$\mathrm{d}X_t = (rX_t - C_t)\,\mathrm{d}t - (1+\lambda)\,\mathrm{d}L_t + (1-\lambda)\,\mathrm{d}M_t, \tag{2.1}$$

$$\mathrm{d}Y_t = \alpha Y_t\,\mathrm{d}t + \sigma Y_t\,\mathrm{d}W_t + \mathrm{d}L_t - \mathrm{d}M_t. \tag{2.2}$$

The constant $\lambda \in (0,1)$ appearing in these equations accounts for proportional transaction costs, which are paid from the money market account.

Define the *solvency region*

$$\mathcal{S} \triangleq \{(x,y);\, x + (1+\lambda)\,y > 0,\, x + (1-\lambda)\,y > 0\}\,.$$

The policy (C,L,M) is *admissible for* (x,y) if (X_t, Y_t) given by (2.1), (2.2) is in $\overline{\mathcal{S}}$ for all $t \geq 0$. We denote by $\mathcal{A}(x,y)$ the set of all such policies. We note that $\mathcal{A}(x,y) \neq \emptyset$ if and only if $(x,y) \in \overline{\mathcal{S}}$; see [43], Remarks 2.1, 2.2.

We introduce the agent's *utility function* U_p defined for all $c \geq 0$ by $U_p(c) \triangleq c^{1-p}/(1-p)$. (An analysis is also possible for $U_0(c) = \log c$.) Let $\beta > 0$ be a positive discount rate and define the value function

$$v(x,y) = \sup_{(C,L,M)\in\mathcal{A}(x,y)} \mathbb{E}\int_0^\infty e^{-\beta t} U_p(C_t)\,\mathrm{d}t, \quad (x,y) \in \overline{\mathcal{S}}.$$

This problem when $\lambda = 0$ was solved by Merton [40], who determined that the optimal policy always keeps a wealth proportion

$$\theta_p = \frac{1}{p}\cdot\frac{\alpha - r}{\sigma^2}, \tag{2.3}$$

in the stock. We call θ_p the *Merton proportion*. For $\lambda = 0$,

$$v(x,y) = \frac{1}{1-p} A^{-p}(p)(x+y)^{1-p}, \tag{2.4}$$

where

$$A(p) \triangleq \frac{\beta - r(1-p)}{p} - \frac{1}{2}\sigma^2\theta_p^2(1-p). \tag{2.5}$$

The optimal consumption in feedback form is $C_t = A(p)(X_t + Y_t)$.

We assume throughout that $A(p) > 0$, which is necessary and sufficient for finiteness of the value function for the zero transaction cost problem.

We introduce the *convex dual function* $\widetilde{U}_p :\ (0,\infty) \mapsto \mathbb{R}$ defined by

$$\widetilde{U}_p(\tilde{c}) \triangleq \sup_{c>0}\{U_p(c) - c\tilde{c}\} = \frac{p}{1-p}\,\tilde{c}^{\,-(1-p)/p}. \tag{2.6}$$

The supremum in (2.6) is attained by $c = \tilde{c}^{\,-1/p}$.

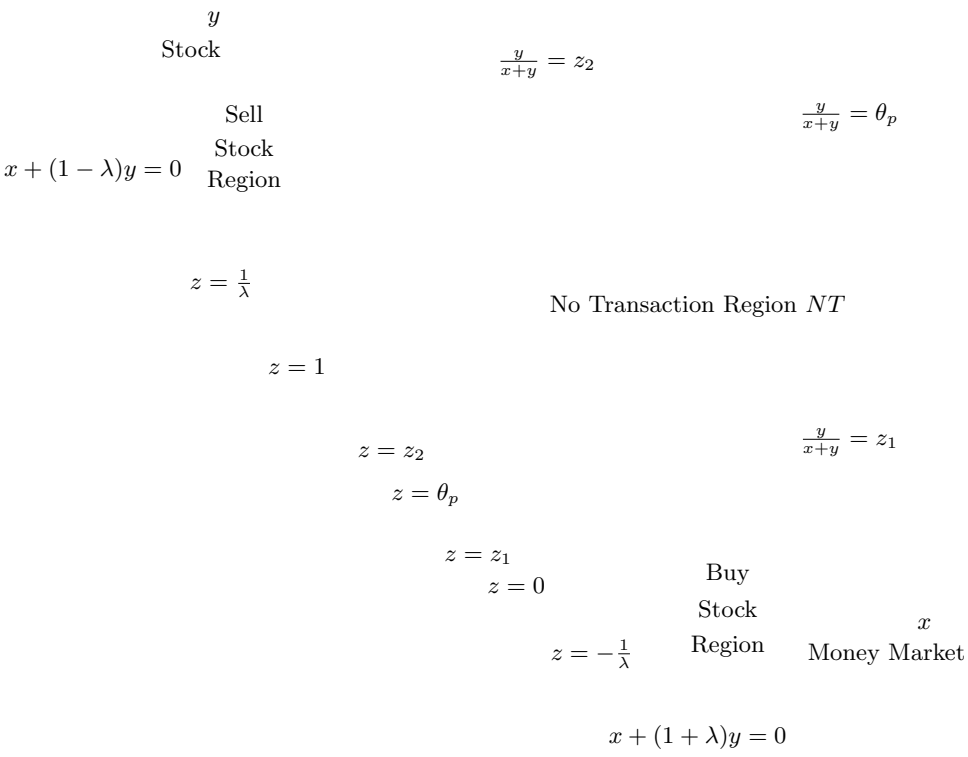

Fig. 1 The solvency region

Shreve and Soner [43] show that the value function is a smooth solution of the *Hamilton-Jacobi-Bellman (HJB) equation*

$$\min\left\{\mathcal{L}v - \widetilde{U}_p(v_x),\ -(1-\lambda)v_x + v_y,\ (1+\lambda)v_x - v_y\right\} = 0, \tag{2.7}$$

where the second-order differential operator $\mathcal{L}$ is given by

$$(\mathcal{L}v)(x,y) = \beta\, v(x,y) - \frac{1}{2}\sigma^2 y^2\, v_{yy}(x,y) - \alpha y\, v_y(x,y) - rx\, v_x(x,y).$$

The optimal policy can be described in terms of two numbers $0 < z_1 < z_2 < 1/\lambda$ which define the *no-trade region* (see [43], Theorem 11.2)

$$NT \triangleq \left\{(x,y) \in \mathcal{S}; z_1 < \frac{y}{x+y} < z_2\right\}.$$

If $Y_t/(X_t+Y_t) < z_1$ one should buy stock to bring this ratio to the boundary $y/(x+y) = z_1$ of NT. If $Y_t/(X_t+Y_t) > z_2$ one should sell stock to bring this ratio to the other boundary $y/(x+y) = z_2$ of NT. For $\theta_p < 1$ we must have $0 < z_1 < \theta_p < z_2 < 1$, so that NT is in the first quadrant. For $\theta_p = 1$, we have $0 < z_1 < z_2 = 1$. In this paper we show that for $\theta_p > 1$ and sufficiently small λ, $1 < z_1 < \theta_p < z_2$, so NT is in the second quadrant.

Power utility functions lead to *homotheticity* of the value function: for $\gamma > 0$,

$$v(\gamma x, \gamma y) = \gamma^{1-p}\, v(x,y), \quad (x,y) \in \mathcal{S}. \tag{2.8}$$

This is because $(C, L, M) \in \mathcal{A}(x, y) \Leftrightarrow (\gamma C, \gamma L, \gamma M) \in \mathcal{A}(\gamma x, \gamma y)$. Consequently, the problem reduces to that of a single variable. With $\mathcal{T} \triangleq (-1/\lambda, 1/\lambda)$, we define

$$u(z) = v(1 - z, z), \qquad z \in \mathcal{T}. \tag{2.9}$$

In other words, we make the change of variables $z = y/(x + y)$, $1 - z = x/(x + y)$, which maps the solvency region $\mathcal{S}$ onto the interval $\mathcal{T}$. Then

$$v(x, y) = (x + y)^{1-p}\, u\left(\frac{y}{x + y}\right), \quad (x, y) \in \mathcal{S}. \tag{2.10}$$

The HJB equation corresponding to (2.7) for the function $u(z)$ is

$$\min\Big\{\mathcal{D}u(z) - \widetilde{U}\big((1 - p)\, u(z) - z u'(z)\big), \lambda(1 - p)u(z) + (1 - \lambda z)u'(z), \\ \lambda(1 - p)\, u(z) - (1 + \lambda z)\, u'(z)\Big\} = 0, \tag{2.11}$$

where (see [43] p. 681, substituting λ for μ, $1 + \lambda$ for $1/(1 - \lambda)$, $1 - p$ for p)

$$\mathcal{D}u(z) = \Big(pA(p) + \frac{1}{2}\sigma^2 p(1 - p)(z - \theta_p)^2\Big)\, u(z) \\ + p\sigma^2 z(1 - z)(z - \theta_p)\, u'(z) - \frac{1}{2}\sigma^2 z^2 (1 - z)^2\, u''(z).$$

Because $v(x, y)$ is continuous on $\overline{\mathcal{S}}$ and of class C^2 in $\mathcal{S} \setminus \{(x, y); x = 0\}$ ([43], Corollary 10.2 and Theorem 11.6), the function u is continuous on $\overline{\mathcal{T}}$, C^2 on $\mathcal{T}$ except possibly at $z = 1$, and satisfies (2.11). Moreover, [43] shows

$$\lambda(1 - p)u(z) - (1 + \lambda z)u'(z) \geqslant 0, \quad -\frac{1}{\lambda} < z \leq z_1, \tag{2.12}$$

$$\mathcal{D}u(z) - \widetilde{U}\big((1 - p)u(z) - z u'(z)\big) \leqslant 0, \quad z_1 \leq z \leq z_2, \tag{2.13}$$

$$\lambda(1 - p)u(z) + (1 - \lambda z)u'(z) \geqslant 0, \quad z_2 \leq z < \frac{1}{\lambda}. \tag{2.14}$$

Since the function u is C^2 except possibly at $z = 1$, at each of $z_1 \neq 1$ and $z_2 \neq 1$ two of the above three equations hold. Moreover, v_{yy} is continuous at $x = 0$, and $v(0, y)$ satisfies (2.7) ([43], Corollary 10.3). Furthermore, $(1 - z)^2 u''(z)$ is continuous for $z \in \mathcal{T}$, and $\lim_{z\to 1}(z - 1)^2 u''(z) = 0$ ([43], (A.5)). Thus, two of the above three equations hold at each z_i if $z_i = 1$.

Equations (2.12), (2.14) are consequences of the directional derivative of $v(x, y)$ being zero in the directions of transaction in the regions in which it is optimal to buy and sell stock, respectively. These equations imply

$$u(z) = u(z_1)\left(\frac{1 + \lambda z}{1 + \lambda z_1}\right)^{1-p}, \quad -\frac{1}{\lambda} < z \leq z_1, \tag{2.15}$$

$$u(z) = u(z_2)\left(\frac{1 - \lambda z}{1 - \lambda z_2}\right)^{1-p}, \quad z_2 \leq z < \frac{1}{\lambda}. \tag{2.16}$$

Throughout this paper, we denote by $O(x)$, respectively $o(x)$, any quantity which is bounded by a positive constant times x, respectively lower than any positive constant times x, in a neighborhood of $x = 0$.

Lemma 1 *Assume $p > 0$, $p \neq 1$. For $a > 0$ and $b < a$ we have*

$$\begin{aligned}\widetilde{U}(a-b) &= \frac{p}{1-p}\,(a-b)\left(a^{-\frac{1}{p}} + \frac{b}{p}\,a^{-\frac{1+p}{p}} + O\big(b^2\big)\right)\\ &= \frac{p}{1-p}\,a^{-\frac{1-p}{p}} + b\,a^{-\frac{1}{p}} + O\big(b^2\big).\end{aligned} \tag{2.17}$$

Proof We write

$$\widetilde{U}(a-b) = \frac{p}{1-p}\,\big(a-b\big)^{-\frac{1-p}{p}} = \frac{p}{1-p}(a-b)\,(a-b)^{-\frac{1}{p}}.$$

A Taylor series expansion of the function $f(x) = (a-x)^{-1/p}$ yields $(a-b)^{-1/p} = a^{-1/p} + \frac{1}{p}\,b\,a^{-(1+p)/p} + O\big(b^2\big)$, and we get the desired result.

3 Heuristic derivation by Taylor series

In this section we heuristically derive several terms of a power series expansion of the value function. One can get an idea on the size of the NT wedge by the following argument. When transaction costs are introduced, it is too expensive for an agent to keep the proportion of capital in stock equal to θ_p. Suppose the agent decides to instead keep the proportion inside an interval centered at θ_p having width w. She then incurs an associated cost of transaction which is the product $\lambda\ell$ of the transaction cost λ and the amount of transacting (local time) ℓ accumulated by the state process at the boundaries of the NT wedge. Suppose now that the no-transaction interval has width ρw for some $\rho > 0$. One could multiply the stock volatility by ρ, which is equivalent to scaling the Brownian motion by ρ, and then the local time on the boundary of the NT wedge would also scale by ρ. If one subsequently scales time by $1/\rho^2$, local time is also scaled by $1/\rho^2$, and we have returned to the original volatility. The net effect of these two scalings is to scale local time by $1/\rho$. In other words, the amount of transacting is inversely proportional to the width of the NT wedge.

On the other hand, by permitting the state process to lie in a wedge rather than at the optimal proportion θ_p, the agent loses utility due to displacement from the optimal proportion. This is proportional to the square of the displacement. To see that, one can consider the problem with zero transaction cost and wealth process $X_t + Y_t$ given by

$$\begin{aligned}\mathrm{d}(X_t+Y_t) = {}& r(X_t+Y_t)\,\mathrm{d}t + (\alpha-r)\theta(X_t+Y_t)\,\mathrm{d}t - c(X_t+Y_t)\,\mathrm{d}t\\ &+ \sigma\theta(X_t+Y_t)\,\mathrm{d}W(t),\end{aligned}$$

where θ is a constant proportion of wealth maintained in the stock at all times and c is a constant fraction of wealth being consumed at all times. It is convenient to take θ to be of the form $\theta_p + \epsilon$ and c to be of the form $(1+\delta)A(p)(X_t+Y_t)$, so that $\epsilon = 0$ and $\delta = 0$ provide the optimal solution to the zero transaction cost problem. One can then compute

$$\mathbb{E}(X_t+Y_t)^{1-p} = (X_0+Y_0)^{1-p}\exp\left\{\,\big(\beta - A(p) - B(\epsilon) - (1-p)A(p)\delta\big)\,t\,\right\},$$

where $B(\epsilon) = \frac{1}{2}p(1-p)\sigma^2\epsilon^2$. This yields expected utility

$$\begin{aligned}
&\frac{1}{1-p}\mathbb{E}\int_0^\infty e^{-\beta t}\big((1+\delta)A(p)(X_t+Y_t)\big)^{1-p}\,\mathrm{d}t \\
&= \frac{1}{1-p}(X_0+Y_0)^{1-p}A^{-p}(p)(1+\delta)^{1-p}\left(1+\frac{B(\epsilon)}{A(p)}+(1-p)\delta\right)^{-1}.
\end{aligned}$$

For fixed ϵ, this is maximized by taking $\delta = B\big(\epsilon\big)/\big(pA(p)\big)$, and that value of δ results in expected utility

$$\begin{aligned}
&\frac{1}{1-p}\mathbb{E}\int_0^\infty e^{-\beta t}\big((1+\delta)A(p)(X_t+Y_t)\big)^{1-p}\,\mathrm{d}t \\
&= \frac{1}{1-p}(X_0+Y_0)^{1-p}A^{-p}(p)\left(1+\frac{1}{2}(1-p)\sigma^2A^{-1}(p)\epsilon^2\right)^{-p} \qquad (3.1)\\
&= \frac{1}{1-p}(X_0+Y_0)^{1-p}A^{-p}(p) - \frac{p}{2}(X_0+Y_0)^{1-p}\sigma^2A^{-1-p}(p)\epsilon^2 + O(\epsilon^4).
\end{aligned}$$

The first term above is the value function when there is zero transaction costs. The order ϵ^2 term is the lowest order loss due to displacement.

Suppose an agent faced with transaction costs chooses a no-transaction wedge whose width is order λ^q for some $q > 0$. The amount of transacting will be of order λ^{-q} and the marginal loss due to transacting will be of order λ^{1-q}. On the other hand, the marginal loss due to displacement will be of order λ^{2q}. The agent chooses q to balance these marginal losses, i.e., chooses $q = \frac{1}{3}$ so that the width of the NT wedge is $\lambda^{1/3}$ and the loss in the value function due to the presence of transaction cost λ is of order $\lambda^{2/3}$. In Remark 4.9 we show how to carry this probabilistic discussion farther to identify the coefficients multiplying the terms $\lambda^{\frac{1}{3}}$ in the NT width and $\lambda^{\frac{2}{3}}$ in the value funciton loss. In this section, we identify these coefficients by an analysis based on asymptotic expansion.

There is no explicit solution to (2.13) in the interval $[z_1, z_2]$. Guided by the above discussion, we thus assume that in this region $u(z)$ has an expansion in powers of $\lambda^{1/3}$, and we expect the coefficient of $\lambda^{1/3}$ to be zero. In order to work with this expansion, we need to also include the variable z, and we do that using powers of $z - \theta_p$:

$$\begin{aligned}
u(z) = \gamma_0 &- \gamma_1\,\lambda^{\frac{1}{3}} - \gamma_2\,\lambda^{\frac{2}{3}} - \gamma_3\,\lambda - \gamma_{40}\,\lambda^{\frac{4}{3}} - \gamma_{41}\,(z-\theta_p)\lambda \\
&- \gamma_{42}\,(z-\theta_p)^2\lambda^{\frac{2}{3}} - \gamma_{43}\,(z-\theta_p)^3\lambda^{\frac{1}{3}} - \gamma_{44}\,(z-\theta_p)^4 + O\big(\lambda^{\frac{5}{3}}\big).
\end{aligned} \qquad (3.2)$$

We argued above that $z - \theta_p = O\big(\lambda^{1/3}\big)$ for $z \in [z_1, z_2]$. Because we can always move from $z \in [z_1, z_2]$ to θ_p by paying a transaction cost of $\lambda\,|z-\theta_p| = O\big(\lambda^{4/3}\big)$, we must have $|u(z) - u(\theta_p)| = O\big(\lambda^{4/3}\big)$, and so the larger terms $(z-\theta_p)$, $(z-\theta_p)^2$, $(z-\theta_p)\lambda^{1/3}$, $(z-\theta_p)^3$, $(z-\theta_p)^2\lambda^{1/3}$, $(z-\theta_p)\lambda^{2/3}$ do not appear in (3.2).

Using $z-\theta_p = O\big(\lambda^{1/3}\big)$ together with (3.2) leads to the expansions

$$\begin{aligned} u'(z) =& -\gamma_{41}\,\lambda - 2\gamma_{42}\,(z-\theta_p)\lambda^{\frac{2}{3}} - 3\gamma_{43}\,(z-\theta_p)^2\lambda^{\frac{1}{3}} \\ & -4\gamma_{44}\,(z-\theta_p)^3 + O\big(\lambda^{\frac{4}{3}}\big), \qquad (3.3) \\ u''(z) =& -2\gamma_{42}\,\lambda^{\frac{2}{3}} - 6\gamma_{43}\,(z-\theta_p)\lambda^{\frac{1}{3}} - 12\gamma_{44}\,(z-\theta_p)^2 + O(\lambda). \qquad (3.4) \end{aligned}$$

For $z_1 \le z \le z_2$ we have

$$\begin{aligned} \mathcal{D}u(z) =& pA(p)\gamma_0 - pA(p)\gamma_1\,\lambda^{\frac{1}{3}} - pA(p)\gamma_2\,\lambda^{\frac{2}{3}} + \frac{1}{2}\sigma^2 p(1-p)(z-\theta_p)^2\gamma_0 \qquad (3.5) \\ & +\sigma^2 z^2(1-z)^2\left(\gamma_{42}\,\lambda^{\frac{2}{3}} + 3\gamma_{43}\,(z-\theta_p)\lambda^{\frac{1}{3}} + 6\gamma_{44}\,(z-\theta_p)^2\right) + O(\lambda). \end{aligned}$$

Furthermore,

$$(1-p)\,u(z) - z\,u'(z) = (1-p)\gamma_0 - (1-p)\gamma_1\lambda^{\frac{1}{3}} - (1-p)\gamma_2\lambda^{\frac{2}{3}} + O(\lambda). \qquad (3.6)$$

Setting $a = (1-p)\gamma_0$ and $b = (1-p)\gamma_1\,\lambda^{1/3} + O\big(\lambda^{2/3}\big)$ in Lemma 1, we obtain

$$\begin{aligned} \widetilde{U}\big((1-p)\,u(z) - z\,u'(z)\big) =& \frac{p}{1-p}\big((1-p)\gamma_0\big)^{-\frac{1-p}{p}} \\ & + (1-p)\gamma_1\big((1-p)\gamma_0\big)^{-\frac{1}{p}}\,\lambda^{\frac{1}{3}} + O\big(\lambda^{\frac{2}{3}}\big). \end{aligned} \qquad (3.7)$$

Equating first the $O(1)$ terms and then the $O\big(\lambda^{1/3}\big)$ terms in (3.5) and (3.7) (see (2.13)), we conclude that

$$\gamma_0 = \frac{1}{1-p}A^{-p}(p), \qquad \gamma_1 = 0.$$

Observe that γ_0 is the value $v(1-z,z)$ for zero transaction costs.

Since γ_1 is zero, we can now set $b = (1-p)\gamma_2\,\lambda^{2/3} + O(\lambda)$ in Lemma 1, and obtain (after substituting for γ_0)

$$\widetilde{U}\big((1-p)\,u(z) - z\,u'(z)\big) = \frac{p}{1-p}A^{1-p}(p) + (1-p)\gamma_2 A(p)\,\lambda^{\frac{2}{3}} + O(\lambda). \qquad (3.8)$$

From (2.15) and (2.16) we observe that $u''(z_1) = O\big(\lambda^2\big)$ and $u''(z_2) = O\big(\lambda^2\big)$. From the continuity of $(1-z)^2u''(z)$ we get from (3.5)

$$\mathcal{D}u(z_i) = \frac{p}{1-p}A^{1-p}(p) - pA(p)\gamma_2\,\lambda^{\frac{2}{3}} + \frac{1}{2}\sigma^2 p(z_i-\theta_p)^2A^{-p}(p) + O(\lambda), \quad i = 1,2,$$

where we have omitted the terms in (3.5) arising from $u''(z_i) = O\big(\lambda^2\big)$. Setting this equal to (3.8) at z_i results in $(z_i-\theta_p)^2 = \frac{1}{4}\nu^2\,\lambda^{\frac{2}{3}} + O(\lambda)$, with

$$\nu = \sqrt{\frac{8}{p\sigma^2}A^{1+p}(p)\gamma_2}. \qquad (3.9)$$

We thus have

$$z_1 - \theta_p = -\frac{1}{2}\nu\,\lambda^{\frac{1}{3}} + O\big(\lambda^{\frac{2}{3}}\big), \quad z_2 - \theta_p = \frac{1}{2}\nu\,\lambda^{\frac{1}{3}} + O\big(\lambda^{\frac{2}{3}}\big). \tag{3.10}$$

We may also equate the $\lambda^{2/3}$ terms in (3.5) and (3.8) at $z = \theta_p$, $z = z_1$, and $z = z_2$, and substitute for $z_i - \theta_p$ from (3.10), to obtain

$$\begin{aligned}
A(p)\gamma_2 &= \sigma^2\,\theta_p^2(1-\theta_p)^2\gamma_{42},\\
A(p)\gamma_2 &= \frac{1}{8}\sigma^2 pA^{-p}(p)\nu^2 + \sigma^2\,\theta_p^2(1-\theta_p)^2\left(\gamma_{42} - \frac{3}{2}\gamma_{43}\nu + \frac{3}{2}\gamma_{44}\nu^2\right),\\
A(p)\gamma_2 &= \frac{1}{8}\sigma^2 pA^{-p}(p)\nu^2 + \sigma^2\,\theta_p^2(1-\theta_p)^2\left(\gamma_{42} + \frac{3}{2}\gamma_{43}\nu + \frac{3}{2}\gamma_{44}\nu^2\right),
\end{aligned}$$

which implies

$$\gamma_{42} = \frac{A(p)\gamma_2}{\sigma^2\,\theta_p^2(1-\theta_p)^2}, \quad \gamma_{43} = 0, \quad \gamma_{44} = -\frac{pA^{-p}(p)}{12\theta_p^2(1-\theta_p)^2}. \tag{3.11}$$

Finally, we observe from (2.15) and from (3.2) that

$$u'(z_1) = \frac{\lambda(1-p)}{1+\lambda z_1}u(z_1) = \lambda(1-p)u(z_1) + O\big(\lambda^2\big) = A^{-p}(p)\,\lambda + O\big(\lambda^{\frac{5}{3}}\big),$$

and similarly for $u'(z_2)$. On the other hand, (3.3) and (3.10) imply that

$$\begin{aligned}
u'(z_1) &= -\gamma_{41}\,\lambda + \gamma_{42}\nu\,\lambda - \frac{3}{4}\gamma_{43}\nu^2\,\lambda + \frac{1}{2}\gamma_{44}\nu^3\,\lambda + O\big(\lambda^{\frac{4}{3}}\big),\\
u'(z_2) &= \gamma_{41}\,\lambda + \gamma_{42}\nu\,\lambda + \frac{3}{4}\gamma_{43}\nu^2\,\lambda + \frac{1}{2}\gamma_{44}\nu^3\,\lambda + O\big(\lambda^{\frac{4}{3}}\big).
\end{aligned}$$

It follows that

$$-\gamma_{41} + \gamma_{42}\nu - \frac{3}{4}\gamma_{43}\nu^2 + \frac{1}{2}\gamma_{44}\nu^3 = A^{-p}(p) = \gamma_{41} + \gamma_{42}\nu + \frac{3}{4}\gamma_{43}\nu^2 + \frac{1}{2}\gamma_{44}\nu^3.$$

Since $\gamma_{43} = 0$ (see (3.11)), we conclude

$$\gamma_{42}\nu + \frac{1}{2}\gamma_{44}\nu^3 = A^{-p}(p), \quad \gamma_{41} = 0. \tag{3.12}$$

We have thus written the value function as

$$u(z) = \frac{1}{1-p}A^{-p}(p) - \gamma_2\,\lambda^{\frac{2}{3}} + O(\lambda).$$

Furthermore, we can solve for γ_2 and the width of the NT interval ν from equations (3.9), (3.11), and (3.12) to obtain

$$\gamma_2 = \left(\frac{9}{32}p\theta_p^4(1-\theta_p)^4\right)^{\frac{1}{3}} A^{-1-p}(p)\sigma^2, \quad \nu = \left(\frac{12}{p}\theta_p^2(1-\theta_p)^2\right)^{\frac{1}{3}}. \tag{3.13}$$

4 Rigorous asymptotic expansion

Definition 1 *Let $w\colon \mathcal{T} \to \mathbb{R}$ be continuous, taking values in $(0,\infty)$ if $p \in (0,1)$ and in $(-\infty,0)$ if $p > 1$. Assume $0 < \zeta_1 < \zeta_2 < 1/\lambda$ satisfy*

$$w(z) = w(\zeta_1)\left(\frac{1+\lambda z}{1+\lambda\zeta_1}\right)^{1-p}, \quad -\frac{1}{\lambda} < z \le \zeta_1, \tag{4.1}$$

$$w(z) = w(\zeta_2)\left(\frac{1-\lambda z}{1-\lambda\zeta_2}\right)^{1-p}, \quad \zeta_2 \le z < \frac{1}{\lambda}. \tag{4.2}$$

Assume further that w is of class C^2 on (ζ_1,ζ_2) and the one-sided limits of w'' exist at the endpoints of this interval. If

$$\lambda(1-p)w(z) + (1-\lambda z)w'(z) \ge 0, \quad -\frac{1}{\lambda} < z < \zeta_2, \tag{4.3}$$

$$\mathcal{D}w(z) - \widetilde{U}\big((1-p)w(z) - zw'(z)\big) \ge 0, \quad -\frac{1}{\lambda} < z < \frac{1}{\lambda}, \tag{4.4}$$

$$\lambda(1-p)w(z) - (1+\lambda z)w'(z) \ge 0, \quad \zeta_1 < z < \frac{1}{\lambda}, \tag{4.5}$$

we say (w,ζ_1,ζ_2) is a supersolution triple.[2] *If (4.1), (4.2) and the reverse of (4.4) for $\zeta_1 < z < \zeta_2$ hold, we say (w,ζ_1,ζ_2) is a* subsolution triple.

If (w,ζ_1,ζ_2) is a supersolution (subsolution) triple, then

$$\varphi(x,y) \triangleq (x+y)^{1-p} w\big(y/(x+y)\big), \quad (x,y) \in \mathcal{S} \tag{4.6}$$

is a supersolution (subsolution) of (2.7).

Lemma 2 *Assume $p > 0$, $p \ne 1$. If (w,ζ_1,ζ_2) is a supersolution (subsolution) triple, then $w(z) \ge u(z)$ ($w(z) \le u(z)$) for $z \in \mathcal{T}$.*

Proof For $0 < p < 1$, this follows from [43], Comparison Theorem A.3. We give the proof for $p > 1$.

Let (w,ζ_1,ζ_2) be a supersolution triple, and let φ be given by (4.6). In light of (2.10), is suffices to prove that $\varphi(x_0,y_0) \ge v(x_0,y_0)$ for fixed but arbitrary $(x_0,y_0) \in \mathcal{S}$. Following the construction of Theorem 9.2 of [43], we let (C,L,M) be an optimal policy for this initial condition, i.e., with (X_t,Y_t) given by (2.1), (2.2), we have $(X(0),Y(0)) \in \partial NT$ if $(x_0,y_0) \notin NT$, $(X_t,Y_t) \in \overline{NT}$ for all $t \ge 0$, and

$$L_t = \int_0^t \mathbb{I}_{\{Y_s/(X_s+Y_s)=z_1\}}\,\mathrm{d}L_s, \quad M_t = \int_0^t \mathbb{I}_{\{Y_s/(X_s+Y_s)=z_2\}}\mathrm{d}M_s,$$

$$C_t = \big(v_x(X_t,Y_t)\big)^{-1/p}, \quad t \ge 0.$$

Then $v(x_0,y_0) = \mathbb{E}\int_0^\infty e^{-\beta t}U_p(C_t)\,\mathrm{d}t$.

[2] For functions of the form (4.1), (4.2), the validity of (4.3) for $-1/\lambda < z < \zeta_2$ implies its validity for all $z \in (-1/\lambda, 1/\lambda)$, and similarly for (4.5).

The function φ is C^2 in $\mathcal{S}$ except possibly on the lines $y/(x+y) = \zeta_i$, $i = 1, 2$. We can mollify φ to obtain a C^2 function, apply Itô's rule to the mollified function, and pass to the limit. On the two lines where φ may not be C^2, (X_t, Y_t) spends Lebesgue-measure zero time. We thus obtain Itô's rule for $\varphi(X_t, Y_t)$. An exception would be if $\zeta_2 = z_2 = 1$, because in this case (X_t, Y_t) remains on the $y-$axis once the y-axis is reached. A direct computation shows, however, that $\varphi_{yy}(x, y)$ is continuous there, even if $\zeta_2 = z_2 = 1$, and hence we will still obtain Itô's rule. Therefore,

$$\begin{aligned}
\mathrm{d}\left(\mathrm{e}^{-\beta t}\,\varphi(X_t, Y_t)\right) &= -\,\mathrm{e}^{-\beta t}\Big[\mathcal{L}\varphi(X_t, Y_t)\;\mathrm{d}t + C_t\varphi_x(X_t, Y_t) \\
&\quad + \big((1+\lambda)\varphi_x(X_t, Y_t) - \varphi_y(X_t, Y_t)\big)\,\mathrm{d}L_t + \big(-(1-\lambda)\varphi_x(X_t, Y_t) \\
&\quad + \varphi_y(X_t, Y_t)\big)\,\mathrm{d}M_t\Big] + \mathrm{e}^{-\beta t}\,\varphi_y(X_t, Y_t)\,\sigma\,\mathrm{d}W_t \\
&\le -\,\mathrm{e}^{-\beta t}\,U_p(C_t)\;\mathrm{d}t + \mathrm{e}^{-\beta t}\,\varphi_y(X_t, Y_t)\,\sigma\,\mathrm{d}W_t,
\end{aligned} \tag{4.7}$$

where we have used the supersolution property, the fact that $\mathcal{L}\varphi = \mathcal{D}u$, and the fact (see(2.6)) that

$$\widetilde{U}_p\big(\varphi_x(X_t, Y_t)\big) \ge U_p(C_t) - C_t\varphi_x(X_t, Y_t). \tag{4.8}$$

We want to integrate (4.7) and eventually argue that the expected value of the Itô integral is zero. Define $\tau_n = \inf\{t \ge 0; |X_t + Y_t| \le 1/n\}$, to obtain

$$\begin{aligned}
&\int_0^{n\wedge\tau_n} e^{-\beta t}U_p(C_t)\;\mathrm{d}t + e^{-\beta(n\wedge\tau_n)}\varphi(X_{n\wedge\tau_n}, Y_{n\wedge\tau_n}) \\
&\qquad\qquad \le \varphi(x_0, y_0) + \sigma\int_0^{n\wedge\tau_n} e^{-\beta t}\varphi_y(X_t, Y_t)\,\mathrm{d}W_t.
\end{aligned} \tag{4.9}$$

We can perform a similar analysis for the value function. The function v is C^2 except on the y-axis. The x-axis is not in $\overline{NT}$, which contains (X_t, Y_t) for all $t \ge 0$. At the y-axis, v_{yy} is continuous ([43], Corollary 10.3). This permits us to compute the differential of $v(X_t, Y_t)$, and we obtain

$$\begin{aligned}
&\int_0^{n\wedge\tau_n} e^{-\beta t}U_p(C_t)\;\mathrm{d}t + e^{-\beta(n\wedge\tau_n)}v(X_{n\wedge\tau_n}, Y_{n\wedge\tau_n}) \\
&\qquad\qquad = v(x_0, y_0) + \sigma\int_0^{n\wedge\tau_n} e^{-\beta t}v_y(X_t, Y_t)\,\mathrm{d}W_t.
\end{aligned} \tag{4.10}$$

As $n \to \infty$, we have either $\tau_n \to \infty$ or $\tau_n \to \tau_0 \triangleq \inf\{t \ge 0 : X_t = Y_t = 0\} < \infty$. On the set $\{\lim_{n\to\infty}\tau_n = \tau_0 < \infty\}$, equation (4.10), the inequality $U_p \le 0$ and the fact that $v(0, 0) = -\infty$ imply $\int_0^{\tau_0} e^{-\beta t}v_y(X_t, Y_t)\,\mathrm{d}W_t = -\infty$. This is impossible; Itô integrals are either finite or have $\limsup = -\liminf = \infty$ (see [30], Chap. 3, Problem 4.11 and p. 232). We conclude that $\tau_n \to \infty$ almost surely. Taking expectations in (4.10) and letting $n \to \infty$, we obtain

$$\lim_{n\to\infty}\mathbb{E}e^{-\beta(n\wedge\tau_n)}v(X_{n\wedge\tau_n}, Y_{n\wedge\tau_n}) = v(x_0, y_0) - \mathbb{E}\int_0^{\infty} e^{-\beta t}U_p(C_t)\;\mathrm{d}t = 0. \tag{4.11}$$

Because of (2.10), (4.6) and the boundedness of u and w for $z \in NT$, either $\varphi(x,y) = 0$, or there are constants c_1 and c_2 such that

$$c_1\varphi(x,y) \le v(x,y) \le c_2\varphi(x,y) \text{ whenever } z_1 \le \frac{y}{x+y} \le z_2.$$

From (4.11) we conclude that $\lim_{n\to\infty} \mathbb{E}e^{-\beta(n\wedge\tau_n)}\varphi(X_{n\wedge\tau_n}, Y_{n\wedge\tau_n}) = 0$. Taking expectations and the limit in (4.9), we obtain

$$v(x_0,y_0) = \mathbb{E}\int_0^\infty e^{-\beta t}U_p(C_t)\,\mathrm{d}t \le \varphi(x_0,y_0).$$

Now let (w,ζ_1,ζ_2) be a subsolution triple. We show for fixed but arbitrary $(x_0,y_0) \in \mathcal{S}$ that $\varphi(x_0,y_0) \le v(x_0,y_0)$. This time we construct a (suboptimal) policy (C,L,M) for which $Y(0)/\big(X(0)+Y(0)\big)$ is either ζ_1 or ζ_2 if $y_0/(x_0+y_0) \notin (\zeta_1,\zeta_2)$, $(X_t,Y_t) \in [\zeta_1,\zeta_2]$ for all $t \ge 0$, and

$$L_t = \int_0^t \mathbb{I}_{\{Y_s/(X_s+Y_s)=\zeta_1\}}\,dL_s, \quad M_t = \int_0^t \mathbb{I}_{\{Y_s/(X_s+Y_s)=\zeta_2\}}\,dM_s,$$

$$C_t = \big(\varphi_x(X_t,Y_t)\big)^{-1/p}, \quad t \ge 0.$$

For this policy, we have the reverse equality in (4.9).[3] Using this inequality in place of (4.10), we argue as before that $\lim_{n\to\infty}\tau_n = \infty$ almost surely. $\varphi \le 0$, (4.9) with the inequality reversed implies $\mathbb{E}\int_0^{n\wedge\tau_n} e^{-\beta t}U_p(C_t)\,\mathrm{d}t \ge \varphi(x_0,y_0)$ and letting $n\to\infty$, we obtain $v(x_0,y_0) \ge \mathbb{E}\int_0^\infty e^{-\beta t}U_p(C_t)\,\mathrm{d}t \ge \varphi(x_0,y_0)$.

Theorem 1 *Assume $p > 0$, $p \ne 1$ and $A(p) > 0$. Then the value function u satisfies*

$$u(\theta_p) = \frac{1}{1-p}A^{-p}(p) - \left(\frac{9}{32}p\,\theta_p^4(1-\theta_p)^4\right)^{\frac{1}{3}} A^{-1-p}(p)\,\sigma^2\,\lambda^{\frac{2}{3}} + O(\lambda). \quad (4.12)$$

Proof We assume $\theta_p \ne 1$. For $\theta_p = 1$ see Remark 1.

Step 1: Choice of constants and variables

We recall the constants γ_2 and ν of (3.13). Set $\xi = \sqrt{\frac{2}{3}(1-p)\gamma_2 A^p(p) + B}$, where B is a constant chosen to make the expression under the square root positive. We next define

$$h(\delta) = \frac{3}{2}\,\delta^2\lambda^{\frac{2}{3}} - \frac{1}{\nu^2}\cdot\delta^4 + \frac{3}{2}B\,\delta^2\lambda^{\frac{2}{3}}$$

[3] We used the optimal policy for the supersolution argument to get equality in (4.10). For the subsolution argument we get the appropriate inequality in (4.10) by using a suboptimal policy. However, we need to pick the trading policy so that the terms containing the integrals $\mathrm{d}L_t$ and $\mathrm{d}M_t$ in (4.7) are zero, and the consumption policy so that we get equality in (4.8).

and choose a positive constant M satisfying

$$M > \theta_p A^{-p}(p) + \frac{1}{4}\sigma^2 p\nu^2\xi^2 A^{-p-1}(p). \tag{4.13}$$

We define functions

$$\begin{aligned} f_1^{\pm}(\delta) &= \nu\lambda^{\frac{1}{3}} - (1-p)\nu\gamma_2 A^p(p)\,\lambda \pm (1-p)\nu M A^p(p)\,\lambda^{\frac{4}{3}} - (1-p)h(-\delta)\,\lambda^{\frac{1}{3}} \\ &\quad + \Big(\lambda^{-\frac{2}{3}} + (\theta_p - \delta)\lambda^{\frac{1}{3}}\Big)h'(-\delta), \end{aligned} \tag{4.14}$$

$$\begin{aligned} f_2^{\pm}(\delta) &= \nu\,\lambda^{\frac{1}{3}} - (1-p)\nu\gamma_2 A^p(p)\,\lambda \pm (1-p)\nu M A^p(p)\,\lambda^{\frac{4}{3}} - (1-p)h(\delta)\,\lambda^{\frac{1}{3}} \\ &\quad + \Big(-\lambda^{-\frac{2}{3}} + (\theta_p + \delta)\lambda^{\frac{1}{3}}\Big)h'(\delta). \end{aligned} \tag{4.15}$$

It is shown in Appendix A that there are numbers

$$\delta_1^{\pm} = \frac{1}{2}\nu\lambda^{\frac{1}{3}}(1 - \xi\lambda^{\frac{1}{3}}) + o\big(\lambda^{\frac{2}{3}}\big), \quad \delta_2^{\pm} = \frac{1}{2}\nu\lambda^{\frac{1}{3}}(1 - \xi\lambda^{\frac{1}{3}}) + o\big(\lambda^{\frac{2}{3}}\big) \tag{4.16}$$

satisfying $f_i^{\pm}(\delta_i^{\pm}) = 0$, $i = 1, 2$.

Step 2: Construction of super/subsolutions.

Choose $\lambda > 0$ small enough that $\zeta_1^{\pm} \triangleq \theta_p - \delta_1^{\pm}$ and $\zeta_2^{\pm} \triangleq \theta_p + \delta_2^{\pm}$ all lie in $(0, 1/\lambda)$. (We have $\theta_p > 0$ since $\alpha > r$.) Define

$$w_{\pm}(z) = \begin{cases} \Big(\frac{A^{-p}(p)}{1-p} - \gamma_2\,\lambda^{\frac{2}{3}} \pm M\lambda - \frac{A^{-p}(p)}{\nu}h(\zeta_1^{\pm} - \theta_p)\Big)\Big(\frac{1+\lambda z}{1+\lambda\zeta_1^{\pm}}\Big)^{1-p}, \\ \frac{1}{1-p}A^{-p}(p) - \gamma_2\,\lambda^{\frac{2}{3}} \pm M\lambda - \frac{A^{-p}(p)}{\nu}h(z - \theta_p), \\ \Big(\frac{A^{-p}(p)}{1-p} - \gamma_2\,\lambda^{\frac{2}{3}} \pm M\lambda - \frac{A^{-p}(p)}{\nu}h(\zeta_2^{\pm} - \theta_p)\Big)\Big(\frac{1-\lambda z}{1-\lambda\zeta_2^{\pm}}\Big)^{1-p}, \end{cases}$$

for $-1/\lambda \le z \le \zeta_1^{\pm}$, $\zeta_1^{\pm} \le z \le \zeta_2^{\pm}$, $\zeta_2^{\pm} \le z \le 1/\lambda$ respectively. The reader can verify that if M were zero, then in the region $[\zeta_1^{\pm}, \zeta_2^{\pm}]$ the formula for $w_{\pm}(z)$ agrees with the power series expansion

$$\gamma_0 - \gamma_2\lambda^{\frac{2}{3}} - \gamma_{42}(z-\theta_p)^2\lambda^{\frac{2}{3}} - \gamma_{44}(z-\theta_p)^4,$$

where the coefficients γ_0, γ_2, γ_{42}, and γ_{44} are those worked out in the previous section. The term $\pm M\lambda$ in the definition of $w_{\pm}$ will be used to create supersolution and subsolution triples. We have the derivative formula

$$w'_{\pm}(z) = \begin{cases} \frac{\lambda}{1+\lambda z}(1-p)w_{\pm}(z), & -\frac{1}{\lambda} < z \le \zeta_1^{\pm}, \\ -\nu^{-1}A^{-p}(p)\,h'(z-\theta_p), & \zeta_1^{\pm} \le z \le \zeta_2^{\pm}, \\ -\frac{\lambda}{1-\lambda z}(1-p)w_{\pm}(z), & \zeta_2^{\pm} \le z < \frac{1}{\lambda}. \end{cases}$$

The equations $f_1^{\pm}(\delta_1) = 0$ and $f_2^{\pm}(\delta_2) = 0$ guarantee that $w'_{\pm}$ is defined and continuous at $\zeta_1^{\pm}$ and $\zeta_2^{\pm}$. We also have

$$w''_{\pm}(z) = \begin{cases} -\frac{\lambda^2}{(1+\lambda z)^2}\,p(1-p)w_{\pm}(z), & -\frac{1}{\lambda} < z < \zeta_1^{\pm}, \\ -\nu^{-1}A^{-p}(p)\,h''(z-\theta_p), & \zeta_1^{\pm} < z < \zeta_2^{\pm}, \\ -\frac{\lambda^2}{(1-\lambda z)^2}p(1-p)w_{\pm}(z), & \zeta_2^{\pm} < z < \frac{1}{\lambda}. \end{cases}$$

The function $w_\pm(z)$ is $\mathcal{C}^2$ except at $\zeta_1^\pm$ and $\zeta_2^\pm$, and at these points the one-sided second derivatives exist and equal the one-sided second derivative.

Step 3: Verification that $(w_-, \zeta_1^-, \zeta_2^-)$ is a subsolution triple.

It suffices to verify

$$\mathcal{D}w_-(z) - \widetilde{U}\big((1-p)w_-(z) - zw_-'(z)\big) \le 0, \quad \zeta_1^- < z < \zeta_2^-. \tag{4.17}$$

To do this, we simultaneously work with both w_- and w_+. We thereby develop an inequality for w_+ needed in the subsequent supersolution verification. We use the facts that $z - \theta_p = O\big(\lambda^{1/3}\big)$, so $h(z-\theta_p) = O\big(\lambda^{4/3}\big)$, $h'(z-\theta_p) = O(\lambda)$.

For $\zeta_1^\pm < z < \zeta_2^\pm$ we have $(1-p)w_\pm(z) - zw_\pm'(z) = a - b$, where $a = A^{-p}(p)$ and $b = (1-p)\gamma_2\lambda^{2/3} \mp (1-p)M\lambda - A^{-p}(p)\nu^{-1}zh'(z-\theta_p) + O\big(\lambda^{4/3}\big)$. Lemma 1 implies

$$\begin{aligned}
&\widetilde{U}\big((1-p)w_\pm(z) - zw_\pm'(z)\big) \qquad (4.18)\\
&= \frac{p}{1-p}A^{1-p}(p) + A(p)\left[(1-p)\gamma_2\lambda^{\frac{2}{3}} \mp (1-p)M\lambda - \frac{A^{-p}(p)}{\nu}zh'(z-\theta_p)\right]\\
&\qquad + O\big(\lambda^{\frac{4}{3}}\big)\\
&= pA(p)w(z) + \gamma_2 A(p)\lambda^{\frac{2}{3}} \mp M\lambda - \frac{A^{1-p}(p)}{\nu}zh'(z-\theta_p) + O\big(\lambda^{\frac{4}{3}}\big).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathcal{D}w_\pm(z) - \widetilde{U}\big((1-p)w_\pm(z) - zw_\pm'(z)\big)\\
&= \frac{1}{2}\sigma^2 p(1-p)(z-\theta_p)^2 w_\pm(z) - \frac{1}{2}\sigma^2 z^2(1-z)^2 w_\pm''(z)\\
&\qquad -\gamma_2 A(p)\lambda^{\frac{2}{3}} \pm M\lambda + \frac{1}{\nu}A^{1-p}(p)zh'(z-\theta_p) + O\big(\lambda^{\frac{4}{3}}\big)\\
&= \frac{1}{2}\sigma^2 pA^{-p}(p)(z-\theta_p)^2 + \frac{1}{2\nu}\sigma^2 z^2(1-z)^2 A^{-p}(p)h''(z-\theta_p)\\
&\qquad -\gamma_2 A(p)\lambda^{\frac{2}{3}} \pm M\lambda + \frac{1}{\nu}A^{1-p}(p)zh'(z-\theta_p) + O\big(\lambda^{\frac{4}{3}}\big).
\end{aligned}$$

Writing $z = \theta_p + (z-\theta_p)$ and $1 - z = 1 - \theta_p - (z-\theta_p)$, we derive the relation

$$z^2(1-z)^2 = \theta_p^2(1-\theta_p)^2 + 2\theta_p(1-\theta_p)(1-2\theta_p)(z-\theta_p) + O\big(\lambda^{\frac{2}{3}}\big).$$

Using this and the formulas $h'(\delta) = 3\delta\lambda^{2/3} - 4\delta^3/\nu^2 + 3B\delta\lambda^{2/3}$, $h''(\delta) = 3\lambda^{2/3} - 12\delta^2/\nu^2 + 3B\lambda^{2/3}$, we obtain

$$\begin{aligned}
&\mathcal{D}w_\pm(z) - \widetilde{U}\big((1-p)w_\pm(z) - zw_\pm'(z)\big)\\
&= \left[\frac{1}{2}\sigma^2 pA^{-p}(p) - \frac{6}{\nu^3}\sigma^2\theta_p^2(1-\theta_p)^2A^{-p}(p)\right](z-\theta_p)^2\\
&\qquad + \left[\frac{3}{2\nu}\sigma^2\theta_p^2(1-\theta_p)^2A^{-p}(p) - \gamma_2 A(p)\right]\lambda^{\frac{2}{3}}
\end{aligned}$$

$$+\sigma^2\theta_p(1-\theta_p)(1-2\theta_p)A^{-p}(p)\left[\frac{3(z-\theta_p)\lambda^{\frac{2}{3}}}{\nu}-\frac{12(z-\theta_p)^3}{\nu^3}\right]$$
$$+\theta_pA^{-p}(p)\left[\frac{3(z-\theta_p)\lambda^{\frac{2}{3}}}{\nu}-\frac{4(z-\theta_p)^3}{\nu^3}\right]\pm M\lambda+O\big(\lambda^{\frac{4}{3}}\big).$$

The definitions of ν and γ_2 imply that the first two terms on the right-hand side are zero. Because

$$\frac{z-\theta_p}{\nu}=\frac{1}{2}\lambda^{\frac{1}{3}}+O\big(\lambda^{\frac{2}{3}}\big),$$

the third term is $O\big(\lambda^{4/3}\big)$, and we can simplify the fourth term to obtain

$$\mathcal{D}w_\pm(z)-\widetilde{U}\big((1-p)w_\pm(z)-zw'_\pm(z)\big)=\theta_pA^{-p}(p)\lambda\pm M\lambda+O\big(\lambda^{\frac{4}{3}}\big)\quad(4.19)$$

for $z_1^\pm<z<z_2^\pm$. From (4.13) we have (4.17) for all sufficiently small $\lambda>0$. This completes the verification that $(w_-,\zeta_1^-,\zeta_2^-)$ is a subsolution triple.

Step 4: Verification that $(w_+,\zeta_1^+,\zeta_2^+)$ is a supersolution triple

Step 4a: Interval $(-1/\lambda,\,\zeta_1^+)$.

We must show that that (4.3) and (4.4) hold in this interval. Since $(1-p)w_+(z)>0$ for all small $\lambda>0$, so is $w'_+(z)$, and thus (4.3) holds.

It remains to verify that for sufficiently small λ

$$\mathcal{D}w_+(z)-\widetilde{U}\big((1-p)w_+(z)-zw'_+(z)\big)\ge 0,\quad -\frac{1}{\lambda}<z<\zeta_1^+.\qquad(4.20)$$

In this interval,

$$\begin{aligned}(1-p)w_+(z)-zw'_+(z)&=\frac{1-p}{1+\lambda z}w_+(z)\\&=\frac{1}{1+\lambda\zeta_1^+}\left(\frac{1+\lambda z_1^+}{1+\lambda z}\right)^p\Big(A^{-p}(p)-(1-p)\gamma_2\lambda^{\frac{2}{3}}\\&\qquad+(1-p)M\lambda-\frac{1}{\nu}(1-p)A^{-p}(p)h(\zeta_1^+-\theta_p)\Big).\end{aligned}$$

We use the equality $\widetilde{U}(\alpha\tilde{c})=\alpha^{-(1-p)/p}\,\widetilde{U}(\tilde{c})$ and the first equality in Lemma 1:

$$\begin{aligned}&\widetilde{U}\big((1-p)w_+(z)-zw'_+(z)\big)\\&=(1+\lambda z_1^+)^{\frac{1-p}{p}}\left(\frac{1+\lambda z}{1+\lambda\zeta_1^+}\right)^{1-p}\frac{p}{1-p}\\&\qquad\times\left(A^{-p}(p)-(1-p)\gamma_2\lambda^{\frac{2}{3}}+(1-p)M\lambda-\frac{1-p}{\nu}A^{-p}(p)h(\zeta_1^+-\theta_p)\right)\\&\qquad\times\left(A(p)+\frac{1-p}{p}\gamma_2A^{p+1}(p)\lambda^{\frac{2}{3}}-\frac{1-p}{p}MA^{p+1}(p)\lambda+O\big(\lambda^{\frac{4}{3}}\big)\right)\\&=(1+\lambda\zeta_1^+)^{\frac{1-p}{p}}w_+(z)\\&\qquad\times\left(pA(p)+(1-p)\gamma_2A^{p+1}(p)\lambda^{\frac{2}{3}}-(1-p)MA^{p+1}(p)\lambda+O\big(\lambda^{\frac{4}{3}}\big)\right).\end{aligned}$$

But $(1+\lambda\zeta_1^+)^{(1-p)/p} = 1+\frac{1-p}{p}\lambda\zeta_1^+ + O\big(\lambda^2\big) = 1+\frac{1-p}{p}\lambda\theta_p + O\big(\lambda^{4/3}\big)$; thus

$$\begin{aligned}
&\widetilde{U}\big((1-p)w_+(z) - zw_+'(z)\big)\\
&= w_+(z)\Big(pA(p) + (1-p)\gamma_2 A^{p+1}(p)\lambda^{\frac{2}{3}} - (1-p)MA^{p+1}(p)\lambda\\
&\qquad +(1-p)\theta_p A(p)\lambda + O\big(\lambda^{\frac{4}{3}}\big)\Big). \qquad (4.21)
\end{aligned}$$

It is easy to verify that for $\lambda > 0$ sufficiently small, the function $k(z) \triangleq (z-\theta_p) + \lambda z(1-z)/(1+\lambda z)$ attains its maximum over $(-1/\lambda, \zeta_1^+]$ at ζ_1^+ and $k(\zeta_1^+) < 0$. Therefore

$$k^2(z) \geq k^2(\zeta_1^+) = (\delta_1^+)^2 + \delta_1^+\, O(\lambda) = \frac{1}{4}\nu^2\lambda^{\frac{2}{3}} - \frac{1}{2}\nu^2\xi\lambda + o\big(\lambda\big),\ -\frac{1}{\lambda} < z \leq \zeta_1^+.$$

It follows that for sufficiently small $\lambda > 0$

$$\begin{aligned}
&\mathcal{D}w_+(z) - \widetilde{U}\big((1-p)w_+(z) - zw_+'(z)\big)\\
&= \left\{\frac{1}{2}p\,\sigma^2\Big((z-\theta_p) + z(1-z)\frac{\lambda}{1+\lambda z}\Big)^2 - \gamma_2 A^{p+1}(p)\,\lambda^{\frac{2}{3}} + MA^{p+1}(p)\lambda\right.\\
&\qquad \left. -\theta_p A(p)\lambda + O\big(\lambda^{\frac{4}{3}}\big)\right\}(1-p)w_+(z)\\
&\geq \left\{\frac{1}{2}p\,\sigma^2 k(\zeta_1^+) - \gamma_2 A^{p+1}(p)\,\lambda^{\frac{2}{3}} + MA^{p+1}(p)\lambda - \theta_p A(p)\lambda + O\big(\lambda^{\frac{4}{3}}\big)\right\}\\
&\qquad \times(1-p)w_+(z)\\
&= \left\{\left(\frac{1}{8}p\sigma^2\nu^2 - \gamma_2 A^{p+1}(p)\right)\lambda^{\frac{2}{3}} + \left(MA^{p+1}(p) - \frac{1}{4}p\sigma^2\nu^2\xi - \theta_p A(p)\right)\lambda\right.\\
&\qquad \left. +o(\lambda)\right\}(1-p)w_+(z)\\
&= \left\{\left(MA^{p+1}(p) - \frac{1}{4}p\sigma^2\nu^2\xi - \theta_p A(p)\right)\lambda + o(\lambda)\right\}(1-p)w_+(z)\ \geq\ 0,
\end{aligned}$$

where we have used (4.13) in the last step.

<u>*Step 4b:*</u> *Interval* $[\zeta_2^+,\, 1/\lambda)$. This is analogous to *Step 4a.*

<u>*Step 4c:*</u> *Interval* (ζ_1^+, ζ_2^+). From (4.19) and (4.13) we have

$$\mathcal{D}w_+(z) - \widetilde{U}\big((1-p)w_+(z) - zw_+(z)\big) = \big(\theta_p A^{-p}(p) + M\big)\,\lambda + O\big(\lambda^{\frac{4}{3}}\big) \geq 0.$$

We must also show that

$$g_1(z) \triangleq \lambda(1-p)w_+(z) - (1+\lambda z)w_+'(z) \geq 0, \quad \zeta_1^+ < z < \zeta_2^+. \qquad (4.22)$$

For $z \in (\zeta_1^+, \zeta_2^+)$, we have $z - \theta_p = O\big(\lambda^{1/3}\big)$. Using this fact, we compute

$$\begin{aligned} g_1(z) &= A^{-p}(p)\lambda - (1-p)\gamma_2\lambda^{\frac{5}{3}} + (1-p)M\lambda^2 \\ &\quad -\frac{3(1-p)}{2\nu}A^{-p}(p)(z-\theta_p)^2\lambda^{\frac{5}{3}} + \frac{1-p}{\nu^3}A^{-p}(p)(z-\theta_p)^4\lambda \\ &\quad +\frac{3(1+\lambda z)}{\nu}A^{-p}(p)(z-\theta_p)\lambda^{\frac{2}{3}} - \frac{4(1+\lambda z)}{\nu^3}A^{-p}(p)(z-\theta_p)^3, \\ g_1'(z) &= \frac{12}{\nu}A^{-p}(p)\lambda^{\frac{2}{3}}\left[\frac{1}{4} - \left(\frac{z-\theta_p}{\nu\lambda^{\frac{1}{3}}}\right)^2 + O(\lambda)\right]. \end{aligned}$$

We know that $g_1(\zeta_1^+) = 0$ and thus, to prove (4.22), it suffices to show that g_1' is positive on $[\zeta_1^+, \zeta_2^+]$. Because $-(z-\theta_p)^2$ is a concave function of z, it suffices to check the endpoints. We have for $i = 1, 2$ that

$$\left(\frac{\zeta_i^+ - \theta_p}{\nu\lambda^{\frac{1}{3}}}\right)^2 = \left(\frac{1}{2}(1 - \xi\lambda^{\frac{1}{3}}) + o\big(\lambda^{\frac{1}{3}}\big)\right)^2 = \frac{1}{4} - \frac{1}{2}\xi\lambda^{\frac{1}{3}} + o\big(\lambda^{\frac{1}{3}}\big).$$

Therefore,

$$g_1'(\zeta_i^+) = \frac{12}{\nu}A^{-p}(p)\lambda^{\frac{2}{3}}\left[\frac{1}{2}\xi\lambda^{\frac{1}{3}} + o\big(\lambda^{\frac{1}{3}}\big)\right] > 0$$

for sufficiently small $\lambda > 0$. The proof that

$$g_2(z) = \lambda(1-p)w_+(z) + (1-\lambda z)w_+'(z)$$

is positive for $z \in [\zeta_1^+, \zeta_2^+]$ is analogous. This completes the proof that $(w_+, \zeta_1^+, \zeta_2^+)$ is a supersolution triple.

Conclusion:

We note that $w_\pm(\theta_p) = \frac{1}{1-p}A^{-p}(p) - \gamma_2\,\lambda^{\frac{2}{3}} \pm M\lambda$, and so Lemma 2 implies

$$\frac{1}{1-p}A^{-p}(p) - \gamma_2\lambda^{\frac{2}{3}} - M\lambda \le u(\theta_p) \le \frac{1}{1-p}A^{-p}(p) - \gamma_2\lambda^{\frac{2}{3}} + M\lambda.$$

Corollary 1 *Assume $p > 0$, $p \neq 1$ and $A(p) > 0$. For fixed $z \in \mathcal{T}$, the value function satisfies*

$$u(z) = \frac{1}{1-p}A^{-p}(p) - \Big(\frac{9}{32}p\,\theta_p^4(1-\theta_p)^4\Big)^{\frac{1}{3}}A^{-1-p}(p)\,\sigma^2\,\lambda^{\frac{2}{3}} + O(\lambda). \quad (4.23)$$

Proof In the proof of Theorem 1 we constructed a supersolution w_+ and a subsolution w_- such that $w_+(z) - w_-(z) = O(\lambda)$, $w_\pm(z) = w_\pm(\theta) + O(\lambda)$ for fixed $z \in \mathcal{T}$. It follows that $u(z) = w_\pm(z) + O(\lambda) = w_\pm(\theta) + O(\lambda) = u(\theta) + O(\lambda)$.

Theorem 2 *Assume $p > 0$, $p \neq 1$, $A(p) > 0$ and $\theta_p \neq 1$. Then with ν given by (3.13), we have*

$$z_1 = \theta_p - \frac{1}{2}\nu\lambda^{\frac{1}{3}} + O\big(\lambda^{\frac{2}{3}}\big), \quad z_2 = \theta_p + \frac{1}{2}\nu\lambda^{\frac{1}{3}} + O\big(\lambda^{\frac{2}{3}}\big).$$

Proof The value function u is concave, so u' is monotone. By taking the derivative of u in the BS and SS regions (see (2.15) and (2.16)) we see that

$$u'(z) = O(\lambda) \quad \forall z \in [z_1, z_2]. \tag{4.24}$$

It follows that for $z = z_1$ and $z = z_2$, and hence for all $z \in [z_1, z_2]$,

$$\widetilde{U}\big((1-p)u(z) - zu'(z)\big) = \frac{p}{1-p}A^{1-p}(p) + (1-p)\gamma_2 A(p)\lambda^{\frac{2}{3}} + O(\lambda).$$

We also know that $u''(z)$ is continuous for $z \in \mathcal{T} \setminus \{1\}$. From equations (2.15), (2.16) it follows that $u''(z_i) = O\big(\lambda^2\big)$, $z_i \neq 1$. We thus have

$$\begin{aligned}\mathcal{D}u(z_i) &= \Big(pA(p) + \frac{1}{2}\sigma^2 p(1-p)(z_i - \theta_p)^2\Big)u(z_i) + O(\lambda) \\ &= \frac{p}{1-p}A^{1-p}(p) - p\gamma_2 A(p)\lambda^{\frac{2}{3}} + \frac{1}{2}\sigma^2 p A^{-p}(p)(z_i - \theta_p)^2 + O(\lambda).\end{aligned}$$

If $z_i = 1$, the term $(1-z_i)^2 u''(z_i)$ is set equal to zero (see [43], equation (A.5)), and the above equation still holds. To satisfy (2.11) we must have

$$\begin{aligned}0 &= \mathcal{D}u(z_i) - \widetilde{U}\big((1-p)u(z_i) - z_i u'(z_i)\big) \\ &= -\gamma_2 A(p)\,\lambda^{\frac{2}{3}} + \frac{1}{2}\sigma^2 p A^{-p}(p)(z_i - \theta_p)^2 + O(\lambda).\end{aligned} \tag{4.25}$$

It follows that

$$\begin{aligned}z_i &= \theta_p \pm \lambda^{\frac{1}{3}}\sqrt{\frac{2}{p\sigma^2}A^{p+1}(p)\gamma_2} + O\big(\lambda^{\frac{2}{3}}\big) \\ &= \theta_p \pm \left(\frac{3}{2p}\theta_p^2(1-\theta_p)^2\right)^{\frac{1}{3}}\ \lambda^{\frac{1}{3}} + O\big(\lambda^{\frac{2}{3}}\big),\end{aligned}$$

where the $-$ sign is for z_1 and the $+$ sign is for z_2.

Remark 1 The proof of Theorem 1 is valid so long as $\theta_p \neq 1$. The case of $\theta_p = 1$ can be considered a *singular* case for which the parameter ν appearing in a denominator in $h(\delta)$ is zero. The intuition is that if the optimal proportion in the risky asset is 100% of wealth, then as soon as the agent's position reachs the optimal 100% in stock, the investor no longer incurs transaction costs from adjusting it. However, to consume the agent must transfer money from stock to money market and pay the transaction cost. We could regard this cost as a *consumption tax*. It follows that the loss in value function is of order λ rather than $\lambda^{2/3}$. In [43], Theorems 11.2, 11.6 show that in this case $z_2 = \theta_p = 1 > z_1 > 0$, and Corollary 9.10 asserts

$$v(x,y) = \frac{1}{1-p}A^{-p}(p)\big(x + (1-\lambda)y\big)^{1-p}, \quad (x,y) \in \mathcal{S}, x \leq 0,$$

or equivalently,

$$u(z) = v(1-z, z) = \frac{1}{1-p}A^{-p}(p)(1-\lambda z)^{1-p}, \quad 1 \leq z < \frac{1}{\lambda}.$$

We see then that

$$u(\theta_p) = \frac{1}{1-p}A^{-p}(p) - A^{-p}(p)\lambda + O\big(\lambda^2\big), \tag{4.26}$$

which is consistent with (4.12).

The proof of Theorem 2 is valid for $\theta_p = 1$, except the last step. When $\theta_p = 1$, we have $\gamma_2 = 0$ and (4.25) reduces to

$$z_1 - \theta_p = z_1 - 1 = O\big(\lambda^{\frac{1}{2}}\big).$$

Remark 2 The optimal consumption policy c is given by $v_x^{-1/p}$ (see (2.7) and (2.6)). Since $v(x,y) = (x+y)^{1-p}u\big(y/(x+y)\big)$ (see (2.10)), we have

$$\begin{aligned} v_x(x,y) &= (1-p)(x+y)^{-p}u\left(\frac{y}{x+y}\right) - y(x+y)^{-1-p}u'\left(\frac{y}{x+y}\right) \\ &= (1-p)(x+y)^{-p}u\left(\frac{y}{x+y}\right) + O(\lambda), \end{aligned}$$

where we have used (4.24). Therefore, Theorem 1 implies that the optimal consumption in feedback form is

$$\begin{aligned} c &= \Big(A^{-p}(p) - (1-p)\gamma_2\lambda^{\frac{2}{3}}\Big)^{-\frac{1}{p}}(x+y) + O(\lambda) \\ &= \Big(A(p) + \frac{1-p}{p}\gamma_2\lambda^{\frac{2}{3}}\Big)(x+y) + O(\lambda). \end{aligned}$$

For sufficiently small λ the existence of transaction costs *increases* the size of consumption for $p \in (0,1)$, while the consumption is *decreased* for $p > 1$. This is explained by the fact that the *index of intertemporal substitution*, $1/p$, is high for small p. An agent with index of intertemporal substitution greater than 1 will avoid some transaction cost by consuming faster.

Remark 3 There is considerable evidence that

$$u(\theta_p) = \frac{1}{1-p}A^{-p}(p) - \gamma_2\lambda^{\frac{2}{3}} - \theta_p A^{-p}(p)\,\lambda + O\big(\lambda^{\frac{4}{3}}\big). \tag{4.27}$$

We have just seen in (4.26) that this is the case when $\theta_p = 1$ (and consequently $\gamma_2 = 0$). In the proof of Theorem 1, any choice of $M > \theta_p A^{-p}(p)$ gives us a subsolution of the form

$$w_-(\theta_p) = \frac{1}{1-p}A^{-p}(p) - \gamma_2\lambda^{\frac{2}{3}} - M\lambda + O\big(\lambda^{\frac{4}{3}}\big). \tag{4.28}$$

(The second term on the right-hand side of (4.13) is needed only for the supersolution argument.) Finally, the coefficient $-\theta_p A^{-p}(p)$ on λ can be obtained by a tedious heuristic analysis along the lines of Section 3.

In Remark 1 we introduced the concept of *consumption tax* as an interpretation of the transaction cost an agent must pay in order to move capital from stock to money market before consuming it. In that remark,

the agent was100% invested in stock. If instead the agent seeks to hold a proportion $\theta_p \neq 1$ in stock, then consuming proportionally from stock and money market, the agent would pay a consumption tax $\lambda\theta_p$ times the total consumption. To highest order, the optimal consumption level is thus $(1-\lambda\theta_p)$ of what it would be if there were no transaction cost, and this multiplies the value function by

$$(1-\lambda\theta_p)^{1-p} = 1-(1-p)\theta_p\,\lambda + O(\lambda^2).$$

The value function for zero transaction cost when wealth is 1 is $A^{-p}(p)/(1-p)$, and so after this multiplication, the value function has been reduced by $\theta_p A^{-p}(p)\,\lambda$, which is the order λ term we see in (4.27). We thus expect the value function for the problem with transaction cost λ to be reduced from the zero-transaction cost value function $A^{-p}(p)/(1-p)$ by at least $\theta_p A^{-p}(p)\lambda$, this reduction being due solely to the cost of moving capital from stock to money market in order to consume. There is also a cost of trading to stay in the NT wedge, which reduces the value function by $\gamma_2\lambda^{\frac{2}{3}}$, but cannot further reduce the value function by an order λ term because then the value function would fall below the lower bound (4.28).

Under the assumption that (4.27) holds, one can improve the calculations in Theorem 2. We already have from Theorem 2 and its proof that $z_i-\theta_p = O(\lambda^{1/3})$ and that $u'(z) = O(\lambda)$ for $z \in [z_1, z_2]$. The mean-value theorem gives

$$u(z_i) = u(\theta_p) + O\big(\lambda^{\frac{4}{3}}\big),$$

and

$$\begin{aligned}
(1-p)u(z_1) - z_1u'(z_1) &= \frac{(1-p)u(z_1)}{1+\lambda z_1}\\
&= A^{-p}(p) - (1-p)\gamma_2\,\lambda^{\frac{2}{3}} - (2-p)\theta_p A^{-p}(p)\,\lambda + O\big(\lambda^{\frac{4}{3}}\big),\\
(1-p)u(z_2) - z_2u'(z_2) &= \frac{(1-p)u(z_2)}{1-\lambda z_2}\\
&= A^{-p}(p) - (1-p)\gamma_2\,\lambda^{\frac{2}{3}} + p\theta_p A^{-p}(p)\,\lambda + O\big(\lambda^{\frac{4}{3}}\big).
\end{aligned}$$

Using Lemma 1, we obtain

$$\begin{aligned}
&\widetilde{U}\big((1-p)u(z_2) - z_2u'(z_2)\big)\\
&= \frac{p}{1-p}A^{1-p}(p) + (1-p)A(p)\gamma_2\,\lambda^{\frac{2}{3}} + (2-p)\theta_p A^{1-p}(p)\,\lambda + O\big(\lambda^{\frac{4}{3}}\big),\\
&\widetilde{U}\big((1-p)u(z_1) - z_1u'(z_1)\big)\\
&= \frac{p}{1-p}A^{1-p}(p) + (1-p)A(p)\gamma_2\,\lambda^{\frac{2}{3}} - p\theta_p A^{1-p}(p)\,\lambda + O\big(\lambda^{\frac{4}{3}}\big).
\end{aligned}$$

We write for z_1 and z_2 similarly as in Theorem 2,

$$\begin{aligned}
\mathcal{D}u(z_i) = \frac{p}{1-p}A^{1-p}(p) - pA(p)\gamma_2\,\lambda^{\frac{2}{3}} - p\theta_p A^{1-p}(p)\,\lambda\\
+\frac{1}{2}\sigma^2 pA^{-p}(p)(z_i-\theta_p)^2 + O\big(\lambda^{\frac{4}{3}}\big).
\end{aligned}$$

From $\mathcal{D}u(z_i) = \widetilde{U}\big((1-p)u(z_i) - z_i u'(z_i)\big)$, we now see that

$$(z_1 - \theta_p)^2 = \frac{2}{p\sigma^2}\left(A^{1+p}(p)\gamma_2\,\lambda^{\frac{2}{3}} + 2\theta_p A(p)\,\lambda\right) + O\big(\lambda^{\frac{4}{3}}\big), \tag{4.29}$$

$$(z_2 - \theta_p)^2 = \frac{2}{p\sigma^2}A^{1+p}(p)\gamma_2\,\lambda^{\frac{2}{3}} + O\big(\lambda^{\frac{4}{3}}\big), \tag{4.30}$$

which implies

$$z_1 - \theta_p = -\frac{1}{2}\nu\,\lambda^{\frac{1}{3}} - \frac{4\theta_p A(p)}{\sigma^2 p\,\nu}\,\lambda^{\frac{2}{3}} + O(\lambda),$$
$$z_2 - \theta_p = \frac{1}{2}\nu\,\lambda^{\frac{1}{3}} + O(\lambda).$$

This suggests that the optimal policy is to keep a wider wedge on the right side of the Merton proportion θ_p. This extra width makes sense because consumption reduces the money market position.

We can do the same calculation for $\theta_p = 1$, in which case $\gamma_2 = \nu = 0$. Taking the square roots in (4.29), (4.30), we now have

$$z_1 - \theta_p = z_1 - 1 = -\sqrt{\frac{4\theta_p A(p)}{\sigma^2 p}}\,\lambda^{\frac{1}{2}} + O\big(\lambda^{\frac{5}{6}}\big), \qquad z_2 - \theta_p = z_2 - 1 = O(\lambda).$$

In fact, when $\theta_p = 1$, we have $z_2 = \theta_p$, as we saw in Remark 1.

Remark 4 In the key formulas derived in this paper, the transaction cost parameter λ appears in combination with $\Gamma \triangleq \theta_p(1-\theta_p)$. According to Theorem 1, the highest order loss in the value function due to transaction costs is

$$\left(\frac{9p}{32}\right)^{\frac{1}{3}} A^{-1-p}(p)\sigma^2\left(\Gamma^2\lambda\right)^{\frac{2}{3}}. \tag{4.31}$$

From Theorem 2, we see that

$$z_i - \theta_p = (-1)^i\left(\frac{3}{2p}\right)^{\frac{1}{3}}\left(\Gamma^2\lambda\right)^{\frac{1}{3}} + O\big(\lambda^{\frac{2}{3}}\big). \tag{4.32}$$

One way to see the intrinsic nature of the quantity Γ^2 is to define the proportion of capital in stock, $\theta_t = Y_t/(X_t + Y_t)$, and apply Itô's formula when (X_t, Y_t) is generated by the optimal triple (C, L, M), for which L and M are continuous, to derive the equation

$$\begin{aligned}\mathrm{d}\theta_t = \theta_t(1-\theta_t)\frac{\mathrm{d}S_t}{S_t} - r\theta_t(1-\theta_t)\,\mathrm{d}t - \sigma^2\theta_t^2(1-\theta_t)\,\mathrm{d}t + \frac{\theta_t C_t}{X_t+Y_t}\,\mathrm{d}t \\ + \frac{\lambda\theta_t + 1}{X_t+Y_t}\,\mathrm{d}L_t + \frac{\lambda\theta_t - 1}{X_t+Y_t}\,\mathrm{d}M_t.\end{aligned} \tag{4.33}$$

We see that the response of θ_t to relative changes in the stock price is $\theta_t(1-\theta_t)$. When replicating an option by trading, the position held by the hedging portfolio, denominated in shares of stock, is called the *delta* of the

option, and the sensitivity of the delta to changes in the stock price is the *gamma*. We have here a similar situation, except that θ_t is the *proportion* of capital held in stock, rather than the *number of shares* of stock, and $\theta_t(1-\theta_t)$ is the sensitivity of this proportion to *relative* changes in the stock price.

We can now obtain the quantities in (4.31) and (4.32) by the following heuristic argument. Suppose when the transaction cost is $\lambda > 0$ we choose to keep θ_t in an interval $[\theta_p - w, \theta_p + w]$, trading in order to create reflection at the endpoints of this interval. As compared to the optimal expected utility corresponding to zero transaction costs, this entails two losses: the loss due to displacement (not always having $\theta_t = \theta_p$) and the loss due to transaction costs. The first increases with w while the second decreases. We wish to choose w to balance these losses, i.e., to minimize their sum.

In (3.1) we computed the loss due to displacement ϵ to be

$$L(\epsilon) = \frac{p}{2}(X_0 + Y_0)^{1-p}\sigma^2 A^{-1-p}(p)\epsilon^2.$$

In the present case, the displacement is a random, time-varying quantity taking values in $[-w, w]$. However, for small $w > 0$, for each positive t the distribution of θ_t in $[-w, w]$ is approximately uniform. Therefore, the loss due to displacement in the present model is approximately

$$DL(w) = \frac{1}{2w}\int_{-w}^{w} L(\epsilon)\, d\epsilon = \frac{p}{6}(X_0 + Y_0)^{1-p}\sigma^2 A^{-1-p}(p)w^2.$$

Consider a Brownian motion doubly reflected in order to stay in $[-w, w]$, i.e., $Z_t = W_t + L_t - M_t$, where $Z_t \in [-w, w]$ for all t, L_t grows only when $Z_t = -w$ and M_t grows only when $Z_t = w$. Then $\lim_{t\to\infty} \mathbb{E}L_t/t = \lim_{t\to\infty} \mathbb{E}M_t/t = 1/(4w)$. (Use Itô's formula to write

$$\mathbb{E}f(Z_t) = f(Z_0) + \frac{1}{2}\int_0^t \mathbb{E}f''(Z_u)\, du + f'(-w)\mathbb{E}L_t - f'(w)\mathbb{E}M_t,$$

divide by t, and let $t \to \infty$ to conclude

$$\frac{1}{4w}\int_{-w}^{w} f''(x)\, dx + f'(-w)\lim_{t\to\infty}\frac{1}{t}\mathbb{E}L_t - f'(w)\lim_{t\to\infty}\frac{1}{t}\mathbb{E}M_t = 0$$

for any C^2 function f.) The amount of "transacting" done at both boundaries grows at rate $1/(2w)$ per unit time. Because the $\mathrm{d}W(t)$ term in (4.33) is multiplied by $\theta_t(1-\theta_t)\sigma$, and scaling Brownian motion by a constant is like scaling time by the square of the constant, the amount of transacting done to keep θ_t in $[-w, w]$ grows at rate $\Gamma^2\sigma^2/(2w)$ per unit time. This causes a relative loss of capital

$$\eta = \frac{\Gamma^2\sigma^2\lambda}{2w}$$

per unit time. Incurring this loss is equivalent to having zero transaction cost but with the mean rate of return of both the stock and the money

market reduced by η. According to (2.3)–(2.5), the optimal expected utility in such a problem is

$$\frac{1}{1-p}\left(A(p)+\frac{\eta(1-p)}{p}\right)^{-p}(X_0+Y_0)^{1-p}$$
$$=\frac{1}{1-p}A^{-p}(p)(X_0+Y_0)^{1-p}-\eta A^{-p-1}(p)(X_0+Y_0)^{1-p}+O(\eta^2).$$

Therefore, the loss due to transaction costs is approximately

$$TC(w)=\eta A^{-p-1}(p)(X_0+Y_0)^{1-p}=\frac{1}{2w}\Gamma^2\sigma^2\lambda A^{-p-1}(p)(X_0+Y_0)^{1-p}.$$

The total loss is approximately $TL(w) = DL(w) + TC(w)$. Setting $TL'(w) = 0$ and solving for w, we obtain

$$w=\left(\frac{3}{2p}\right)^{\frac{1}{3}}(\Gamma^2\lambda)^{\frac{1}{3}},$$

the leading term in (4.32). With this value of w, we have

$$TL(w)=\left(\frac{9p}{32}\right)^{\frac{1}{3}}A^{-1-p}(p)\sigma^2(\Gamma^2\lambda)^{\frac{2}{3}}(X_0+Y_0)^{1-p},$$

thereby obtaining the term in (4.31).

It is interesting to note that the quantity $\Gamma^2\lambda$ also plays a fundamental role in the formal asymptotic expansions of Whalley and Wilmott [49]. In fact, even the constants $\left(\frac{9}{32}\right)^{\frac{1}{3}}$ and $\left(\frac{3}{2}\right)^{\frac{1}{3}}$ in (4.31) and (4.32) appear in [49], the first at the end of Section 3.3 and the second in equation (3.10).

A Width of the *NT* interval

We shall only consider δ of the form $O\big(\lambda^{1/3}\big)$. For such δ, we may write the terms of order λ and lower in $f_1^{\pm}(\delta)$ of (4.14) as

$$\begin{aligned}
&f_1^{\pm}(\delta)\\
&=\nu\,\lambda^{\frac{1}{3}}-(1-p)\gamma_2\nu A^p(p)\,\lambda-h'(\delta)\,\lambda^{-\frac{2}{3}}-\theta_p h'(\delta)\,\lambda^{\frac{1}{3}}+h'(\delta)\,\delta\lambda^{\frac{1}{3}}+O\big(\lambda^{\frac{4}{3}}\big)\\
&=\nu\,\lambda^{\frac{1}{3}}-(1-p)\gamma_2\nu A^p(p)\,\lambda-3\delta+\frac{4}{\nu^2}\cdot\frac{\delta^3}{\lambda^{\frac{2}{3}}}-3B\delta\lambda^{\frac{2}{3}}+O\big(\lambda^{\frac{4}{3}}\big).
\end{aligned}$$

Consider $\delta_0 \triangleq \frac{1}{2}\nu\lambda^{1/3}\left(1-\xi_0\lambda^{1/3}\right)$. Then

$$\begin{aligned}
f_1^{\pm}(\delta_0)&=\nu\,\lambda^{\frac{1}{3}}-(1-p)\gamma_2\nu A^p(p)\,\lambda-\frac{3}{2}\nu\,\lambda^{\frac{1}{3}}+\frac{3}{2}\nu\xi_0\,\lambda^{\frac{2}{3}}\\
&\quad+\frac{1}{2}\nu\,\lambda^{\frac{1}{3}}\left(1-3\xi_0\lambda^{\frac{1}{3}}+3\xi_0^2\lambda^{\frac{2}{3}}\right)-\frac{3}{2}B\nu\,\lambda+O\big(\lambda^{\frac{4}{3}}\big)\\
&=\nu\Big(\frac{3}{2}\xi_0^2-(1-p)\gamma_2A^p(p)-\frac{3}{2}B\Big)\,\lambda+O\big(\lambda^{\frac{4}{3}}\big).
\end{aligned}$$

With $\xi = \sqrt{\frac{2}{3}(1-p)\gamma_2 A^p(p) + B} > 0$, we take $\xi_0 = \sqrt{\xi^2 + \eta}$, where $|\eta| < \xi^2$. Then $f_1^{\pm}(\delta_0) = \frac{3}{2}\nu\eta\,\lambda + O\big(\lambda^{\frac{4}{3}}\big)$. Thus, for $\eta > 0$ we have $f_1^{\pm}(\delta_0) > 0$, and for $\eta < 0$ we have $f_1^{\pm}(\delta_0) < 0$ for sufficiently small $\lambda > 0$. Therefore, for every $\eta \in \big(0, \xi^2\big)$ and sufficiently small $\lambda > 0$, there exists

$$\delta_1^{\pm} \in \left(\frac{1}{2}\nu\,\lambda^{\frac{1}{3}}\Big(1 - \lambda^{\frac{1}{3}}\sqrt{\xi^2 - \eta}\Big), \frac{1}{2}\nu\,\lambda^{\frac{1}{3}}\Big(1 - \lambda^{\frac{1}{3}}\sqrt{\xi^2 + \eta}\Big)\right)$$

satisfying $f_1^{\pm}(\delta_1^{\pm}) = 0$. In other words, $\delta_1^{\pm} = \frac{1}{2}\nu\lambda^{\frac{1}{3}}\left(1 - \xi\,\lambda^{\frac{1}{3}}\right) + o\big(\lambda^{\frac{2}{3}}\big)$. The proof of the existence of $\delta_2^{\pm}$ is analogous.

References

1. Akian, M., Menaldi, J. L. & Sulem, A., On an investment consumption model with transaction costs, *SIAM J. Control* **34**, 329–364 (1996).
2. Atkinson, C. & Wilmott, P., Portfolio management with transaction costs: an asymptotic analysis of the Morton and Pliska model, *Math. Finance* **5**, 357–367 (1995).
3. Avellaneda, M. & Parás, A., Dynamic hedging portfolios for derivative securities in the presence of large transaction costs, *Appl. Math. Finance* **1**, 165–194 (1994).
4. Barles, G. & Soner, H. M., Option pricing with transaction costs and a nonlinear Black-Scholes equation, *Finance Stoch.* **2**, 369–397 (1998).
5. Bensaid, B., Lesne, J. P., Pagés, H. & Scheinkman, J., Derivative asset pricing with transaction costs, *Math. Finance* **2**, 63–86 (1992).
6. Bouchard, B. & Touzi, N., Explicit solution of the multivariable super-replication problem under transaction costs, *Ann. Appl. Probab.* **10**, 685–708 (2000).
7. Boyle, P. P. & Vorst, T., Option replication in discrete time with transaction costs, *J. Finance* **47**, 272–293 (1992).
8. Constantinides, G., Capital market equilibrium with transaction costs, *J. Political Economy* **94**, 842–862 (1986).
9. Constantinides, G. & Zariphopoulou, T., Bounds on prices of contingent claims in an intertemporal economy with proportional transaction costs and general preferences, *Finance Stoch.* **3**, 345–369 (1999).
10. Constantinides, G. & Zariphopoulou, T., Bounds on derivative prices in an intertemporal setting with proportional transaction costs and multiple securities, *Math. Finance* **11**, 331–346 (2001).
11. Cvitanič, J. & Karatzas, I., Hedging and portfolio optimization under transaction costs, *Math. Finance* **6**, 113–165 (1996).
12. Cvitanič, J. & Wang, H., On optimal terminal wealth under transaction costs, *J. Mathematical Economics* **35**, 223–232. (2001)
13. Cvitanič, J., Pham, H. & Touzi, N, A closed-form solution to the problem of super-replicating under transaction costs, *Finance Stoch.* **3**, 35–54 (1999).
14. Davis, M. H. A. & Norman, A., Portfolio selection with transaction costs, *Math. Oper. Res.* **15**, 676–713 (1990).
15. Davis, M. H. A., Panas, V. G. & Zariphopoulou, T., European option pricing with transaction costs, *SIAM J. Control* **31**, 470–493 (1993).

16. Deelstra, G., Pham, H. & Touzi, N., Dual formulation of the utility maximization problem under transaction costs, *Ann. Appl. Probab.* **11**, 1353–1383 (2001).
17. Delbaen, F., Kabanov, Yu. & Valkeila, E., Hedging under transaction costs in currency markets: a discrete-time model, *Math. Finance* **12**, 45–61 (2002).
18. Edirisinghe, C., Naik, V. & Uppal, R., Optimal replication of options with transaction costs and trading restrictions, *J. Finan. Quant. Anal.* **28**, 117–138 (1993).
19. Figlewski, S., Options arbitrage in imperfect markets, *J. Finance* **44**, 1289–1311 (1989).
20. Framstad, N., Øksendal, B. & Sulem, A., Optimal consumption and portfolio in a jump-diffusion market with proportional transaction costs, *J. Math. Econ.* **35**, 233–257 (2001).
21. Grannan, E. R. & Swindle, G., Minimizing transaction costs of option hedging strategies, *Math. Finance* **6**, 341–364 (1996).
22. Hodges, S. & Neuberger, A., Option replication of contingent claims under transaction costs, *Rev. Futures Markets* **8**, 222–239 (1989).
23. Hoggard, T., Whalley, A. & Wilmott, P. , Hedging option portfolios in the presence of transaction costs, *Adv. Fut. Oper. Res.* **7**, 21–35 (1994).
24. Jouni, E. , Market imperfections, equilibrium, and arbitrage, in *Lecture Notes in Mathematics 1656*, Springer-Verlag, Berlin, pp. 247–307 (1997).
25. Jouni, E. & Kallal, H., Martingales and arbitrage in securities markets with transaction costs, *J. Econ. Theory* **66**, 178–197 (1995).
26. Kabanov, Yu., Hedging and liquidation under transaction costs, *Finance Stoch.* **3**, 237–348 (1999).
27. Kabanov, Yu. & Klüppelberg, C. , A geometric approach to portfolio optimization in models with transaction costs, *Submitted for publication.*
28. Kabanov, Yu. & Last, G. , Hedging under transaction costs in currency markets, *Math. Finance* **12**, 63–70 (2002).
29. Kabanov, Yu. & Stricker, C. Hedging of contingent claims under transaction costs. In: Sandman, K., Schönbucher, P. (eds.) *Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann.* Springer 2002.
30. Karatzas, I. & Shreve, S. *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, 1988.
31. Koehl, P. F., Pham, H. & Touzi, N., Hedging in discrete-time under transaction costs and continuous-time limit, *J. Appl. Probab.* **36**, 163–178 (1999).
32. Koehl, P. F., Pham, H. & Touzi, N., On super-replication under transaction costs in general discrete-time models, *Theory of Probability and Applications* **45**, 783–788 (1999).
33. Korn, R., Portfolio optimisation with strictly positive transaction costs and impulse control, *Finance Stoch.* **2**, 85–114 (1998).
34. Korn, R. & Laue, S., Portfolio optimisation with transaction costs and exponential utility, preprint *FB Mathematik, Univ. Kaiserslautern* (2003).
35. Kusuoka, S., Limit theorem option replication with transaction costs, *Ann. Appl. Probab.* **5**, 198–221 (1995).
36. Leland, H., Option pricing and replication with transaction costs, *J. Finance* **40**, 1283–1301 (1985).
37. Leventhal, S. & Skorohod, A., On the possibility of hedging options in the presence of transaction costs, *Ann. Appl. Probab.* **7**, 410–443 (1997).

38. Loewenstein, M., On optimal portfolio trading strategies for an investor facing transaction costs in a continuous trading market, *J. Math. Econ.* **33**, 209–228 (2000).
39. Magill, M. J. .P. and Constantinides, G. M., Portfolio selection with transaction costs, *J. Econ. Theory* **13**, 245–263 (1976).
40. Merton, R., Optimum consumption and portfolio rules in a continuous-time case, *J. Econ. Theory* **3**, 373–413 (1971) [Erratum **6**, 213–214 (1973)].
41. Morton, A. & Pliska, S. R., Optimal portfolio management with fixed transaction costs, *Math. Finance* **5**, 337–357 (1995).
42. Samuelson, P. , Why we should not make mean log of wealth big though years to act are long, *J. Banking Finance* **3**, 305–307 (1979).
43. Shreve, S. & Soner, H. M., Optimal investment and consumption with transaction costs, *Ann. Applied Probab.* **4**, 609–692 (1994).
44. Soner, H. M., Shreve, S. & Cvitanič, J., There is no nontrivial hedging portfolio for option pricing with transaction costs, *Ann. Appl. Probab.* **5**, 327–355 (1995).
45. Toft, K. B., On the mean-variance tradeoff in option replication with transaction costs, *J. Financial Quant. Anal.* **31**, 233–263 (1996).
46. Tourin, A. & Zariphopoulou, T., Numerical schemes for investment models with singular transactions, *Computat. Econ.* **7**, 287–307 (1994).
47. Tourin, A. & Zariphopoulou, T., Viscosity solutions and numerical schemes for investment/consumption models with transaction costs, in *Numerical Methods in Finance*, L. C. G. Rogers and D. Talay, eds., Isaac Newton Institute Publications (1997).
48. Touzi, N., Super-replication under proportional transaction costs: from discrete-to-continuous time models, *Math. Methods Oper. Res.* **50**, 297–320 (1999).
49. Whalley, A. E. & Wilmott, P., An asymptotic analysis of an optimal hedging model for option pricing under transaction costs, *Math. Finance* **7**, 307–324 (1997).