Finance Stoch (2015) 19:363–414
DOI 10.1007/s00780-015-0261-3

Asymptotics for fixed transaction costs

Albert Altarovici[1] · Johannes Muhle-Karbe[1] · Halil Mete Soner[1]

Received: 12 June 2013 / Accepted: 1 September 2014 / Published online: 18 February 2015

Abstract An investor with constant relative risk aversion trades a safe and several risky assets with constant investment opportunities. For a small *fixed* transaction cost, levied on each trade regardless of its size, we explicitly determine the leading-order corrections to the frictionless value function and optimal policy.

Keywords Fixed transaction costs · Optimal investment and consumption · Homogenization · Viscosity solutions · Asymptotic expansions

Mathematics Subject Classification 91G10 · 91G80 · 35K55 · 60H30

JEL Classification G11

1 Introduction

Market frictions play a key role in portfolio choice, "drastically reducing the frequency and volume of trade" [9]. These imperfections manifest themselves in various forms. Trading costs *proportional* to the traded volume affect all investors in the form of bid–ask spreads. In addition, *fixed* costs, levied on each trade regardless of its size, also play a key role for small investors.

Proportional transaction costs have received most of the attention in the literature. On the one hand, this is due to their central importance for investors of all sizes.

✉ J. Muhle-Karbe
johannes.muhle-karbe@math.ethz.ch

A. Altarovici
albert.altarovici@math.ethz.ch

H.M. Soner
hmsoner@ethz.ch

1 Departement für Mathematik, ETH Zürich, Rämistrasse 101, 8092, Zürich, Switzerland

On the other hand, this stems from their relative analytical tractability: by their very definition, proportional costs are “scale invariant” in that their effect scales with the number of shares traded. With constant relative or absolute risk aversion and a constant investment opportunity set, this leads to a no-trade region of constant width around the frictionless target position [32, 9, 11, 12, 42]. Investors remain inactive while their holdings lie inside this region, and engage in the minimal amount of trading to return to its boundaries once these are breached. The trading boundaries can be determined numerically by solving a free-boundary problem [11]. In the limit for small costs, the no-trade region and the corresponding utility loss can be determined explicitly at the leading order; cf. Shreve and Soner [42], Whalley and Wilmott [44], Janeček and Shreve [22], and many more recent studies [6, 18, 43, 38, 7]. Extensions to more general preferences and stochastic opportunity sets have been studied numerically by Balduzzi, Lynch, and Tan [30, 3, 31]. Corresponding formal asymptotics have been determined by Goodman and Ostrov [19], Martin [33], Kallsen and Muhle-Karbe [25, 24], and Soner and Touzi [43]. The last study [43], also contains a rigorous convergence proof for general utilities, which is extended to several risky assets by Possamaï, Soner, and Touzi [38].

Proportional costs lead to infinitely many small transactions. In contrast, fixed costs only allow a finite number of trades over finite time intervals. However, the optimal policy again corresponds to a no-trade region. In this setting, trades of all sizes are penalized equally; therefore, rebalancing takes place by a bulk trade to the optimal frictionless target inside the no-trade region [13]. These “simple” policies involving only finitely many trades are appealing from a practical point of view. However, fixed costs destroy the favorable scaling properties that usually allow one to reduce the dimensionality of the problem for utilities with constant relative or absolute risk aversion. In particular, the boundaries of the no-trade region are no longer constant, even in the simplest settings with constant investment opportunities and constant absolute or relative risk aversion. Accordingly, the literature analyzing the impact of fixed trading costs is much more limited than for proportional costs: on the one hand, there are a number of numerical studies [41, 28] that iteratively solve the dynamic programming equations. On the other hand, Korn [27] and Lo, Mamaysky, and Wang [29] have obtained formal asymptotic results for investors with constant *absolute* risk aversion. For small costs, these authors find that constant trading boundaries are optimal at the leading order. Thus, these models are tractable but do not allow us to study how the impact of fixed trading costs depends on the size of the investor under consideration. The same applies to the “quasi-fixed” costs proposed by Morton and Pliska [34] and analyzed in the small-cost limit by Atkinson and Wilmott [2]. In their model, each trade—regardless of its size—incurs a cost proportional to the investors' current wealth, leading to a scale-invariant model where investors of all sizes are affected by the “quasi-fixed” costs to the same extent. Similarly, the asymptotically efficient discretization rules developed by Fukasawa [16, 17] and Rosenbaum and Tankov [40] also do not take into account that the effect of fixed trading costs should depend on the “size” of the investor under consideration.[1]

[1]Indeed, these schemes asymptotically correspond to constant absolute risk version; see [17] for more details.

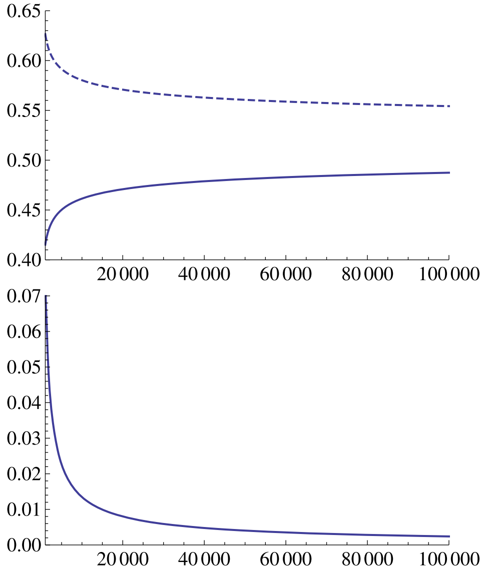


Fig. 1 Trading boundaries (*top panel*) and equivalent proportional costs (*bottom panel*) for a \$1 fixed transaction cost as functions of the investor's wealth. Parameters are $\gamma = 6$, $\mu - r = 8\ \%$, and $\sigma = 16\ \%$, so that the Merton proportion is $\pi_m = 52\ \%$

The present study helps to overcome these limitations by providing rigorous asymptotic expansions for investors with constant *relative* risk aversion.[2] In the standard infinite-horizon consumption model with constant investment opportunities, we obtain explicit formulas for the leading-order welfare effect of small fixed costs and a corresponding almost optimal trading policy. These shed new light on the differences and similarities compared to proportional transaction costs.

A universal theme is that as for proportional transaction costs [22, 33, 25, 24], the crucial statistic of the optimal frictionless policy turns out to be its "portfolio gamma," which trades off the local variabilities of the strategy and the market (see (2.6)). The latter is also crucial in the asymptotic analysis of finely discretized trading strategies [45, 5, 21, 16, 17, 40]. Therefore, it appears to be an appealingly robust proxy for the sensitivity of trading strategies to small frictions.

A fundamental departure from the corresponding results for proportional transaction costs is that the effect of small fixed costs is inversely proportional to investors' wealth. That is, doubling the fixed cost has the same effect on investors' welfare and trading boundaries as halving their wealth.[3] This quantifies the extent to which fixed costs can be neglected by large institutional entities or, contrarily, need to be taken into account by small private investors. For example, for typical market parameters (see Fig. 1), a fixed transaction cost of \$1 per trade leads to trading boundaries of 45 % and 59 % around the frictionless Merton proportion of 52 % if the investor's

[2] For our *formal* derivations, we consider general utilities like in recent independent work of Alcala and Fahim [1].

[3] Here, both quantities are measured in relative terms, as is customary for investors with constant relative risk aversion. That is, trading boundaries are parameterized by the fractions of wealth held in the risky asset, and the welfare effect is described by the relative certainty equivalent loss, that is, the fraction of the initial endowment the investor would be willing to give up to trade without frictions.

wealth is \$5000. If wealth increases to \$100,000, however, the trading boundaries narrow to 49 % and 55 %, respectively. Our results also show that asymptotically for small costs, fixed transaction costs are equivalent—both in terms of the no-trade region and the corresponding welfare loss—to a suitable "equivalent proportional cost." Since the effect of the fixed costs varies with investors' wealth, this equivalent proportional cost is not constant, but decreases with the investors' wealth level. For example, with typical market parameters (see Fig. 1), a \$1 fixed cost corresponds to a proportional cost of 2.3 % if the investor's wealth is \$5000, but to only 0.24 % if wealth is \$100,000. In a similar spirit, our results are also formally linked to those of Atkinson and Wilmott [2]: their trading costs, taken to be a constant fraction of the investors' current wealth, formally lead to the same results as substituting a stochastic fixed cost proportional to current wealth into our formulas.

A second novelty is that our results readily extend to a multivariate setting with several risky assets. This is in contrast to the models with proportional transaction costs, where optimal no-trade regions for several risky assets can only be determined numerically by solving a multidimensional nonlinear free-boundary problem, even in the limit for small costs [38]. With small fixed costs, the optimal no-trade region with several risky assets turns out to be an ellipsoid centered around the frictionless target, whose precise shape is easily determined even in high dimensions by the solution of a matrix-valued algebraic Riccati equation. This is again in line with the quasi-fixed costs studied by Atkinson and Wilmott [2], up to rescaling the transaction cost by current wealth. Qualitatively, the shape of our ellipsoid resembles the one for the parallelogram-like regions computed numerically for proportional transaction costs by Muthuraman and Kumar [35] and Possamaï, Soner, and Touzi [38]. On a quantitative level, however, we find that the shape of the ellipsoid is much more robust with respect to correlation among the risky assets.

Finally, the present study provides the first rigorous proofs for asymptotics with small fixed costs, complementing earlier partially heuristic results [27, 29, 1], rigorous analyses of the related problem of optimal discretization [16, 17, 40], and rigorous asymptotics with proportional costs (see [42, 22, 6, 18, 43, 38, 7]). As for proportional costs [43], our approach is based on the theories of viscosity solutions and homogenization, in particular, the weak-limits technique of Barles and Perthame [4] and Evans [14]. However, substantial new difficulties have to be overcome because i) the value function is not concave, ii) the usual dimensionality reduction techniques fail even in the simplest models, iii) the set of controls is not scale-invariant, and iv) the dynamic programming equation involves a nonlocal operator here. In order not to drown these new features in further technicalities, we leave for future research the extension to more general preferences and asset price and cost dynamics as in [43, 25, 24] for proportional costs, and also the analysis of the joint impact of proportional and fixed costs.[4]

The remainder of the article is organized as follows. The model, the main results, and their implications are presented in Sect. 2. Subsequently, we derive the results in an informal manner. This is done in some detail to explain the general procedure that is likely to be applicable for a number of related problems. In particular,

[4] See [27, 1] for corresponding formal asymptotics.

we explain how to come up with the scaling in powers of $\lambda^{1/4}$ by heuristic arguments as in [22, 39] and discuss how to use homogenization techniques to derive the corrector equations describing the first-order approximations of the exact solution. Sections 4–8 then make these formal arguments rigorous by providing a convergence proof. Some technical estimates are deferred to Appendix A. Finally, Appendix B presents a self-contained proof of the weak dynamic programming principle in the spirit of Bouchard and Touzi [8], which in turn leads to the viscosity solution property of the value function for the problem at hand.

Throughout, we denote by $x^\top$ the transpose of a vector or matrix x; we also set $\mathbf{1}_d := (1,\dots,1)^\top \in \mathbb{R}^d$ and write I_d for the identity matrix on $\mathbb{R}^d$. For a vector $x \in \mathbb{R}^d$, the diagonal matrix with diagonal elements $x^1,\dots,x^d$ is denoted by $\operatorname{diag}[x]$. We also use the notation D_p and D_p^2 for the gradient and Hessian with respect to a variable p. Finally, $x \cdot y$ denotes the standard inner product of two Euclidean vectors, and we write $x \otimes y = x_i y_j$ for the matrix with entries $x_i y_j$.

2 Model and main results

2.1 Market, trading strategies, and wealth dynamics

Consider a financial market consisting of a safe asset earning a constant interest rate $r > 0$ and of d risky assets with expected excess returns $\mu^i - r > 0$ and invertible infinitesimal covariance matrix $\sigma\sigma^\top$, that is,

$$dS_t^0 = S_t^0 r\,dt, \quad dS_t = S_t \mu\,dt + S_t \sigma\,dW_t,$$

for a d-dimensional standard Brownian motion $(W_t)_{t\geq 0}$ defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$, where $(\mathscr{F}_t)_{t\geq 0}$ denotes the augmentation of the filtration generated by $(W_t)_{t\geq 0}$. Each trade incurs a *fixed transaction cost* $\lambda > 0$, regardless of its size or the number of assets involved. As a result, portfolios can only be rebalanced finitely many times over finite time intervals, and trading strategies can be described by pairs (τ, m), where the trading times $\tau = (\tau_1, \tau_2, \dots)$ are a sequence of stopping times increasing toward infinity, and the $\mathscr{F}_{\tau_i}$-measurable, $\mathbb{R}^d$-valued random variables collected in $m = (m_1, m_2, \dots)$ describe the transfers at each trading time. More specifically, m_i^j represents the monetary amount transferred from the safe to the jth risky asset at time τ_i. Each trade is assumed to be self-financing, and the fixed costs are deducted from the safe asset account. Thus, the safe and risky positions evolve as

$$(x, y) = (x, y^1, \dots, y^d) \mapsto \left(x - \sum_{j=1}^d m_i^j - \lambda, y^1 + m_i^1, \dots, y^d + m_i^d \right)$$

for each trade m_i at time τ_i. The investor also consumes from the safe account at some rate $(c_t)_{t\geq 0}$. Starting from an initial position $(X_{0-}, Y_{0-}) = (x, y) \in \mathbb{R} \times \mathbb{R}^d$, the wealth dynamics corresponding to a *consumption–investment strategy* $\nu = (c, \tau, m)$

are therefore given by

$$X_t = x + \int_0^t (rX_s - c_s)\,ds - \sum_{k=1}^{\infty}\left(\lambda + \sum_{j=1}^{d} m_k^j\right)1_{\{\tau_k \le t\}},$$

$$Y_t^i = y^i + \int_0^t Y_s^i \frac{dS_s^i}{S_s^i} + \sum_{k=1}^{\infty} m_k^i 1_{\{\tau_k \le t\}}.$$

We write $(X, Y)^{\nu,x,y}$ for the solution of this equation. The *solvency region*

$$\mathrm{K}_\lambda := \left\{(x, y) \in \mathbb{R}^{d+1} : \max\left\{x + y \cdot \mathbf{1}_d - \lambda, \min_{i=1,\dots,d}\{x, y^i\}\right\} \ge 0\right\}$$

is the set of positions with nonnegative liquidation value. A strategy $\nu = (c, \tau, m)$ starting from the initial position (x, y) is called *admissible* if it remains solvent at all times, that is, $(X_t^{\nu,x}, Y_t^{\nu,y}) \in \mathrm{K}_\lambda$ for all $t \ge 0$ $\mathbb{P}$-a.s. The set of all admissible strategies is denoted by $\Theta^\lambda(x, y)$.

2.2 Preferences

In the above market with constant investment opportunities (r, μ, σ) and fixed transaction costs λ, an investor with *constant relative risk aversion* $\gamma > 0$, that is, with utility function $U_\gamma : (0, \infty) \to \mathbb{R}$ of either logarithmic or power type,

$$U_\gamma(c) = \begin{cases} c^{1-\gamma}/(1-\gamma), & 0 < \gamma \ne 1, \\ \log c, & \gamma = 1, \end{cases}$$

and *impatience rate* $\beta > 0$ trades to maximize the expected utility from consumption over an infinite horizon, starting from an initial endowment of $X_{0-} = x$ in the safe and $Y_{0-} = y$ in the risky assets, respectively.[5] So we consider

$$v^\lambda(x, y) = \sup_{(c,\tau,m)\in\Theta^\lambda(x,y)} \mathbb{E}\left[\int_0^\infty e^{-\beta t} U_\gamma(c_t)\,dt\right]. \tag{2.1}$$

Theorem 2.1 *The value function* v^λ *of the problem with fixed costs* $\lambda > 0$ *is a* (*possibly*) *discontinuous viscosity solution of the dynamic programming equation* (3.7) *in the domain*

$$\mathcal{O}_\lambda = \{(x, y) \in \mathrm{K}_\lambda \ : \ x + y \cdot \mathbf{1}_d > 2\lambda\}.$$

For our asymptotic results, it suffices to obtain this result for $\mathcal{O}_\lambda$ rather than for the full solvency region K_λ. This is because any fixed initial allocation $(x, y) \in \mathbb{R}^{d+1}$ with $x + y \cdot \mathbf{1}_d > 0$ will satisfy $(x, y) \in \mathcal{O}_\lambda$ for sufficiently small λ.

For the definition of a discontinuous viscosity solution, we refer the reader to [10, 15, 23, 37]. Øksendal and Sulem [37] study the existence and uniqueness for

[5] By convention, the value of the integral is set to minus infinity if its negative part is infinite.

one risky asset and power utility with risk aversion $\gamma \in (0,1)$ under the additional assumption $\beta > (1-\gamma)\mu$, a sufficient condition for the finiteness of the frictionless value function. The proof of Theorem 2.1 is given in Appendix B by establishing a weak dynamic programming principle in the spirit of Bouchard and Touzi [8]. We believe that in analogy to corresponding results for proportional costs [42], Theorem 2.1 and a comparison result hold in the entire solvency region for all utility functions whenever the transaction cost value function v^λ is finite. However, this extension is not needed here.

2.3 Main results

Let us first collect the necessary inputs from the frictionless version of the problem (see, e.g., [15]). Denote by

$$\pi_m = (\sigma\sigma^\top)^{-1}(\mu - r\mathbf{1}_d)/\gamma$$

the optimal frictionless target weights, that is, the Merton proportions, in the risky assets. Write

$$c_m(\gamma) = \frac{1}{\gamma}\beta + \left(1 - \frac{1}{\gamma}\right)\left(r + \frac{(\mu - r\mathbf{1}_d)^\top(\sigma\sigma^\top)^{-1}(\mu - r\mathbf{1}_d)}{2\gamma}\right)$$

for the frictionless optimal consumption rate and let

$$v(z) = \begin{cases} \frac{z^{1-\gamma}}{1-\gamma} c_m^{-\gamma}, & \gamma \neq 1, \\ \frac{1}{\beta}\log(\beta z) + \frac{1}{\beta^2}\left(r + \frac{(\mu - r\mathbf{1}_d)^\top(\sigma\sigma^\top)^{-1}(\mu - r\mathbf{1}_d)}{2} - \beta\right), & \gamma = 1, \end{cases} \tag{2.2}$$

be the value function for the frictionless counterpart of (2.1) with initial wealth $z = x + y \cdot \mathbf{1}_d$. The latter is finite, provided that $c_m > 0$, **which we assume throughout.** Moreover, we also suppose that the following matrix is invertible:

$$\alpha = (I_d - \pi_m \mathbf{1}_d^\top)\,\mathrm{diag}[\pi_m]\sigma. \tag{2.3}$$

Remark 2.2 Assuming (2.3) to be invertible ensures that the asymptotically optimal no-trade region in Theorem 2.4 below is nondegenerate. This is tantamount to a nontrivial investment in each of the $d+1$ assets.

Our main results are the leading-order corrections for small fixed transaction costs λ; their interpretation and connections to the literature are discussed in Sect. 2.4.

Theorem 2.3 (Expansion of the value function) *For all solvent initial endowments* $(x,y) \in \mathbb{R}^{d+1}$ *with* $z = x + y \cdot \mathbf{1}_d > 0$, *we have*

$$v^\lambda(x,y) = v(z) - \lambda^{1/2}u(z) + o(\lambda^{1/2}),$$

that is,

$$u^\lambda(x,y) := \frac{v(z) - v^\lambda(x,y)}{\lambda^{1/2}} \longrightarrow u(x+y),$$

locally uniformly as $\lambda \to 0$. *Here*,

$$u(z) = u_0 z^{1/2-\gamma}$$

for a constant $u_0 > 0$ *determined by the* corrector equations *from Definition* 3.1. *For a single risky asset* $(d = 1)$, *we have*

$$u_0 = \sigma^2 \left(\frac{\gamma}{3}\pi_m^2 (1-\pi_m)^2\right)^{1/2} \frac{c_m(\gamma)^{-\gamma}}{c_m(2\gamma)};$$

see Sect. 3.6 *for the multivariate case.*

This determines the leading-order relative certainty equivalent loss, *that is*, *the fraction of her initial endowment the investor would give up to trade the risky asset without transaction costs*, *as follows*:

$$v^\lambda(x, y) = v\left(z\left(1 - u_0 c_m(\gamma)^\gamma \frac{\lambda^{1/2}}{z^{1/2}}\right)\right) + o(\lambda^{1/2}). \tag{2.4}$$

The leading-order optimal performance from Theorem 2.3 is achieved by the following "almost optimal policy":

Theorem 2.4 (Almost optimal policy) *Fix a solvent initial portfolio allocation. Define the no-trade region*

$$\mathrm{NT}^\lambda = \left\{(x, y) \in \mathbb{R}^{d+1} : \frac{y}{x + y \cdot \mathbf{1}_d} \in \pi_m + \frac{\lambda^{1/4}}{(x + y \cdot \mathbf{1}_d)^{1/4}} \mathscr{I}\right\}$$

for the ellipsoid $\mathscr{I} = \{\rho \in \mathbb{R}^d : \rho^\top M \rho < 1\}$ *from Sect.* 3.6. *Consider the strategy that consumes at the frictionless Merton rate*, *does not trade while the current position lies in the above no-trade region*, *and jumps to the frictionless Merton proportion once its boundaries are breached. Then*, *for any* $\delta > 0$, *the utility obtained from following this strategy until wealth falls to level* δ *and then switching to a leading-order optimal strategy for* (2.1) *is optimal at the leading order* $\lambda^{1/2}$ (*see Sect.* 8.2 *for more details*).

For a single risky asset, *the above no-trade region simplifies to the following interval around the frictionless Merton proportion*:

$$\mathrm{NT}^\lambda = \left\{(x, y) \in \mathbb{R}^2 : \left|\frac{y}{x+y} - \pi_m\right| \le \left(\frac{12}{\gamma}\pi_m^2 (1-\pi_m)^2 \frac{\lambda}{x+y}\right)^{1/4}\right\}. \tag{2.5}$$

Remark 2.5 Unlike for proportional transaction costs, trading only after leaving the above asymptotic no-trade region is not admissible for any given fixed cost $\lambda > 0$. This is because wealth can fall below the level λ needed to perform a final liquidating trade. Hence, this region is only "locally" optimal in that one needs to switch to the unknown optimal policy after wealth falls below a given threshold.

2.4 Interpretations and implications

In this section, we discuss a number of interpretations and implications of our main results. We first focus on the simplest case of one safe and one risky asset, before turning to several correlated securities.

2.4.1 Small frictions and portfolio gammas

The transactions of the optimal policies for proportional and fixed costs are radically different. For proportional costs, there is an infinite number of small trades of "local-time type," whereas fixed costs lead to finitely many bulk trades over finite time intervals. Nevertheless, the respective no-trade regions—that indicate when trading is initiated—turn out to be determined by exactly the statistics summarizing the market and preference parameters.

Indeed, just as for proportional transaction costs [22], the width of the leading-order optimal no-trade region in (2.5) is determined by a power of $\pi_m^2(1-\pi_m)^2$ rescaled by the investor's risk tolerance $1/\gamma$. This term quantifies the sensitivity of the current risky weight with respect to changes in the price of the risky asset; cf. [22, Remark 4]. Compared to the corresponding formula for proportional transaction costs in [22], it enters through its quartic rather than cubic root and is multiplied by a different constant. Nevertheless, most qualitative features remain the same: the leading-order no-trade region vanishes if a full safe or risky investment is optimal in the absence of frictions ($\pi_m = 0$ or $\pi_m = 1$, respectively), and the effect on optimal strategies increases significantly in the presence of leverage ($\pi_m > 1$, cf. [18]).

As in [33, 25, 24] for proportional costs, the no-trade region can also be interpreted in terms of the activities of the frictionless optimizer and the market as follows. Let $\varphi_m(t) = \pi_m Z_t/S_t$ be the frictionless optimal strategy for current wealth Z_t, expressed in terms of the number of shares held in the risky asset. Then the frictionless wealth dynamics $dZ_t = Z_t\pi_m\, dS_t/S_t - c_t\, dt$ and Itô's formula yield

$$\frac{d\langle\varphi_m\rangle_t}{dt} = \frac{\pi_m^2(1-\pi_m)^2\sigma^2 Z_t^2}{S_t^2}.$$

As a result, the maximal deviations (2.5) from the frictionless target can be rewritten in numbers of risky shares as

$$\pm\left(\frac{12}{\gamma}\frac{d\langle\varphi_m\rangle_t}{d\langle S\rangle_t}\frac{\lambda}{Z_t}\right)^{1/4}.$$

Our formal results from Sect. 3.5 suggest that an analogous result remains valid also for more general preferences. Then, the frictionless target $\varphi_m(t) = \theta(Z_t)/S_t$ (see Sect. 3.1) no longer corresponds to a constant target weight, and Itô's formula yields

$$\frac{d\langle\varphi_m\rangle_t}{dt} = \frac{\sigma^2\theta^2(Z_t)(1-\theta_z^2(Z_t))^2}{S_t^2},$$

so that the maximal deviations (3.19) from the frictionless target $\varphi_m(t)$ can be written as

$$\pm\left(\frac{12}{-v_{zz}(z)/v_z(z)}\frac{d\langle\varphi_m\rangle_t}{d\langle S\rangle_t}\lambda\right)^{1/4}, \tag{2.6}$$

in terms of numbers of risky shares. Up to changing the power and the constant, this is the same formula as for proportional transaction costs [25, 33, 24]: the width of the no-trade region is determined by the transaction cost, times the (squared) *portfolio gamma* $d\langle\varphi_m\rangle_t/d\langle S\rangle_t$, times the risk tolerance of the indirect utility function of the frictionless problem. The portfolio gamma also is the key driver in the analysis of finely discretized trading strategies [45, 5, 21, 16, 17, 40]. Hence, it appears to be an appealingly robust measure for the sensitivity of trading strategies to small frictions.

2.4.2 Wealth dependence and equivalent proportional costs

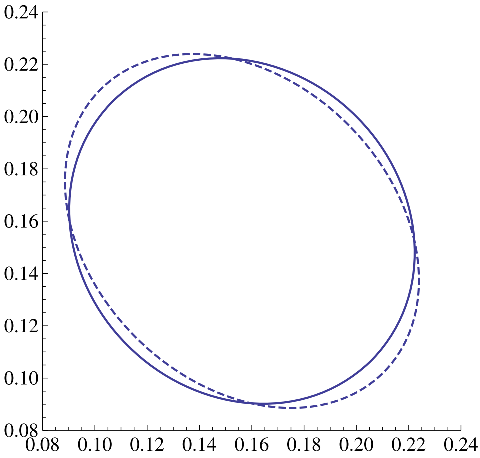
A fundamental departure from the corresponding results for proportional transaction costs is that the impact of fixed costs depends on investors' wealth. Indeed, the fixed cost λ is normalized by the investors' current wealth, both in the asymptotically optimal trading boundaries (2.5) and in the leading-order relative welfare loss (2.4); see Fig. 1 for an illustration. This makes precise to what extent fixed costs can indeed be neglected for large institutional traders, but play a key role for small private investors: ceteris paribus, doubling the investors' wealth reduces the impact of fixed trading costs in exactly the same way as halving the costs themselves. As a result, a constant fixed cost leads to a no-trade region that fluctuates with the investors' wealth. In contrast, for proportional transaction costs, this only happens if these evolve stochastically. The formal results of Kallsen and Muhle-Karbe [24] shed more light on this connection. It turns out that a constant fixed cost λ is equivalent—both in terms of the associated no-trade region *and* the corresponding welfare loss—to a random and time-varying proportional cost given by

$$\lambda_t^{\text{equiv}} = \left(\frac{1024\gamma}{3\pi_m^2(1-\pi_m)^2}\right)^{1/4}\left(\frac{\lambda}{Z_t}\right)^{3/4}$$

for current total wealth Z_t.[6] Note that this formula is independent of the impatience parameter β and only depends on the market parameters (μ, σ, r) through the Merton proportion $\pi_m = (\mu - r)/\gamma\sigma^2$. This relation clearly shows that a fixed cost corresponds to a larger proportional cost if rebalancing trades are small because i) the investors' wealth Z_t is small or ii) the no-trade region is narrow because the frictionless optimal position π_m is close to a full safe or risky position ($\pi_m = 0$ or $\pi_m = 1$). In contrast, for large investors and a frictionless position sufficiently far away from full risky or safe investment, the effect of fixed costs becomes negligible (see Fig. 1 for an illustration). For sufficiently high risk aversion γ, the equivalent proportional cost is increasing in risk aversion (as higher risk aversion leads to smaller trades),

[6] To see this, formally let the time horizon tend to infinity in [24, Sects. 4.1 and 4.2] and insert the explicit formulas for the optimal consumption rate and risky weight. This immediately yields that the leading-order no-trade regions coincide; for the corresponding welfare effects, this follows after integrating.

Fig. 2 No-trade ellipsoid for two identical risky assets with excess returns 5 %, correlation 0 (*solid*) resp. 44 % (*dashed*) and corresponding volatilities 40 % resp. 33 % chosen so that the Merton proportion remains constant at $(5/32, 5/32)$. Risk aversion is $\gamma = 2$, wealth is \$50,000, and the fixed cost is \$3.41



in line with the numerical findings of Liu [28] for exponential utility. Here, however, one can additionally assess the impact of changing wealth over time endogenously, rather than by having to vary the investors' risk aversion.

Our asymptotic formulas for fixed costs also allow us to relate these to the fixed *fraction* of current wealth charged per transaction in the model of Morton and Pliska [34]. Their "quasi-fixed" costs are scale-invariant in that they lead to constant trading boundaries around the Merton proportion π_m, whose asymptotics have been derived by Atkinson and Wilmott [2]. Formally, these trading boundaries coincide with ours if the ratio of their time-varying trading cost and our fixed fee is given by the investors' current wealth.

2.4.3 Multiple stocks

For multiple stocks, Theorem 2.4 shows that it is approximately optimal to keep the portfolio weight in an ellipsoid around the frictionless Merton position π_m. Whereas nonlinear free-boundary problems have to be solved to determine the optimal no-trade region for proportional costs even if these are small [38], the asymptotically optimal no-trade ellipsoid with fixed costs is determined by a matrix-valued algebraic Riccati equation, which is readily evaluated numerically even in high dimensions (see Sect. 3.6 for more details). Qualitatively, this is again in analogy to the asymptotic results of Atkinson and Wilmott [2] for the Merton and Pliska model [2], but—as for a single risky asset—the trading boundary varies with investors' wealth for the fixed costs considered here.

To shed some light on the quantitative features of the solution, Fig. 2 depicts the no-trade ellipsoid for two identical risky assets with varying degrees of correlation.[7] Qualitatively, correlation deforms the shape of the no-trade region similarly as in Muthuraman and Kumar [35, Fig. 6.8] for proportional costs: in the space of risky

[7] To facilitate comparison, we use the same market parameters μ, σ, r and risk aversion γ as in Muthuraman and Kumar [35]. The fixed cost and the current wealth are chosen so that the one-dimensional no-trade region for each asset corresponds to the one for their 1 % proportional cost.

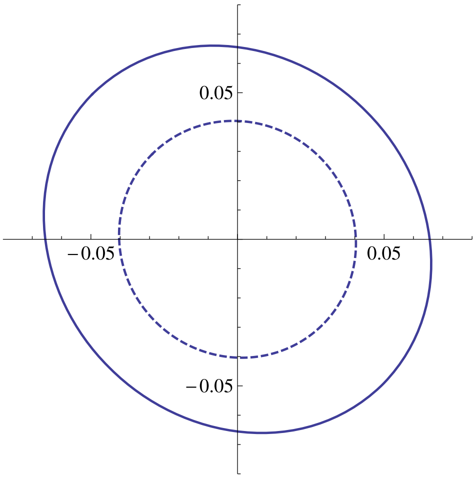


Fig. 3 Maximal deviations (in percentages of wealth held in the risky assets) from the frictionless Merton proportion for two identical risky assets with excess returns 5 %, correlation 0, volatilities 40 %, and risk aversions 2 (*solid*) and 6 (*dashed*). Wealth is \$50,000, and the fixed cost is \$3.41

asset weights, the no-trade region shrinks in the $(1,1)$ direction but widens in the $(1,-1)$-direction because investors use the positively correlated assets as partial substitutes for each other.

On a quantitative level, however, the impact of correlation turns out to be considerably less pronounced for fixed costs. This is because whenever *any* trade happens, all stocks can be traded with no extra cost, weakening the incentive to use substitutes for hedging. Also notice that the no-trade region is not rotationally symmetric even for two identical uncorrelated stocks. This is in contrast to the results for exponential utilities, for which the investor's maximization problem factorizes into a number of independent subproblems [28]. Note, however, that as risk aversion increases, the optimal no-trade region for uncorrelated identical stocks quickly becomes more and more symmetric, in line with the high risk aversion asymptotics linking power utilities to their exponential counterparts.[8] This is illustrated in Fig. 3.

3 Heuristic derivation of the solution

In this section, we explain how to use the homogenization approach to determine the small-cost asymptotics on an informal level. The derivations are similar to the ones for proportional costs [43].

Since this entails few additional difficulties on a *formal* level, we consider general utilities U defined on the positive half-line in this section. For the rigorous convergence proofs in Sect. 4, we focus on utilities U_γ with constant relative risk aversion in order not to drown the arguments in technicalities.

[8]Compare Nutz [36] for a general frictionless setting and Guasoni and Muhle-Karbe [20] for a model with proportional transaction costs. A similar result for fixed costs is more difficult to formulate because the investor's wealth does not factor out of the trading policy in this case.

3.1 The frictionless problem

The starting point for the present asymptotic analysis is the solution of the frictionless version of the problem at hand. Since trades are costless in that setting, the corresponding value function does not depend separately on the positions x, y in the safe and the risky assets, but only on total wealth $z = x + y \cdot \mathbf{1}_d$. As is well known (see e.g. [15, Chapter X]), the frictionless value function solves the dynamic programming equation

$$0 = \widetilde{U}\big(v_z(z)\big) - \beta v(z) + \mathscr{L}_0 v(z), \tag{3.1}$$

where

$$\mathscr{L}_0 v(z) = v_z(z) z r + v_z(z)(\mu - r\mathbf{1}_d) T\theta(z) + \frac{1}{2} v_{zz}(z)|\sigma^\top \theta(z)|^2, \tag{3.2}$$

and the corresponding optimal consumption rate and risky positions are given by

$$\kappa(z) := (U')^{-1}\big(v_z(z)\big) \tag{3.3}$$

and

$$\theta(z) := -\frac{v_z(z)}{v_{zz}(z)} (\sigma\sigma^\top)^{-1}(\mu - r\mathbf{1}_d). \tag{3.4}$$

For power or logarithmic utilities $U_\gamma(z)$, which have constant relative risk aversion $-zU''_\gamma(z)/U'_\gamma(z) = \gamma$, this leads to the explicit formulas from Sect. 2.3 because the value function is homothetic in this case: we have $v(z) = z^{1-\gamma} v(1)$ (if $\gamma \neq 1$) resp. $v(z) = \frac{1}{\beta} \log z + v(1)$ (if $\gamma = 1$).

3.2 The frictional dynamic programming equation

For the convenience of the reader, we now recall how to heuristically derive the dynamic programming equation with fixed trading costs. We start from the ansatz that the value function $v^\lambda(x, y)$ for our infinite-horizon problem with constant model parameters should only depend on the positions in each of the assets. Evaluated along the positions X_t, Y_t corresponding to any admissible policy $\nu = (c, \tau, m)$, Itô's formula in turn yields

$$\begin{aligned} &dv^\lambda(X_t, Y_t) \\ &= \Big(v^\lambda_x(X_t, Y_t)(rX_t - c_t) + \mu \cdot \mathbf{D}_y v^\lambda(X_t, Y_t) + \frac{1}{2}\mathrm{Tr}[\sigma\sigma^\top \mathbf{D}_{yy} v^\lambda(X_t, Y_t)]\Big)\, dt \\ &\quad + \mathbf{D}_y v^\lambda(X_t, Y_t)^\top \sigma\, dW_t + \sum_{\tau_i \le t} \big(v^\lambda(X_{\tau_i} - m_i \cdot \mathbf{1}_d - \lambda, Y_{\tau_i} + m_i) - v^\lambda(X_{\tau_i}, Y_{\tau_i})\big), \end{aligned} \tag{3.5}$$

where

$$\mathbf{D}^i_y = y^i \frac{\partial}{\partial y_i}, \quad \mathbf{D}^{ij}_{yy} = y^i y^j \frac{\partial^2}{\partial y^i \partial y^j}, \quad i, j = 1, \ldots, d.$$

By the martingale optimality principle of stochastic control, the utility

$$\int_0^t e^{-\beta s} U(c_s)\, ds + e^{-\beta t} v^\lambda(X_t, Y_t)$$

obtained by applying an arbitrary policy ν until some intermediate time t and then trading optimally should always lead to a supermartingale and to a martingale if the optimizer is used all along. Between trades—in the policy's "no-trade region"—this means that the absolutely continuous drift should be nonpositive and zero for the optimizer. After taking into account (3.5), using integration by parts and cancelling the common factor $e^{-\beta t}$, this leads to

$$\begin{aligned} 0 = \sup_{c>0} \Big(& -\beta v^\lambda(x, y) + U(c) + (rx - c) v_x^\lambda(x, y) \\ & + \mu \cdot \mathbf{D}_y v^\lambda(x, y) + \frac{1}{2} \mathrm{Tr}[\sigma\sigma^\top \mathbf{D}_{yy}] v^\lambda(x, y) \Big). \end{aligned} \tag{3.6}$$

By definition, the value function can only be decreased by admissible bulk trades at any time; so

$$0 \geq \sup_{m \in \mathbb{R}^d} \big(v^\lambda(x - m \cdot \mathbf{1}_d - \lambda, y + m) - v^\lambda(x, y) \big),$$

and this inequality should become an equality for the optimal transaction once the boundaries of the no-trade region are breached. Combining this with (3.6) and switching the sign yields the *dynamic programming equation*

$$0 = \min\big(\beta v^\lambda - \widetilde{U}(v_x^\lambda) - \mathscr{L} v^\lambda, v^\lambda - \mathbf{M} v^\lambda\big), \tag{3.7}$$

where $\widetilde{U}(\tilde{c}) = \sup_{c>0}(U(c) - c\tilde{c})$ is the convex dual of the utility function U, the differential operator $\mathscr{L}$ is defined as

$$\mathscr{L} = rx \frac{\partial}{\partial x} + \mu \cdot \mathbf{D}_y + \frac{1}{2} \mathrm{Tr}[\sigma\sigma^\top \mathbf{D}_{yy}],$$

and $\mathbf{M}$ denotes the nonlocal *intervention operator*

$$\mathbf{M}\psi(x, y) = \sup_{m \in \mathbb{R}^d} \{\psi(x', y') : (x', y') = (x - m \cdot \mathbf{1}_d - \lambda, y + m) \in \mathrm{K}_\lambda\}. \tag{3.8}$$

3.3 Identifying the correct scalings

The next step is to determine heuristically how the optimal no-trade region around the frictionless solution and the corresponding utility loss should scale with a *small* transaction cost λ. This can be done by adapting the heuristic argument in [22, 39]. Indeed, the welfare effect of any trading cost is composed of two parts, namely the direct costs incurred due to actual trades, and the displacement loss due to having to deviate from the frictionless optimum. Since the frictionless value function is locally quadratic around its maximum, Taylor's theorem suggests that the displacement effect should be of order x^2 for any small cost that only causes a small displacement x.

Where the various cost structures differ is in the losses due to actual trades. Proportional transaction costs lead to trading of local-time type, which scales with the inverse of the width of the no-trade region [22, Sect. 3]. This leads to a total welfare loss proportional to

$$Cx^2 + \lambda/x$$

for some constant $C > 0$. Minimizing this expression leads to a no-trade region with width of order $\lambda^{1/3}$ and a corresponding welfare loss of order $\lambda^{2/3}$. In contrast, trades of all sizes are penalized alike by fixed costs. This leads to a bulk trade to the optimal frictionless position, and therefore a transaction cost of λ, whenever the boundaries of the no-trade region are reached. On the short time interval before leaving a narrow no-trade region, any diffusion resembles a Brownian motion at the leading order. Hence, the first exit time can be approximated by the one of a Brownian motion from the interval $[-x, x]$, which scales with x^2. After the subsequent jump to the midpoint of the no-trade region, this procedure is repeated, so that the number of trades approximately scales with $1/x^2$. As a result, the total welfare loss due to small fixed costs λ is proportional to

$$Cx^2 + \lambda/x^2$$

for some constant $C > 0$. Minimizing this expression in x then leads to an optimal no-trade region of order $\lambda^{1/4}$ and a corresponding welfare loss of order $\lambda^{1/2}$.

3.4 Derivation of the corrector equations

In view of the previous considerations, we expect the leading-order utility loss due to small transaction costs λ to be of order $\lambda^{1/2}$, whereas the deviations of the optimal policy from its frictionless counterpart should be of order $\lambda^{1/4}$. This motivates for the asymptotic expansion of the transaction cost value function the ansatz

$$v^\lambda(x, y) = v(z) - \lambda^{1/2}u(z) - \lambda w(z, \xi) + o(\lambda^{3/4}). \tag{3.9}$$

Here, v is the frictionless value function from Sect. 3.1, the functions u and w are to be calculated, and we change variables from the safe and risky positions x, y to the total wealth

$$z := x + y \cdot \mathbf{1}_d$$

and the deviations

$$\xi := \big(y - \theta(x + y)\big)/\lambda^{1/4}$$

of the risky positions from their frictionless targets, normalized to be of order $O(1)$ as $\lambda \downarrow 0$. The function λw is included, even though it only contributes at the higher order λ itself because its second derivatives with respect to the y-variables are of order $\lambda^{1/2}$.

To determine u and w, insert the postulated expansion (3.9) into the dynamic programming equation (3.7). This leads to two separate equations in the no-trade and trade regions, respectively.

3.4.1 No-trade region

To ease notation, we illustrate the calculations for the case of a single risky asset ($d=1$) and merely state the multidimensional results at the end.[9] In the no-trade region, we have to expand the elliptic operator from (3.7) in powers of λ. To this end, Taylor expansion (3.3), and $\widetilde{U}' = -(U')^{-1}$ yield

$$\widetilde{U}\big(v_x^\lambda(x,y)\big) = \widetilde{U}\big(v_z(z)\big) + \lambda^{1/2}\kappa(z)u_z(z) + o(\lambda^{3/4}).$$

Moreover, also taking into account that $y = \theta(z) + \lambda^{1/4}\xi$, it follows that

$$\begin{aligned}
&\beta v^\lambda(x,y) - \widetilde{U}\big(v_x^\lambda(x,y)\big) - \mathscr{L}v^\lambda(x,y)\\
&\quad = \beta v(z) - \widetilde{U}\big(v_z(z)\big) - \mathscr{L}_0 v(z)\\
&\qquad - \lambda^{1/4}\xi\big(\mu v_z(z) + \sigma^2\theta(z)v_{zz}(z)\big)\\
&\qquad - \lambda^{1/2}\Big(\beta u(z) - \mathscr{L}_0 u(z) + \kappa(z)u_z(z) + \frac{\sigma^2}{2}\xi^2 v_{zz}(z)\\
&\qquad\qquad - \frac{\sigma^2}{2}\theta(z)^2\big(1-\theta_z(z)\big)^2 w_{\xi\xi}(z,\xi)\Big)\\
&\qquad + o(\lambda^{1/2})
\end{aligned}$$

for the differential operator $\mathscr{L}_0$ from (3.2). The $O(\lambda^{1/4})$-terms in this expression vanish by definition (3.4) of the frictionless optimal weight; the same holds for the $O(1)$-terms by the frictionless dynamic programming equation (3.1). Satisfying the elliptic part of (3.7) between bulk trades—at the leading order $O(\lambda^{1/2})$—is therefore tantamount to

$$\begin{aligned}
0 &= \beta u(z) - \mathscr{L}_0 u(z) + \kappa(z)u_z(z)\\
&\quad + \frac{\sigma^2}{2}\xi^2 v_{zz}(z) - \frac{\sigma^2}{2}\theta(z)^2\big(1-\theta_z(z)\big)^2 w_{\xi\xi}(z,\xi).
\end{aligned} \tag{3.10}$$

3.4.2 Trade region

Now, turn to the second part of the frictional dynamic programming equation (3.7), which should vanish when a bulk trade becomes optimal outside the no-trade region. Suppose that $(x,y)\in \mathrm{K}_\lambda$ and $v^\lambda(x,y) = \mathbf{M}v^\lambda(x,y)$. Then, inserting the expansion for v^λ yields

$$v(z) - \lambda^{1/2}u(z) - \lambda w(z,\xi) = v(z-\lambda) - \lambda^{1/2}u(z-\lambda) - \lambda\cdot\inf_{\hat{\xi}} w(z-\lambda,\hat{\xi}),$$

[9] The full multidimensional derivation can be found in [43]. In the no-trade region, the calculations are identical.

where the infimum is over deviations $\hat{\xi}$ attainable from the current position (z, ξ) by a single trade. Taylor expansion yields

$$0 = \lambda\big(v_z(z) - w(z, \xi) + \inf_{\hat{\xi}} w(z - \lambda, \hat{\xi})\big) + o(\lambda).$$

If $w(z, \xi) = w(z - \lambda, \xi) + o(\lambda)$, where $o(\lambda)$ only depends on z,[10] this simplifies to

$$0 = \lambda\big(v_z(z) - w(z, \xi) + \inf_{\hat{\xi}} w(z, \hat{\xi})\big) + o(\lambda).$$

In the ansatz (3.9), the function w is multiplied by a higher-order λ-term. Therefore, its value at a particular point is irrelevant at the leading order $\lambda^{1/2}$, and we may assume that $w(z, 0) = 0$. As a result, we expect that

$$\inf_{\hat{\xi}} w(z, \hat{\xi}) = w(z, 0) = 0$$

because a zero deviation $\xi = 0$ from the frictionless position should lead to the smallest utility loss. Consequently, the leading-order dynamic programming equation outside the no-trade region reads as

$$0 = v_z(z) - w(z, \xi). \tag{3.11}$$

Note that this derivation remains valid for several risky assets.

3.4.3 Corrector equations

Together with (3.10), (3.11) shows that—at the leading order $\lambda^{1/2}$—the dynamic programming equation (3.7) can be written as

$$\max\Big(\mathscr{A}u(z) + \frac{\sigma^2}{2}\xi^2 v_{zz}(z) - \frac{\sigma^2}{2}\theta^2(z)\big(1 - \theta_z(z)\big)^2 w_{\xi\xi}(z, \xi), w(z, \xi) - v_z(z)\Big)$$
$$= 0, \tag{3.12}$$

where we set

$$\mathscr{A}u(z) := \beta u(z) - \mathscr{L}_0 u(z) + \kappa(z) u_z(z). \tag{3.13}$$

To solve (3.12), we first treat the z-variable as constant and solve (3.12) as a function of ξ only to get

$$0 = \max\Big(\frac{\sigma^2}{2}\xi^2 v_{zz}(z) - \frac{\sigma^2}{2}\theta^2(z)\big(1 - \theta_z(z)\big)^2 w_{\xi\xi}(z, \xi) + a(z), w(z, \xi) - v_z(z)\Big)$$

for some $a(z)$ that only depends on z but not on ξ. Then, take $a(z)$ as given and solve for the function u of z to get

$$\mathscr{A}u(z) = a(z).$$

[10] This will turn out to be consistent with the results of our calculations below; see Sects. 3.5 and 3.6.

If both of these "corrector equations" are satisfied, then (3.12) evidently holds as well. For several risky assets, the corresponding analogues read as follows.

Definition 3.1 (Corrector equations) For a given $z > 0$, the *first corrector equation* for the unknown pair $(a(z), w(z, \cdot)) \in \mathbb{R}_+ \times C^2(\mathbb{R}_+)$ is

$$\max\Big(-\frac{|\sigma^\top \xi|^2}{2}\big(-v_{zz}(z)\big) - \frac{1}{2}\operatorname{Tr}[\alpha(z)\alpha(z)^\top w_{\xi\xi}] + a(z),\ w(z,\xi) - v_z(z)\Big)$$
$$= 0, \qquad \forall \xi \in \mathbb{R}^d, \tag{3.14}$$

together with the normalization $w(z, 0) = 0$, where

$$\alpha(z) := \big(I_d - \theta_z(z)\mathbf{1}_d^\top\big)\operatorname{diag}[\theta(z)]\sigma.$$

The *second corrector equation* uses the function $a(z)$ from the first corrector equation and is a simple linear equation for the function $u : \mathbb{R}_+ \to \mathbb{R}$; it reads

$$\mathcal{A}u(z) = a(z) \qquad \forall\, z \in \mathbb{R}_+, \tag{3.15}$$

where $\mathcal{A}$, defined in (3.13) and (3.2), is the infinitesimal generator of the optimal wealth process for the frictionless problem.

Remark 3.2 As for proportional costs [43, Remark 3.3], the first corrector equation is the dynamic programming equation of an ergodic control problem. Indeed, for fixed z and for an increasing sequence of stopping times $\tau = (\tau_k)_{k\in\mathbb{N}}$ and impulses $m = (m_k)_{k\in\mathbb{N}} \in \mathbb{R}^d$, we define the cost functional by

$$J(z, m, \tau) := v_z(z) \limsup_{T\to\infty} \frac{1}{T}\, \mathbb{E}\bigg[\int_0^T \frac{(-v_{zz}(z))}{2v_z(z)} |\sigma^\top \xi_s|^2\, ds + \sum_{k=1}^{\infty} 1_{\{\tau_k \le T\}}\bigg],$$

where the state process ξ is given by

$$\xi_t^i = \xi_0^i + \sum_{j=1}^{d} \alpha^{i,j}(z) B_t^j + \sum_{k=1}^{\infty} m_k 1_{\{\tau_k \le t\}}, \quad t \ge 0, i = 1, \dots, d,$$

with a d-dimensional standard Brownian motion B.

The structure of this problem implies that the optimal strategy is determined by a region $\mathcal{C}$ enclosing the origin. The optimal stopping times are the hitting times of ξ to the boundary of $\mathcal{C}$. When ξ hits $\partial\mathcal{C}$, it is optimal to move it to the origin. Hence, the optimal stopping times (τ_k) are the hitting times of ξ to the boundary of $\mathcal{C}$ and $m_k = -\xi_{\tau_k-}$, so that $\xi_{\tau_k} = 0$ for each $k = 1, 2, \dots$. Put differently, the region $\mathcal{C}$ provides the asymptotic shape of the no-trade region. In the power and log utility case, it is an ellipsoid as in Fig. 2.

The function a is the optimal value,

$$a(z) := \inf_{(\tau,m)} J(z, m, \tau).$$

Then, the Feynman–Kac formula for the linear equation $\mathscr{A}u = a$ for u implies

$$u(z) = \mathbb{E}\left[\int_0^\infty e^{-\beta t}\, a(Z_t^{m,z})\, dt\right],$$

where $Z^{m,z}$ is the optimal wealth process for the frictionless Merton problem with initial value $Z_0^{m,z} = z$.

3.5 Solution in one dimension

If there is only a single risky asset ($d = 1$), the asymptotically optimal no-trade region is the interval $\{z : |\xi| \leq \xi_0(z)\}$. The first corrector equation can then be readily solved explicitly by imposing smooth pasting at the boundaries, similarly as for proportional transaction costs [43]. Matching values and first derivatives across the trading boundaries $\pm\xi_0(z)$ leads to two conditions for a symmetric function $w(z, \cdot)$, in addition to the actual optimality equation in the interior of the no-trade region. Thus, the lowest-order polynomials in ξ capable of fulfilling these requirements are of order four. Since we have imposed $w(z, 0) = 0$, this motivates the ansatz

$$w(z,\xi) = \begin{cases} A(z)\xi^2 - B(z)\xi^4, & |\xi| \leq \xi_0(z), \\ v_z(z), & |\xi| \geq \xi_0(z). \end{cases}$$

Inside the no-trade region, inserting this ansatz into the first corrector equation (3.14) gives

$$0 = a(z) + \frac{\sigma^2}{2}\xi^2 v_{zz}(z) - \alpha^2(z)A(z) + 6\alpha^2(z)B(z)\xi^2,$$

where $\alpha(z) := \sigma\theta(z)(1 - \theta_z(z))$ as in Definition 3.1. Since this equation should be satisfied for any value of ξ, comparison of coefficients yields

$$B(z) = \frac{-\sigma^2 v_{zz}(z)}{12\alpha^2(z)}, \quad A(z) = \frac{a(z)}{\alpha^2(z)}. \tag{3.16}$$

Next, the smooth pasting condition $0 = w_\xi(z, \xi_0(z)) = 2A(z)\xi_0(z) - 4B(z)\xi_0^3(z)$ at the trading boundary $\xi = \xi_0(z)$ implies

$$\xi_0^2(z) = \frac{A(z)}{2B(z)}. \tag{3.17}$$

Finally, the value matching conditions $v_z(z) = A(z)\xi_0^2(z) - B(z)\xi_0^4(z)$ evaluated at $\xi = \xi_0(z)$ give

$$v_z(z) = \frac{A^2(z)}{4B(z)} = -\frac{3a^2(z)}{\alpha^2(z)\sigma^2 v_{zz}(z)},$$

and in turn

$$a(z) = v_z(z)\alpha(z)\sigma\sqrt{-\frac{v_{zz}(z)}{3v_z(z)}}. \tag{3.18}$$

In view of (3.17) and (3.16), the optimal trading boundaries are therefore determined as

$$\xi_0(z) = \left(\frac{12}{-v_{zz}(z)/v_z(z)} \theta(z)^2 \big(1-\theta_z(z)\big)^2 \right)^{1/4}. \tag{3.19}$$

For utilities with constant relative risk aversion $\gamma > 0$, the optimal frictionless risky position is $\theta(z) = \pi_m z$, so that the corresponding trading boundaries are given by

$$\xi_0(z) = \left(\frac{12}{\gamma} \pi_m^2 (1-\pi_m)^2 z^3 \right)^{1/4}.$$

For the maximal deviations of the risky weight from the frictionless target, this yields the formulas from Theorem 2.4, namely

$$\pi_0(z) = \frac{\lambda^{1/4}\xi_0(z)}{z} = \left(\frac{12}{\gamma} \pi_m^2 (1-\pi_m)^2 \frac{\lambda}{z} \right)^{1/4}.$$

With constant relative risk aversion, the homotheticity of the value function (2.2) and (3.18) imply that the second corrector equation $\mathscr{A}u(z) = a(z)$ simplifies to

$$\begin{aligned} &\beta u(z) - rzu_z - \frac{(\mu-r)^2}{\gamma\sigma^2} zu_z(z) - \frac{(\mu-r)^2}{2\gamma^2\sigma^2} z^2 u_{zz}(z) + c_m z u_z \\ &= \sqrt{\frac{\gamma}{3}} c_m^{-\gamma} \sigma^2 \pi_m (1-\pi_m) z^{1/2-\gamma}, \end{aligned}$$

which is solved by

$$u(z) = u_0 z^{1/2-\gamma} \quad \text{with } u_0 = \sigma^2 \left(\frac{\gamma}{3} \pi_m^2 (1-\pi_m)^2 \right)^{1/2} \frac{c_m(\gamma)^{-\gamma}}{c_m(2\gamma)}.$$

This is the formula from Theorem 2.3.

3.6 Solution in higher dimensions

Let us now turn to the solution of the corrector equations for multiple risky assets. To ease the already heavy notation, we restrict ourselves to utilities U_γ with constant relative risk aversion $\gamma > 0$ here. Then we can rescale the corrector equation to obtain a version that is independent of the wealth variable z. Indeed, let

$$\rho = z^{-3/4}\xi,$$

so that setting

$$v_0 = c_m^{-\gamma},$$

we obtain

$$w(z,\xi) = v_z(z) W(z^{-3/4}\xi) = v_0 z^{-\gamma} W(\rho), \quad a(z) = a_0 z^{1/2-\gamma} > 0$$

for some constant $a_0 > 0$ and a function $W(\rho)$ to be determined. We also introduce the matrices

$$A := z^{-2}\alpha(z)\alpha(z)^\top, \quad \Sigma := \sigma\sigma^\top.$$

Then a direct computation shows

$$\begin{aligned}|\sigma^\top\xi|^2 v_{zz}(z) &= (-\Sigma\rho\cdot\rho)\frac{v_0 z^{1/2-\gamma}}{\gamma},\\ \mathrm{Tr}[\alpha(z)\alpha(z)^\top w_{\xi\xi}(z,\xi)] &= \mathrm{Tr}[AW_{\rho\rho}(\rho)](v_0 z^{1/2-\gamma}).\end{aligned}$$

The resulting rescaled equation for the pair $(W(\cdot), a_0)$, with independent variable $\rho \in \mathbb{R}^d$, is

$$\max\Big(-\frac{1}{2}\Sigma\rho\cdot\rho - \frac{1}{2}\mathrm{Tr}[AW_{\rho\rho}(\rho)] + a_0, -1 + W(\rho)\Big) = 0, \tag{3.20}$$

together with the normalization $W(0) = 0$. Following Atkinson and Wilmott [2], we postulate a solution of the form

$$W^*(\rho) = 1 - (M\rho\cdot\rho - 1)^2$$

for a symmetric matrix M to be computed. Then,

$$W^*_{\rho\rho}(\rho) = -4(M\rho\cdot\rho - 1)M - 8M\rho\otimes M\rho.$$

Hence,

$$\begin{aligned}-\frac{1}{2}\mathrm{Tr}[AW^*_{\rho\rho}(\rho)] &= 2(M\rho\cdot\rho - 1)\,\mathrm{Tr}[AM] + 4MAM\rho\cdot\rho\\ &= (2M\,\mathrm{Tr}[AM] + 4MAM)\rho\cdot\rho - 2\,\mathrm{Tr}[AM]\\ &= \frac{1}{2}\Sigma\rho\cdot\rho - a_0,\end{aligned}$$

provided that $a_0 = 2\,\mathrm{Tr}[AM]$ and M solves the algebraic Riccati equation

$$4M\,\mathrm{Tr}[AM] + 8MAM = \Sigma. \tag{3.21}$$

Remarkably, this is exactly Eq. (3.7) obtained by Atkinson and Wilmott [2] in their asymptotic analysis of the Morton and Pliska model [34] with trading costs equal to a constant fraction of the investors' current wealth. Atkinson and Wilmott [2] argue that one may take A to be the identity without any loss of generality by transforming to a coordinate system in which the second-order operator is the Laplacian. For the convenience of the reader, we provide this transformation here: since A is symmetric positive definite by Assumption (2.3), there is a unitary matrix $O \in \mathbb{R}^{d\times d}$ for which

$$OAO^\top = \mathrm{diag}[\zeta_i],$$

where $\zeta_1, \zeta_2, \ldots, \zeta_d$ denote the eigenvalues of A. Setting

$$\tilde{M} := \operatorname{diag}[\zeta_i^{1/2}] O M O^\top \operatorname{diag}[\zeta_i^{1/2}],$$
$$\tilde{\Sigma} := \operatorname{diag}[\zeta_i^{1/2}] O \Sigma O^\top \operatorname{diag}[\zeta_i^{1/2}],$$

(3.21) becomes

$$4\tilde{M}\operatorname{Tr}[\tilde{M}] + 8\tilde{M}^2 = \tilde{\Sigma}.$$

Using that $\tilde{M}$ and $\tilde{\Sigma}$ have the same eigenvectors, Atkinson and Wilmott (see (3.8)–(3.11) in [2]) obtain simple algebraic equations for the eigenvalues of $\tilde{M}$, thus determining M up to the above coordinate transformation. In summary, given A, Σ positive definite, there exists a positive definite solution M of (3.21). Then, a solution to the corrector equation (3.14) is given by the function

$$W(\rho) := \begin{cases} 1 - (\rho^\top M\rho - 1)^2 & \text{for } \rho \in \mathscr{I}, \\ 1 & \text{for } \rho \notin \mathscr{I}, \end{cases}$$

where $\mathscr{I}$ is the ellipsoid around zero given by

$$\mathscr{I} := \{\rho \in \mathbb{R}^d \ : \ \rho^\top M\rho < 1\}.$$

Reverting to the original variables, it follows that the asymptotically optimal no-trade region should be given by

$$\mathrm{NT}^\lambda = \left\{(x, y) \in \mathbb{R}^{d+1} : \frac{y}{x + y\cdot \mathbf{1}_d} \in \pi_m + \frac{\lambda^{1/4}}{(x + y\cdot \mathbf{1}_d)^{1/4}}\mathscr{I}\right\},$$

in accordance with Theorem 2.4.

Remark 3.3 The following notation will be useful. Given a wealth $z > 0$, define

$$\mathrm{NT}(z) := \{\xi \in \mathbb{R}^d \ : \ \xi \in z^{3/4}\mathscr{I}\}.$$

That is, for any $(z, \xi) = (x + y\cdot \mathbf{1}_d, \frac{y - \pi_m(x + y\cdot \mathbf{1}_d)}{\lambda})$ corresponding to $(x, y) \in \mathrm{K}_\lambda$, we have $(x, y) \in \mathrm{NT}^\lambda$ if and only if $\xi \in \mathrm{NT}(z)$.

4 Existence of the relaxed semi-limits

In the sequel, we turn the previous heuristics into rigorous proofs of our main results, Theorems 2.3 and 2.4, using the general methodology developed by Barles and Perthame [4] and Evans [14] in the context of viscosity solutions. To ease notation by avoiding fractional powers, we write

$$\lambda = \epsilon^4$$

and, with a slight abuse of notation, use a sub- or superscript ϵ to refer to objects pertaining to the transaction cost problem. For instance, v^ϵ refers to v^λ, K_ϵ to K_λ, and so on.

To establish the expansion of the value function asserted in Theorem 2.3, we need to show that

$$u^\epsilon(x,y) = \frac{v(z) - v^\epsilon(x,y)}{\epsilon^2}$$

is locally uniformly bounded from above as $\epsilon \to 0$. To this end, define the *relaxed semi-limits*

$$u_*(x_0,y_0) = \liminf_{\substack{(\epsilon,x,y)\to(0,x_0,y_0)\\(x,y)\in \mathrm{K}_\epsilon}} u^\epsilon(x,y), \quad u^*(x_0,y_0) = \limsup_{\substack{(\epsilon,x,y)\to(0,x_0,y_0)\\(x,y)\in \mathrm{K}_\epsilon}} u^\epsilon(x,y). \quad (4.1)$$

Their existence is guaranteed by the straightforward lower bound $u^\epsilon \geq 0$ and the locally uniform upper bound provided further in Theorem 4.1. Establishing the latter involves an explicit construction of a particular trading strategy and is addressed first. We then show in Sects. 5 and 6 that the relaxed semi-limits u^*, u_* are viscosity sub- and supersolutions, respectively, of the second corrector equation (3.15). Combined with the comparison result in Theorem 7.2 for the second corrector equation provided in Sect. 7, this in turn yields that $u^* \leq u_*$. Since the opposite inequality is satisfied by definition, it follows that $u = u^* = u_*$ is the unique solution of the second corrector equation (3.15). As a consequence, $u^\epsilon \to u$ locally uniformly, verifying the asymptotic expansion of the value function. With the latter at hand, we can in turn verify that the policy from Theorem 2.4 is indeed almost optimal for small costs (see Sect. 8).

4.1 Locally uniform upper bound of u^ϵ

In this section, we show that $u^\epsilon(x,y) = \epsilon^{-2}(v(z) - v^\epsilon(x,y))$ is locally uniformly bounded from above as $\epsilon \to 0$.

Theorem 4.1 *Given any x_0, y_0 with $x_0 + y_0 \cdot \mathbf{1}_d > 0$, there exist $\epsilon_0 > 0$ and $r_0 = r_0(x_0,y_0) > 0$ such that*

$$\sup\{u^\epsilon(x,y) :\ (x,y) \in B_{r_0}(x_0,y_0), \epsilon \in (0,\epsilon_0]\} < \infty. \quad (4.2)$$

Theorem 4.1 is an immediate corollary of Lemma 4.6. To prove the latter, we construct an investment–consumption policy that gives rise to a suitable upper bound. The construction necessitates some technical estimates. The reader can simply read the definition of the strategy and proceed directly to the proof of Lemma 4.6 in order to view the thread of the argument.

4.1.1 Strategy up to a stopping time θ

Given an initial portfolio allocation $(X_{0-}, Y_{0-}) \in \mathrm{K}_\epsilon$, use the trading strategy from Theorem 2.4, corresponding to the no-trade region NT^ϵ, from time 0 until a stopping time θ to be defined further.

More specifically, let $(\tau_1, \tau_2, \dots; m_1, m_2, \dots)$, where τ_i is the ith time the portfolio process hits the boundary $\partial \mathrm{NT}^\epsilon$ of the no-trade region. The corresponding reallocations $m_1, m_2, \dots$ are chosen so that after taking into account transaction costs, the portfolio process is at the frictionless Merton proportions, that is,

$$\frac{Y^\epsilon_{\tau_i}}{X^\epsilon_{\tau_i} + Y^\epsilon_{\tau_i} \cdot \mathbf{1}_d} = \pi_m.$$

Until θ, the investor consumes the optimal frictionless proportion of her current wealth,

$$c_t = c_m Z^\epsilon_t \quad \forall t \le \theta,$$

so that her wealth process is governed until time θ by the stochastic differential equations

$$\begin{aligned} X^\epsilon_t &= X_{0-} + \int_0^t (r X^\epsilon_s - c_s)\, ds - \sum_{k=1}^\infty \left(\epsilon^4 + \sum_{j=1}^d m^j_k \right) 1_{\{\tau_k \le t\}}, \\ Y^\epsilon_t &= Y_{0-} + \int_0^t Y^\epsilon_s \frac{dS_s}{S_s} + \sum_{k=1}^\infty m_k 1_{\{\tau_k \le t\}}. \end{aligned} \tag{4.3}$$

The stopping time θ must be chosen so that the investor's position remains solvent at all times, that is, $(X^\epsilon_t, Y^\epsilon_t) \in \mathrm{K}_\epsilon$ $\forall t \le \theta$ P-almost surely. Therefore, we use the first time the investor's wealth falls below some threshold, which needs to be large enough to permit the execution of a final liquidating bulk trade.

4.1.2 At time θ and beyond

Define $\theta = \theta^{\eta,\epsilon}$ to be the exit time of the portfolio process from the set

$$\begin{aligned} K^{\eta,\epsilon} := \big\{ (z,\xi) \in \mathbb{R}_+ \times \mathbb{R}^d : &\text{ either } z > (\eta+1)\epsilon^4 \text{ and } \xi \in \mathbb{R}^d \\ &\text{ or } z \in \big(\eta\epsilon^4, (\eta+1)\epsilon^4\big] \text{ and } \xi \in \mathrm{NT}(z) \big\}. \end{aligned}$$

Within $K^{\eta,\epsilon}$, the above policy is used, and the portfolio process follows (4.3). At time θ, the investor liquidates all risky assets, leading to a safe position of at least $(\eta-1)\epsilon^4$. Afterward, she consumes at half the interest rate, thereby remaining solvent forever. The resulting portfolio process satisfies a deterministic integral equation with stochastic initial data, namely

$$X^\epsilon_{\theta+t} = X^\epsilon_\theta + \int_0^t \frac{r}{2} X^\epsilon_{\theta+s}\, ds, \quad Y^\epsilon_{\theta+t} = 0, \quad t \ge 0. \tag{4.4}$$

Let $(X^{\eta,\epsilon}_t, Y^{\eta,\epsilon}_t)$, $t \ge 0$, be the portfolio produced by concatenating the controlled stochastic process (4.3) and the deterministic process (4.4) at time θ.

Remark 4.2 For any $\eta > 1$, the optimal value $v^\epsilon(\eta\epsilon^4, \xi)$ must be greater than or equal to the utility obtained from the immediate liquidation of all risky assets and then running the deterministic policy (4.4). Since the latter can be computed explicitly, this provides a crude lower bound for $v^\epsilon(\eta\epsilon^4, \xi)$.

To see this, suppose the investor's wealth after the liquidating trade at time θ is given by $X^\epsilon_\theta \geq (\eta-1)\epsilon^4$. Then $X^\epsilon_{\theta+t} \geq (\eta-1)\epsilon^4 e^{\frac{r}{2}t}$. For power utilities ($0 < \gamma \neq 1$), this yields the lower bound

$$v^\epsilon(\eta\epsilon^4, \xi) \geq \int_0^\infty e^{-\beta t} \frac{(\eta-1)^{1-\gamma}\epsilon^{4-4\gamma} e^{\frac{r}{2}(1-\gamma)t}}{1-\gamma}\, dt$$
$$= \frac{(\eta-1)^{1-\gamma}}{(1-\gamma)(\beta - \frac{r}{2}(1-\gamma))}\epsilon^{4-4\gamma}.$$

The corresponding result for logarithmic utility ($\gamma = 1$) is

$$v^\epsilon(\eta\epsilon^4, \xi) \geq \int_0^\infty e^{-\beta t} \log\left((\eta-1)\epsilon^4 e^{rt/2}\right) dt = \frac{\log((\eta-1)\epsilon^4)}{\beta} + \frac{r}{2\beta^2}. \tag{4.5}$$

4.1.3 Constructing a candidate lower bound

For given $\epsilon, \delta, C > 0$, define the function

$$V^{\epsilon,\delta}_C(z, \xi) = v(z) - \epsilon^2 C u(z) - \epsilon^4(1+\delta) w(z, \xi).$$

We now establish a series of technical lemmas. These will be used in the proof of Lemma 4.6 to verify that, asymptotically, $V^{\epsilon,\delta}_C$ is dominated by the value function v^ϵ in the no-trade region for an appropriate choice of the parameters C and δ.

Lemma 4.3 *Let $\eta > 1$ be given. There exists $C_\eta > 0$, independent of ϵ, such that for all $\bar{z} \in [\eta\epsilon^4, (\eta+1)\epsilon^4]$, we have*

$$v^\epsilon(\bar{z}, \xi) \geq V^{\epsilon,\delta}_{C_\eta}(\bar{z}, \xi) \quad \textit{for all } \xi \in \mathbb{R}.$$

Proof We only consider power utilities ($\gamma \neq 1$); the case of logarithmic utility can be treated similarly. First, notice that since the term $-\epsilon^4(1+\delta)w(\bar{z}, \xi)$ is always negative, it can be ignored. Write $\bar{z} = (\eta + \bar{\lambda})\epsilon^4$ for some $\bar{\lambda} \in [0, 1]$. Using the estimates from Remark 4.2, the goal is to find a sufficiently large C_η so that

$$\frac{(\eta - 1 + \bar{\lambda})^{1-\gamma}}{(1-\gamma)(\beta - \frac{r}{2}(1-\gamma))}\epsilon^{4-4\gamma} \geq v(\bar{z}) - C_\eta \epsilon^2 u(\bar{z})$$
$$= (\eta + \bar{\lambda})^{1-\gamma}\left(\frac{v_0}{1-\gamma} - C_\eta(\eta+\bar{\lambda})^{-1/2}u_0\right)\epsilon^{4-4\gamma}$$

for all $\bar{\lambda} \in [0, 1]$. This follows by observing that we can take

$$C_\eta := \text{const} \cdot \sqrt{\eta} \tag{4.6}$$

for a large enough positive constant that only depends on the model and preference parameters $(\mu, r, \sigma, \gamma, \beta)$ but is independent of ϵ and η. $\square$

Lemma 4.4 *There exists $\delta > 0$ with the property that for all η sufficiently large, there is $\epsilon_0 = \epsilon_0(\eta, \gamma) > 0$ such that*

$$V_{C_\eta}^{\epsilon,\delta}(z - \epsilon^4, 0) - V_{C_\eta}^{\epsilon,\delta}(z, \bar{\xi}) \geq 0, \quad \forall \epsilon \in (0, \epsilon_0],\ z \geq \eta\epsilon^4,\ \bar{\xi} \in \partial\mathrm{NT}(z).$$

Proof Recall that by definition the corrector w satisfies $w(\cdot, 0) = 0$ and $w(z, \bar{\xi}) = v_z(z)$ for $\bar{\xi} \in \partial\mathrm{NT}(z)$.

First, consider the case of power utility. Taylor expansion and evaluation at $z = \eta\epsilon^4$ yield

$$\begin{aligned}
&V_{C_\eta}^{\epsilon,\delta}(z - \epsilon^4, 0) - V_{C_\eta}^{\epsilon,\delta}(z, \bar{\xi}) \\
&\quad = v(z - \epsilon^4) - v(z) - \epsilon^2 C_\eta\big(u(z - \epsilon^4) - u(z)\big) + \epsilon^4(1 + \delta)v_z(z) \\
&\quad = \delta\epsilon^4 v_z(z) + \epsilon^6 C_\eta u'(z) - \epsilon^8 v_{zz}(\tilde{z}_1) + \epsilon^{10} C_\eta u''(\tilde{z}_2) \\
&\quad = \epsilon^{4-4\gamma}\Big(v_0(\delta\eta^{-\gamma} + \gamma\tilde{\eta}_1^{-1-\gamma}) \\
&\qquad\qquad + C_\eta u_0(1/2 - \gamma)\big(\eta^{-1/2-\gamma} + (1/2 + \gamma)\tilde{\eta}_2^{-3/2-\gamma}\big)\Big),
\end{aligned} \tag{4.7}$$

where the points $\tilde{z}_1, \tilde{z}_2 \in [z - \epsilon^4, z]$ are determined by the Taylor remainders of v and u, respectively, and $\tilde{\eta}_1, \tilde{\eta}_2 \in [\eta - 1, \eta]$ satisfy $\tilde{z}_i = \tilde{\eta}_i\epsilon^4$. Considering (4.7) as a function of η, the dominant term is of order $O(\eta^{-\gamma})$. Since $C_\eta = C\sqrt{\eta}$, where C only depends on the model parameters $(\mu, r, \sigma, \gamma, \beta)$, the term $C_\eta\eta^{-1/2-\gamma}$ also contributes at the order $O(\eta^{-\gamma})$. Consequently, choosing

$$\delta > \frac{Cu_0|1/2 - \gamma|}{v_0} \tag{4.8}$$

ensures that the leading-order coefficient is positive, independently of η. For sufficiently large η, the assertion follows.

In the case of logarithmic utility ($\gamma = 1$), the argument is the same because the expressions involved are all power-type functions. The same choice (4.8) of δ works as well. $\square$

Lemma 4.5 *For sufficiently large η, there exists $\epsilon_0 = \epsilon_0(\eta, \gamma) > 0$ such that*

$$\beta V_{C_\eta}^{\epsilon,\delta}(z, \xi) - \mathscr{L} V_{C_\eta}^{\epsilon,\delta}(z, \xi) \leq U(c_m z), \quad \forall \epsilon \in (0, \epsilon_0],\ z \geq \eta\epsilon^4,\ \xi \in \mathrm{NT}(z),$$

where δ is given by (4.8).

Proof We consider only the power utility case since the argument also works mutatis mutandis for logarithmic utility. To ease notation, we write V^ϵ instead of $V_{C_\eta}^{\epsilon,\delta}$.

Throughout the proof, $(x,y)\in \mathrm{K}_\epsilon$ satisfies $z=x+y\cdot\mathbf{1}_d=\eta\epsilon^4$. Decompose

$$\begin{aligned}\beta V^\epsilon(x,y)-\mathscr{L}V^\epsilon(x,y)&=\big(\beta v(z)-\mathscr{L}v(z)\big)-\epsilon^2 C_\eta\big(\beta u(z)-\mathscr{L}u(z)\big)\\&\quad-\epsilon^4(1+\delta)\big(\beta w(z,\xi)-\mathscr{L}w(z,\xi)\big)\\&=: I_1(z)-I_2(z)-I_3(z,\xi).\end{aligned}$$

We analyze the asymptotic properties in η of each of the terms I_1, I_2, and I_3. First,

$$\begin{aligned}I_1(z)&=-\frac{1}{2}\epsilon^2|\sigma^\top\xi|^2 v_{zz}(z)+\tilde{U}\big(v_z(z)\big)\\&=-\frac{1}{2}\epsilon^2|\sigma^\top\xi|^2 v_{zz}(z)+U(c_m z)-c_m z v_z(z)\\&\le-\frac{1}{2}\epsilon^2|\sigma^\top\xi|^2 v_{zz}(z)+U(c_m z)-\epsilon^2 C_\eta c_m z u_z(z)\\&\le\epsilon^{4-4\gamma}O(\eta^{1/2-\gamma})+U(c_m z)-\epsilon^2 C_\eta c_m z u_z(z).\end{aligned}$$

Here, we used that $v_z(z)\ge\epsilon^2 C_\eta u_z(z)$ if η is sufficiently large. The estimates in Remark A.2 and the fact that C_η is of order $O(\sqrt{\eta})$ (see (4.6)) give

$$\begin{aligned}I_2(z)+\epsilon^2 C_\eta c_m z u_z(z)&=C_\eta\bigg(\epsilon^2\mathcal{A}u(z)-\epsilon^3\xi\cdot(\mu-r\mathbf{1}_d)u_z(z)\\&\qquad-\epsilon^3\Big(\frac{1}{2}\epsilon|\sigma^\top\xi|^2-\sigma^\top\xi\cdot\sigma^\top\pi_m z\Big)u_{zz}(z)\bigg)\\&\ge C_\eta\big(\epsilon^2 a_0 z^{1/2-\gamma}-K(\epsilon^3 z^{1/4-\gamma}+\epsilon^4 z^{-\gamma})\big)\\&=\epsilon^{4-4\gamma}O(\eta^{1-\gamma}),\end{aligned}$$

where $a_0=a(z)/z^{1/2-\gamma}$. Hence, this term is positive for sufficiently large η. Finally, by Remark A.2 we have

$$\begin{aligned}|I_3(z,\xi)|&\le(1+\delta)K(\epsilon^2 z^{1/2-\gamma}+\epsilon^3 z^{1/4-\gamma}+\epsilon^4 z^{-\gamma}+\epsilon^5 z^{-1/4-\gamma}+\epsilon^6 z^{-1/2-\gamma})\\&=\epsilon^{4-4\gamma}O(\eta^{1/2-\gamma}),\end{aligned}$$

again for all sufficiently large η. In summary,

$$\begin{aligned}\beta V^\epsilon(x,y)-\mathscr{L}V^\epsilon(x,y)&=I_1(z)-I_2(z)-I_3(z,\xi)\\&\le U(c_m z)+\epsilon^{4-4\gamma}\big(O(\eta^{1/2-\gamma})-O(\eta^{1-\gamma})\big)\\&\le U(c_m z)\end{aligned}$$

for sufficiently large η. Equivalently, there exists some $\eta>1$ such that for all $z\ge\eta\epsilon^4$ and $\xi\in\mathrm{NT}(z)$,

$$\beta V^\epsilon(z,\xi)-\mathscr{L}V^\epsilon(z,\xi)\le U(c_m z)\quad\forall\epsilon\in(0,\epsilon_0].$$

This completes the proof. □

4.1.4 Proof of Theorem 4.1

We now have all the ingredients to prove the main result of this section in the next lemma, which in turn yields Theorem 4.1 as a corollary.

Lemma 4.6 *There are constants $C, \delta, \epsilon_0 > 0$ such that for all $\epsilon \in (0, \epsilon_0]$,*

$$v(z) - \epsilon^2 Cu(z) - \epsilon^4 (1+\delta) w(z,\xi) \leq v^\epsilon(z,\xi) \quad \forall (z,\xi) \in \mathrm{K}_\epsilon. \tag{4.9}$$

In particular,

$$u^\epsilon(z,\xi) \leq Cu(z) + o(\epsilon) \quad \forall (z,\xi) \in \mathrm{K}_\epsilon, \tag{4.10}$$

so that (4.2) is satisfied.

Proof Let $(x,y) \in \mathrm{K}_\epsilon$ be given, and let $\eta > 1$ be large enough so that all the previous lemmas are applicable. Without loss of generality, we may assume that $x + y > \eta\epsilon^4$ since we are proving an asymptotic result.

Step 1: Let $(X_t, Y_t) := (X_t^{\eta,\epsilon,x}, Y_t^{\eta,\epsilon,y})$ be the controlled portfolio process with dynamics (4.3), (4.4) that starts from the initial allocation (x,y) and switches to deterministic consumption at half the interest rate at the first time $\theta := \theta^\eta$ the total wealth $Z_t := X_t + Y_t$ falls to level $z = \eta\epsilon^4$. As before, write

$$V^\epsilon(z,\xi) := V_{C_\eta}^{\epsilon,\delta_\eta}(z,\xi).$$

Recall that C_η and δ_η are given by (4.6) and (4.8), respectively. Itô's formula yields

$$\begin{aligned} e^{-\beta\theta} V^\epsilon(X_\theta, Y_\theta) = V^\epsilon(x,y) &- \int_0^\theta e^{-\beta s}\big(\beta V^\epsilon(X_s,Y_s) - \mathscr{L} V^\epsilon(X_s,Y_s)\big)\,ds \\ &+ \int_0^\theta e^{-\beta s} \mathbf{D}_y V^\epsilon(X_s,Y_s)^\top \sigma\, dW_s \\ &+ \sum_{t\leq\theta} e^{-\beta t}\big(V^\epsilon(Z_t,\xi_t) - V^\epsilon(Z_{t-},\xi_{t-})\big). \end{aligned}$$

Observe that the summation is at most countable and that in view of Lemma 4.4, each summand satisfies

$$V^\epsilon(Z_t,\xi_t) - V^\epsilon(Z_{t-},\xi_{t-}) \geq 0.$$

Together with Lemma 4.5, this yields

$$\begin{aligned} &e^{-\beta\theta} V^\epsilon(X_\theta, Y_\theta) \\ &\quad\geq V^\epsilon(x,y) - \int_0^\theta e^{-\beta s} U(c_m Z_s)\,ds + \int_0^\theta e^{-\beta s} \mathbf{D}_y V^\epsilon(X_s,Y_s)^\top \sigma\, dW_s. \end{aligned} \tag{4.11}$$

Step 2: For any $(x',y') \in \mathrm{K}_\epsilon$ with $0 < x' + y' \cdot \mathbf{1}_d \leq \eta\epsilon^4$, let $\nu^{(x',y')} \in \Theta_\epsilon(x',y')$ be the strategy of (4.4), that is, liquidation of all risky assets and then deterministic consumption at half the risk-free rate forever. According to Remark 4.2 and the proof of Lemma 4.3,

$$v^\epsilon(X_\theta, Y_\theta) \geq J(\nu^{(X_\theta, Y_\theta)}) \geq V^\epsilon(X_\theta, Y_\theta) \quad \text{on } \{\theta < \infty\},$$

where

$$J(\nu) := \mathbb{E}\left[\int_0^\infty e^{-\beta t} U(c_t)\, dt\right] \quad \text{for any } \nu = (c, \tau, m).$$

Let $(\tau_n)_{n\geq 0}$ be a localizing sequence of stopping times for the local martingale term in (4.11) and set $\theta_n := \theta \wedge \tau_n$. Let us assume for the moment that the family $(e^{-\beta\theta_n} V^\epsilon(X_{\theta_n}, Y_{\theta_n}))_{n\geq 0}$ is uniformly integrable, and therefore it converges in expectation to its pointwise limit. Then the same applies to the integral of the dt-term in (4.11) by dominated convergence. Taking expectations in (4.11), sending $n \to \infty$, and using these observations together with Lemmas 4.3 and 4.5 it follows that

$$\begin{aligned} V^\epsilon(x, y) &\leq \mathbb{E}\left[\int_0^\theta e^{-\beta s} U(c_m Z_s)\, ds + e^{-\beta\theta} V^\epsilon(X_\theta, Y_\theta)\right] \\ &\leq \mathbb{E}\left[\int_0^\theta e^{-\beta s} U(c_m Z_s)\, ds + e^{-\beta\theta} \int_0^\infty e^{-\beta t} U(Z_\theta e^{rt/2})\, dt\right] \\ &\leq v^\epsilon(x, y) \quad \forall \epsilon \in (0, \epsilon_0]. \end{aligned}$$

Since x, y were arbitrary, assertion (4.9) follows.

Step 3: All that remains to show is that $(e^{-\beta\theta_n} V^\epsilon(X_{\theta_n}, Y_{\theta_n}))_{n\geq 0}$ is uniformly integrable. Since the functions and domains are explicit, we can check that there is a constant $M > 0$, independent of ϵ, n, such that

$$|e^{-\beta\theta_n} V^\epsilon(X_{\theta_n}, Y_{\theta_n})| \leq M |e^{-\beta\theta_n} v(Z_{\theta_n})|.$$

Hence, it is sufficient to show that $(e^{-\beta\theta_n} v(Z_{\theta_n}))_{n\geq 0}$ is uniformly integrable. This will follow, for instance, if it is uniformly bounded in $L^{1+q}(\mathbb{P})$ for some $q > 0$. The interesting case is $0 < \gamma \leq 1$; otherwise, $v(z)$ is bounded on the domain under consideration because the wealth process is bounded away from zero and the Merton value function is negative. We just show the power utility case; a similar argument applies for logarithmic utility.

Let $\tilde{Z}_t := \tilde{Z}_t^{\eta,\epsilon,x+y}$ denote the same controlled wealth process as in Step 1, but obtained by not deducting transaction costs or consumption. Evidently, $Z_\tau^{1-\gamma} \leq \tilde{Z}_\tau^{1-\gamma}$ almost surely for any stopping time τ. Moreover, for any $a, b > 0$, we have

$$\begin{aligned} &\frac{d(e^{-a\beta\theta_n} (\tilde{Z}_{\theta_n})^b)}{e^{-a\beta\theta_n} \tilde{Z}_{\theta_n}^b} \\ &\quad = \left(-a\beta + b\left(r + \pi_t \cdot (\mu - r\mathbf{1}_d) + \frac{b-1}{2}|\sigma^\top \pi_t|^2\right)\right) dt + b\pi_t \cdot \sigma\, dW_t, \end{aligned}$$

where $\pi_t := Y_t/Z_t$. When $a = 1$ and $b = 1-\gamma$, the drift term is maximized at the Merton proportion, $\pi_t \equiv \pi_m$, and moreover, by the finiteness criterion for the frictionless value function, we have

$$-\beta + (1-\gamma)\Big(r + \pi_m \cdot (\mu - r\mathbf{1}_d) - \frac{1}{2}|\sigma^\top \pi_m|^2\Big) < 0.$$

Taking $b = (1-\gamma)(1+q)$ and $a = (1+q)$ for sufficiently small $q > 0$, the drift term is maximized at a vector $\pi^{a,b}$ arbitrarily close to π_m, for which

$$A := -a\beta + b\Big(r + \pi^{a,b} \cdot (\mu - r\mathbf{1}_d) - \frac{1}{2}|\sigma^\top \pi^{a,b}|^2\Big) < 0.$$

As a consequence,

$$\frac{d(e^{-a\beta\theta_n}(\tilde{Z}_{\theta_n})^b)}{e^{-a\beta\theta_n}\tilde{Z}^b_{\theta_n}} \le \bigg(- a\beta + b\Big(r + \pi^{a,b} \cdot (\mu - r\mathbf{1}_d) + \frac{b-1}{2}|\sigma^\top \pi^{a,b}|^2\Big)\bigg) dt + b\pi_t \cdot \sigma\, dW_t.$$

Taking expectations, passing to the limit over a localizing sequence of stopping times for the local martingale term, and applying Fatou's lemma, we obtain

$$\big\|e^{-\beta\theta_n}(\tilde{Z}_{\theta_n})^{1-\gamma}\big\|^{1+q}_{L^{1+q}(\mathbb{P})} = \mathbb{E}[e^{-a\beta\theta_n}(\tilde{Z}_{\theta_n})^b] \le \tilde{Z}^b_0 \mathbb{E}[\exp(A\theta_n)] \le z^b \quad \forall n \in \mathbb{N}.$$

Hence, the family is uniformly bounded in $L^{1+q}(\mathbb{P})$ and thus uniformly integrable. □

4.1.5 Relaxed limits are functions of wealth only

We conclude this section by establishing that the relaxed semi-limits u_*, u^* only depend on total wealth and can be realized by restricting to limits taken on the Merton line.

Lemma 4.7 *For any $x_0 + y_0 \cdot \mathbf{1}_d > 0$, we have*

$$u_*(x_0, y_0) = \liminf_{\substack{(\epsilon,x,y)\to(0,x_0,y_0)\\ (x,y)\in \mathrm{K}_\epsilon}} u^\epsilon(z - \pi_m \cdot \mathbf{1}_d z, \pi_m z),$$

$$u^*(x_0, y_0) = \limsup_{\substack{(\epsilon,x,y)\to(0,x_0,y_0)\\ (x,y)\in \mathrm{K}_\epsilon}} u^\epsilon(z - \pi_m \cdot \mathbf{1}_d z, \pi_m z).$$

Proof Given $(x, y) \in \mathrm{K}_\epsilon$, where $z = x + y \cdot \mathbf{1}_d > \epsilon^4$ without loss of generality, we observe that

$$\inf_{x'+y'\cdot\mathbf{1}_d = z+\epsilon^4} v^\epsilon(x', y') \ge v^\epsilon(x, y) \ge \sup_{x'+y'\cdot\mathbf{1}_d = z-\epsilon^4} v^\epsilon(x', y'). \tag{4.12}$$

Therefore,

$$\frac{v(z)-v^\epsilon(x,y)}{\epsilon^2} \le \epsilon^{-2}\Big(v(z-\epsilon^4) - \sup_{x'+y'\cdot \mathbf{1}_d = z-\epsilon^4} v^\epsilon(x',y')\Big) + \epsilon^2 v_z(z-\epsilon^4)$$
$$= \inf_{x'+y'\cdot \mathbf{1}_d = z-\epsilon^4} u^\epsilon(x',y') + \epsilon^2 v_z(z-\epsilon^4)$$

and

$$\inf_{x'+y'\cdot \mathbf{1}_d = z-\epsilon^4} u^\epsilon(x',y') + \epsilon^2 v_z(z-\epsilon^4) \ge u^\epsilon(x,y) \tag{4.13}$$
$$\ge \sup_{x'+y'\cdot \mathbf{1}_d = z+\epsilon^4} u^\epsilon(x',y') - \epsilon^2 v_z(z+\epsilon^4).$$

Let $(\epsilon_n, x_n, y_n) \to (0, x_0, y_0)$ be chosen such that $u^{\epsilon_n}(x_n, y_n) \to u_*(x_0, y_0)$. Setting $z_n' = x_n + y_n \cdot \mathbf{1}_d + \epsilon^4$ and using the previous observations, it follows that

$$u^{\epsilon_n}(x_n, y_n) \ge u^{\epsilon_n}(z_n' - \pi_m \cdot \mathbf{1}_d z_n', \pi_m z_n') - O(\epsilon^2).$$

Taking lim inf as $n \to \infty$ on both sides yields

$$u_*(x_0, y_0) = u_*(z_0 - \pi_m \cdot \mathbf{1}_d z_0, \pi_m z_0),$$

where $z_0 = x_0 + y_0 \cdot \mathbf{1}_d$. The proof for u^* is similar. □

Remark 4.8 For later use, observe that

$$u_*(x_0,y_0) = \liminf_{\substack{(\epsilon,x,y)\to(0,x_0,y_0)\\ (x,y)\in K_\epsilon}} \underline{u}^\epsilon(x,y), \qquad u^*(x_0,y_0) = \limsup_{\substack{(\epsilon,x,y)\to(0,x_0,y_0)\\ (x,y)\in K_\epsilon}} \overline{u}^\epsilon(x,y),$$

where $\underline{u}^\epsilon, \overline{u}^\epsilon$ are the lower- and upper-semicontinuous envelopes of u^ϵ, respectively. Moreover, (4.13) extends to the envelopes in the form

$$\overline{u}^\epsilon(x,y) \le \inf_{x'+y'\cdot \mathbf{1}_d = z-\epsilon^4} \overline{u}^\epsilon(x',y') + \epsilon^2 v_z(z-\epsilon^4), \tag{4.14}$$
$$\underline{u}^\epsilon(x,y) \ge \sup_{x'+y'\cdot \mathbf{1}_d = z+\epsilon^4} \underline{u}^\epsilon(x',y') - \epsilon^2 v_z(z+\epsilon^4).$$

5 Viscosity subsolution property of u^*

Theorem 5.1 *The function u^* is a viscosity subsolution of the second corrector equation* (3.15).

Proof Let $(z_0, \varphi) \in (0,\infty) \times C^2(\mathbb{R}_+)$ with

$$0 = (u^* - \varphi)(z_0) > (u^* - \varphi)(z) \quad \forall z > 0, z \ne z_0.$$

To prove the assertion, we have to show that

$$\mathcal{A}\varphi(z_0) \le a(z_0).$$

Step 1: By Lemma 4.6 there exist $\epsilon_0, r > 0$ depending on (x_0, y_0) such that

$$b^* := \sup_{(x,y)\in B_r, \epsilon\in(0,\epsilon_0]} \overline{u}^\epsilon(x, y) < \infty, \qquad B_r := B_r(x_0, y_0).$$

The radius r can be taken small enough so that B_r does not intersect the line $z = 0$. By Lemma 4.7, $u^*(z_0)$ can be achieved along a sequence $(z^\epsilon, 0)$ on the Merton line, that is,

$$z^\epsilon \to z_0 \quad \text{and} \quad \overline{u}^\epsilon(z^\epsilon, 0) \to u^*(z_0) \qquad \text{as } \epsilon \to 0.$$

Observe that

$$\ell^*_\epsilon := \overline{u}^\epsilon(z^\epsilon) - \varphi(z^\epsilon) \to 0$$

and

$$(x^\epsilon, y^\epsilon) := (z^\epsilon - \pi_m \cdot \mathbf{1}_d z^\epsilon, \pi_m z^\epsilon) \to (x_0, y_0) := (z_0 - \pi_m \cdot \mathbf{1}_d z_0, \pi_m z_0).$$

Due to the strict maximality of $u^* - \varphi$ at z_0, each z^ϵ can be taken to be a maximizer of $\overline{u}^\epsilon(\cdot, 0) - \varphi(\cdot)$ on $[z_0 - r, z_0 + r]$. For $\epsilon \in (0, \epsilon_0]$ and $\delta > 0$, set

$$\psi^{\epsilon,\delta}(z, \xi) := v(z) - \epsilon^2\big(\varphi(z) + \ell^*_\epsilon + C(z - z^\epsilon)^4\big) - \epsilon^4(1 + \delta)w(z, \xi)$$

with $C > 0$ to be chosen later.

Step 2: Now we use the function $\psi^{\epsilon,\delta}$ to touch $\underline{v}^\epsilon$ from below near (x_0, y_0). Set

$$I^{\epsilon,\delta}(z, \xi) := (\underline{v}^\epsilon - \psi^{\epsilon,\delta})(z, \xi).$$

Consider any point (z, ξ) corresponding to $(x, y) \in B_r$. We have

$$\begin{aligned}\epsilon^{-2} I^\epsilon(z, \xi) &= -\overline{u}^\epsilon(z, \xi) + \varphi(z) + \ell^*_\epsilon + C(z - z^\epsilon)^4 + \epsilon^2(1 + \delta)w(z, \xi) \\ &\ge -b^* + \varphi(z) + \ell^*_\epsilon + C(z - z^\epsilon)^4. \end{aligned} \tag{5.1}$$

Thus, $C > 0$ can be chosen large enough to ensure that (5.1) is positive for all sufficiently small $\epsilon > 0$ when $r > |z - z_0| > r/2$.

Next, we show that $I^{\epsilon,\delta}(z, \xi) > 0$ when $|z - z_0| < r$ and $\xi \notin \mathrm{NT}(z)$. To this end, observe that by Taylor expansion, (4.14), and the maximizer property of z^ϵ we have

$$\begin{aligned}\epsilon^{-2} I^{\epsilon,\delta}(z, \xi) &= -\overline{u}^\epsilon(z, \xi) + \varphi(z) + \ell^*_\epsilon + C(z - z^\epsilon)^4 + \epsilon^2(1 + \delta)v_z(z) \\ &\ge -\overline{u}^\epsilon(z - \epsilon^4, 0) + \varphi(z - \epsilon^4) + \ell^*_\epsilon - \epsilon^2 v_z(z - \epsilon^4) + \epsilon^2(1 + \delta)v_z(z) \\ &\quad + O(\epsilon^4) \\ &\ge \delta\epsilon^2 v_z(z) + O(\epsilon^4) \\ &> 0\end{aligned}$$

for all sufficiently small $\epsilon > 0$. Using $I^{\epsilon,\delta}(z^\epsilon, 0) = 0$, we deduce that $I^{\epsilon,\delta}$ attains a local minimum at some point $(\tilde{z}^\epsilon, \tilde{\xi}^\epsilon)$ with $|z_0 - \tilde{z}^\epsilon| < r$ and $\tilde{\xi}^\epsilon \in \mathrm{NT}(\tilde{z}^\epsilon)$ for all $\epsilon > 0$ sufficiently small.

Step 3: Now we derive some limiting identities. Since according to the previous argument, $|\tilde{\xi}^\epsilon|$ is uniformly bounded in ϵ, there is a convergent subsequence $(\tilde{z}^{\epsilon_n}, \tilde{\xi}^{\epsilon_n}) \to (\hat{z}, \hat{\xi})$, where $\hat{z} > 0$ and $\hat{\xi} \in \mathbb{R}$. Then,

$$0 \geq \liminf_{n\to\infty} \epsilon_n^{-2} I^{\epsilon_n,\delta}(\tilde{z}^{\epsilon_n}, \tilde{\xi}^{\epsilon_n}) = -\limsup_{n\to\infty} \overline{u}^\epsilon(\tilde{z}^{\epsilon_n}, \tilde{\xi}^{\epsilon_n}) + \varphi(\hat{z}) + C(\hat{z} - z_0)^4$$

by construction. Moreover,

$$-\limsup_{n\to\infty} \overline{u}^\epsilon(\tilde{z}^{\epsilon_n}, \tilde{\xi}^{\epsilon_n}) + \varphi(\hat{z}) + C(\hat{z} - z_0)^4 \geq -u^*(\hat{z}) + \varphi(\hat{z}) + C(\hat{z} - z_0)^4 \geq 0.$$

So in fact, the inequalities must all be equalities. The strict maximality property of $u^* - \varphi$ at z_0 in turn gives $\hat{z} = z_0$ and $\hat{\xi} \in \overline{\mathrm{NT}(z_0)}$. Having chosen a particular subsequence, we may without loss of generality write ϵ instead of ϵ_n. Using that $\underline{v}^\epsilon$ is a supersolution of the dynamic programming equation (3.7), we obtain

$$\begin{aligned}
0 &\leq \epsilon^{-2}\big(\beta \underline{v}^\epsilon - \mathscr{L}\psi^{\epsilon,\delta} - \tilde{U}(\psi_x^{\epsilon,\delta})\big)(\tilde{z}^\epsilon, \tilde{\xi}^\epsilon) \\
&\leq \epsilon^{-2}\big(\beta \psi^{\epsilon,\delta} - \mathscr{L}\psi^{\epsilon,\delta} - \tilde{U}(\psi_x^{\epsilon,\delta})\big)(\tilde{z}^\epsilon, \tilde{\xi}^\epsilon) \\
&= \frac{(\beta v - \mathscr{L} v - \tilde{U}(v_x))(\tilde{z}^\epsilon, \tilde{\xi}^\epsilon)}{\epsilon^2} + \frac{\tilde{U}(v_x) - \tilde{U}(\psi_x^{\epsilon,\delta})}{\epsilon^2} \\
&\quad - (\beta - \mathscr{L})\big(\varphi(\tilde{z}^\epsilon) + \ell_\epsilon^* + C(\tilde{z}^\epsilon - z^\epsilon)^4\big) \\
&\quad - \epsilon^2(1+\delta)\big(\beta w(\tilde{z}^\epsilon, \tilde{\xi}^\epsilon) - \mathscr{L} w(\tilde{z}^\epsilon, \tilde{\xi}^\epsilon)\big) \\
&=: I_1^\epsilon + I_2^\epsilon - I_3^\epsilon - I_4^\epsilon.
\end{aligned}$$

As $\epsilon \to 0$, we have

$$\begin{aligned}
&I_1^\epsilon \longrightarrow -\tfrac{1}{2}|\sigma^\top \hat{\xi}|^2 v_{zz}(z_0); && I_2^\epsilon \longrightarrow \tilde{U}'\big(v_z(z_0)\big)\varphi_z(z_0) = -c_m z_0 \varphi_z(z_0); \\
&I_3^\epsilon \longrightarrow \mathcal{A}\varphi(z_0) - c_m z_0 \varphi_z(z_0); && I_4^\epsilon \longrightarrow -\tfrac{1}{2}(1+\delta)\,\mathrm{Tr}[\alpha(z_0)\alpha(z_0)^\top w_{\xi\xi}(z_0, \hat{\xi})].
\end{aligned}$$

Step 4: Combining these limits with $\hat{\xi} \in \overline{\mathrm{NT}(z_0)}$ yields the inequality

$$\begin{aligned}
0 &\leq -\frac{1}{2}|\sigma^\top \hat{\xi}|^2 v_{zz}(z_0) - \mathcal{A}\varphi(z_0) + \frac{1}{2}(1+\delta)\,\mathrm{Tr}[\alpha(z_0)\alpha(z_0)^\top w_{\xi\xi}(z_0, \hat{\xi})] \\
&= a(z_0) - \mathcal{A}\varphi(z_0) + \frac{\delta}{2}\,\mathrm{Tr}[\alpha(z_0)\alpha(z_0)^\top w_{\xi\xi}(z_0, \hat{\xi})].
\end{aligned}$$

Finally, letting $\delta \to 0$ and using the local boundedness of $w_{\xi\xi}$ (see Proposition A.3) produce the desired inequality, namely

$$\mathcal{A}\varphi(z_0) \leq a(z_0). \qquad \square$$

6 Viscosity supersolution property of u_*

Theorem 6.1 *u_* is a viscosity supersolution of the second corrector equation* (3.15).

Proof Let $(z_0, \varphi) \in (0, \infty) \times C^2(\mathbb{R})$ be such that

$$0 = (u_* - \varphi)(z_0) < (u_* - \varphi)(z) \quad \forall z > 0,\ z \neq z_0.$$

To prove the assertion, we have to show that

$$\mathcal{A}\varphi(z_0) \geq a(z_0).$$

Step 1: As before, we start by constructing the test function. Recall that

$$\underline{u}^\epsilon(z, \xi) = \frac{v(z) - \overline{v}^\epsilon(z, \xi)}{\epsilon^2},$$

where $\overline{v}^\epsilon$ denotes the upper-semicontinuous envelope of the transaction cost value function v^ϵ. By the definition of the relaxed semilimit u_* and Lemma 4.7 there exists a sequence $(z^\epsilon, 0)$ on the Merton line such that

$$z^\epsilon \to z_0 \quad \text{and} \quad \underline{u}^\epsilon(z^\epsilon, 0) \to u_*(z_0) \qquad \text{as } \epsilon \to 0.$$

Set

$$\ell_\epsilon^* := \underline{u}^\epsilon(z^\epsilon) - \varphi(z^\epsilon) \to 0$$

and

$$(x^\epsilon, y^\epsilon) := (z^\epsilon - \pi_m \cdot \mathbf{1}_d z^\epsilon, \pi_m z^\epsilon) \to (x_0, y_0) := (z_0 - \pi_m \cdot \mathbf{1}_d z_0, \pi_m z_0).$$

We localize by choosing $r > 0$ such that $B_r := B_r(x_0, y_0)$ does not intersect the line $z = 0$. Define

$$\psi^{\epsilon,\delta}(z, \xi) = v(z) - \epsilon^2\big(\varphi(z) + \ell_\epsilon^* - C(z - z^\epsilon)^4\big) - \epsilon^4(1 - \delta)w(z, \xi),$$

where $C > 0$ is chosen so large that

$$I^{\epsilon,\delta}(z, \xi) := \overline{v}^\epsilon(z, \xi) - \psi^{\epsilon,\delta}(z, \xi) < 0, \quad \text{where } |z - z_0| \geq r/2,\ \epsilon \in (0, \epsilon_0].$$

Then for all $\epsilon \in (0, \epsilon_0]$, we have

$$\sup_{\substack{|z - z_0| < r/2 \\ \xi \in \mathbb{R}^d}} I^{\epsilon,\delta}(z, \xi) < \infty, \tag{6.1}$$

and by construction,

$$I^{\epsilon,\delta}(z^\epsilon, 0) = 0.$$

Step 2: A priori, there is no reason that the supremum in (6.1) should be achieved at any particular point, let alone that a maximizing sequence should converge as we send ϵ to zero. As a way out, we perturb the original test function to complete the localization. To this end, fix ϵ, δ and let (z_n, ξ_n) be a maximizing sequence of $I^{\epsilon,\delta}$. (Keep in mind that this sequence depends on ϵ, δ.) Set

$$\alpha_n := \frac{3}{2}\left(\sup_{\substack{|z-z_0|<r/2\\ \xi\in\mathbb{R}^d}} I^{\epsilon,\delta}(z,\xi) - I^{\epsilon,\delta}(z_n,\xi_n)\right)$$

and

$$h_n(\xi) = h(\xi - \xi_n),$$

where

$$h(\xi) = \begin{cases} \exp(1 - \frac{1}{1-|\xi|^2}) & \text{if } |\xi| < 1,\\ 0 & \text{if } |\xi| \geq 1.\end{cases}$$

Notice that $\alpha_n \to 0$ as $\epsilon \to 0$. The modified test function is taken to be

$$\psi^{\epsilon,\delta,n}(z,\xi) = \psi^{\epsilon,\delta}(z,\xi) - \alpha_n h_n(\xi),$$

so that

$$I^{\epsilon,\delta,n}(z,\xi) := (v^{\epsilon,\delta} - \psi^{\epsilon,\delta,n})(z,\xi) = I^{\epsilon,\delta}(z,\xi) + \alpha_n h_n(\xi).$$

By construction each $I^{\epsilon,\delta,n}$ has a maximizer, say $(\hat{z}^\epsilon_n, \hat{\xi}^\epsilon_n) \in [z_0 - r, z_0 + r] \times \mathbb{R}^d$. Observe that the rate of decay of α_n with respect to ϵ can be taken to be as fast as we wish by choosing n large enough. We find it convenient to take $\alpha_{n_\epsilon} \leq \frac{1}{n_\epsilon}\exp(-\epsilon^{-1})$, which can always be accomplished by relabeling if necessary.

For any selection $\epsilon \mapsto \hat{z}^\epsilon_{n_\epsilon}$, it turns out that z_0 is the unique subsequential limit of $(\hat{z}^\epsilon_{n_\epsilon})_\epsilon$ as $\epsilon \to 0$. Indeed, note that since $(\hat{z}^\epsilon_{n_\epsilon})_\epsilon \subset [z_0 - r, z_0 + r]$, it contains a convergent subsequence. If $\hat{z}$ is the limit of such a subsequence, then

$$0 \leq \epsilon_k^{-2} I^{\epsilon_k,\delta,n_k}(z^{\epsilon_k}_{n_k}, \xi^{\epsilon_k}_{n_k}),$$

which implies that

$$0 \leq -u_*(\hat{z}) + \varphi(\hat{z}) - C(\hat{z} - z_0)^4.$$

By the strict minimality of $u_* - \varphi$ at z_0 we must have $\hat{z} = z_0$.

Step 3: Next, we show that any sequence of maximizers $(\hat{z}^{\epsilon_k}_{n_k}, \hat{\xi}^{\epsilon_k}_{n_k})$ of $I^{\epsilon_k,\delta,n_k}$ where $\epsilon_k \to 0$ as $k \to \infty$, satisfying $\alpha_{n_k} \leq \frac{1}{k}\exp(-\epsilon_k^{-1})$, is asymptotically contained in the no-trade region, that is,

$$\underline{(\psi^{\epsilon_k,\delta,n_k} - \mathbf{M}\psi^{\epsilon_k,\delta,n_k})}(\hat{z}^{\epsilon_k}_{n_k}, \hat{\xi}^{\epsilon_k}_{n_k}) > 0$$

for all sufficiently large k. We proceed to show this by contradiction. First, note that by Lemma B.1, $\psi^{\epsilon_k,\delta,n_k} - \mathbf{M}\psi^{\epsilon_k,\delta,n_k}$ is already lower-semicontinuous. Therefore, suppose that

$$0 \geq \psi^{\epsilon_k,\delta,n_k}(\hat{z}_n^{\epsilon_k}, \hat{\xi}_n^{\epsilon_k}) - \mathbf{M}\psi^{\epsilon_k,\delta,n_k}(\hat{z}_n^{\epsilon_k}, \hat{\xi}_n^{\epsilon_k})$$
$$= \psi^{\epsilon_k,\delta,n_k}(\hat{z}_n^{\epsilon_k}, \hat{\xi}_n^{\epsilon_k}) - \psi^{\epsilon_k,\delta,n_k}(\hat{z}_n^{\epsilon_k} - \epsilon^4, \tilde{\xi}_n^{\epsilon_k}) \tag{6.2}$$

for some $\tilde{\xi}_n^{\epsilon_k}$. Such a point exists by the construction of $\psi^{\epsilon_k,\delta,n_k}$. Therefore, we deduce that

$$\begin{aligned}
0 &\geq \psi^{\epsilon_k,\delta,n_k}(\hat{z}_n^{\epsilon_k}, \hat{\xi}_n^{\epsilon_k}) - \psi^{\epsilon_k,\delta,n_k}(\hat{z}_n^{\epsilon_k} - \epsilon^4, \tilde{\xi}_n^{\epsilon_k}) \\
&= \epsilon_k^4 v_z(\hat{z}_n^{\epsilon_k}) + O(\epsilon_k^6) - \epsilon_k^4(1-\delta)\big(w(\hat{z}_n^{\epsilon_k}, \hat{\xi}_n^{\epsilon_k}) - w(\hat{z}_n^{\epsilon_k} - \epsilon_k^4, \tilde{\xi}_n^{\epsilon_k})\big) \\
&\quad - \epsilon_k^2 \alpha_{n_k}\big(h_{n_k}(\tilde{\xi}_n^{\epsilon_k}) - h_{n_k}(\hat{\xi}_n^{\epsilon_k})\big) \\
&\geq \delta \epsilon_k^4 v_z(\hat{z}_n^{\epsilon_k}) + O(\epsilon_k^6) - \frac{2}{k}\epsilon_k^2 \exp(-\epsilon_k^{-1}) \\
&> 0
\end{aligned}$$

for all sufficiently large k because $\delta > 0$. This contradicts (6.2).

By the subsolution property of $\overline{v}^{\epsilon_k}$ at $(\hat{z}_{n_k}^{\epsilon_k}, \hat{\xi}_{n_k}^{\epsilon_k})$, for which we now write $(\hat{z}^{\epsilon_k}, \hat{\xi}^{\epsilon_k})$, we obtain the differential inequality

$$\begin{aligned}
0 &\geq \big(\beta \overline{v}^{\epsilon_k} - \mathscr{L}\psi^{\epsilon_k,\delta,n_k} - \tilde{U}(\psi_x^{\epsilon_k,\delta,n_k})\big)(\hat{z}^{\epsilon_k}, \hat{\xi}^{\epsilon_k}) \\
&\geq \big(\beta \psi^{\epsilon_k,\delta,n_k} - \mathscr{L}\psi^{\epsilon_k,\delta,n_k} - \tilde{U}(\psi_x^{\epsilon_k,\delta,n_k})\big)(\hat{z}^{\epsilon_k}, \hat{\xi}^{\epsilon_k}).
\end{aligned}$$

Step 4: We claim that $|\hat{\xi}^{\epsilon}|$ is uniformly bounded in $\epsilon \in (0, \epsilon_0]$. Expanding the above differential inequality into powers of ϵ_k leads to

$$\begin{aligned}
0 &\geq \epsilon_k^{-2}\big(\beta \psi^{\epsilon_k,\delta,n_k} - \mathscr{L}\psi^{\epsilon_k,\delta,n_k} - \tilde{U}(\psi_x^{\epsilon_k,\delta,n_k})\big)(\hat{z}^{\epsilon_k}, \hat{\xi}^{\epsilon_k}) \\
&= -\frac{1}{2}|\sigma^\top \hat{\xi}^{\epsilon_k}|^2 v_{zz}(\hat{z}^{\epsilon_k}) - \alpha_{n_k}\big(\beta h_{n_k}(\hat{\xi}^{\epsilon_k}) - \mathscr{L}h_{n_k}(\hat{\xi}^{\epsilon_k})\big) \\
&\quad - \Big(\beta\big(\varphi(\hat{z}^{\epsilon_k}) + \ell_{\epsilon_k}^* - C(\hat{z}^{\epsilon_k} - z^{\epsilon_k})^4\big) - \mathscr{L}\big(\varphi(\hat{z}^{\epsilon_k}) + \ell_{\epsilon_k}^* - C(\hat{z}^{\epsilon_k} - z^{\epsilon_k})^4\big)\Big) \\
&\quad - \epsilon_k^2(1-\delta)\big(\beta w(\hat{z}^{\epsilon_k}, \hat{\xi}^{\epsilon_k}) - \mathscr{L}w(\hat{z}^{\epsilon_k}, \hat{\xi}^{\epsilon_k})\big) - \frac{\tilde{U}(\psi_x^{\epsilon_k,\delta,n_k}) - \tilde{U}(v_x)}{\epsilon_k^2}.
\end{aligned}$$

We proceed to estimate each term. To this end, let $K = K(\beta, \mu, \sigma, r, \gamma) > 0$ denote a sufficiently large generic constant. By Proposition A.3, we have

$$\begin{aligned}
&\epsilon_k^2(1-\delta)\big(\beta w(\hat{z}^{\epsilon_k}, \hat{\xi}^{\epsilon_k}) - \mathscr{L}w(\hat{z}^{\epsilon_k}, \hat{\xi}^{\epsilon_k})\big) \\
&\quad = -(1-\delta)\frac{1}{2}\operatorname{Tr}[\alpha(\hat{z}^{\epsilon_k})\alpha(\hat{z}^{\epsilon_k})^\top w_{\xi\xi}(\hat{z}^{\epsilon_k}, \hat{\xi}^{\epsilon_k})] + (1-\delta)\epsilon \mathcal{R}_w(\hat{z}^{\epsilon_k}, \hat{\xi}^{\epsilon_k}) \\
&\quad \leq -(1-\delta)\frac{1}{2}\operatorname{Tr}[\alpha(\hat{z}^{\epsilon_k})\alpha(\hat{z}^{\epsilon_k})^\top w_{\xi\xi}(\hat{z}^{\epsilon_k}, \hat{\xi}^{\epsilon_k})] + K(1 + \epsilon_k|\hat{\xi}^{\epsilon_k}| + \epsilon_k^2|\hat{\xi}^{\epsilon_k}|^2),
\end{aligned}$$

and

$$\alpha_{n_k}|\beta h_{n_k}(\hat{\xi}^{\epsilon_k}) - \mathscr{L} h_{n_k}(\hat{\xi}^{\epsilon_k})| \le \frac{K}{k}\exp(-\epsilon_k^{-1})(\epsilon_k^{-2} + \epsilon_k^{-1}|\hat{\xi}^{\epsilon_k}| + |\hat{\xi}^{\epsilon_k}|^2).$$

Moreover,

$$\left|(\beta - \mathscr{L})\big(\varphi(\hat{z}^{\epsilon_k}) + \ell^*_{\epsilon_k} - C(\hat{z}^{\epsilon_k} - z^{\epsilon})^4\big)\right| \le (1 + \epsilon_k|\hat{\xi}^{\epsilon_k}| + \epsilon_k^2|\hat{\xi}^{\epsilon_k}|^2).$$

Finally,

$$\epsilon_k^{-2}|\tilde{U}(\psi_x^{\epsilon_k,\delta,n_k}) - \tilde{U}(v_x)| \le K.$$

We therefore conclude that

$$0 \ge -\frac{1}{2}|\sigma^\top \hat{\xi}^{\epsilon_k}|^2 v_{zz}(\hat{z}^{\epsilon_k}) - K(1 + \epsilon_k|\hat{\xi}^{\epsilon_k}| + \epsilon_k^2|\hat{\xi}^{\epsilon_k}|^2).$$

Recalling that $v_{zz} < 0$, it follows that the dominant term $-\frac{1}{2}|\sigma^\top \hat{\xi}^{\epsilon_k}|^2 v_{zz}(\hat{z}^{\epsilon_k})$ is non-negative, and therefore $|\hat{\xi}^{\epsilon_k}|$ must be uniformly bounded in ϵ_k. Hence, along some subsequence, we have $\hat{\xi}^{\epsilon_k} \to \hat{\xi}$ and $\hat{z}^{\epsilon_k} \to z_0$. Sending $\epsilon_k \to 0$ gives

$$\begin{aligned} 0 &\ge -\frac{1}{2}|\sigma^\top \hat{\xi}|^2 v_{zz}(z_0) - \mathcal{A}\varphi(z_0) + \frac{1}{2}(1-\delta)\,\mathrm{Tr}[\alpha(z_0)\alpha(z_0)^\top w_{\xi\xi}(z_0,\hat{\xi})] \\ &= a(z_0) - \mathcal{A}\varphi(z_0) - \frac{\delta}{2}\,\mathrm{Tr}[\alpha(z_0)\alpha(z_0)^\top w_{\xi\xi}(z_0,\hat{\xi})]. \end{aligned}$$

Finally, let $\delta \to 0$. Together with the C^2-estimates on w (see Proposition A.3), it follows that the trace term disappears from the inequality. This yields

$$\mathcal{A}\varphi(z_0) \ge a(z_0),$$

thereby completing the proof. □

7 Comparison for the second corrector equation (3.15)

A straightforward computation shows that $\mathcal{A}z^p = \nu_p z^p$ for some constant $\nu_p \in \mathbb{R}$. If the Merton value function is finite, that is, $c_m > 0$, we readily verify that $\nu_{1/2-\gamma} > 0$. Moreover, since the matrix α from (2.3) is assumed to be invertible, the diffusion coefficient A in (3.20) is positive definite, so that $a(z) = a_0 z^{1/2-\gamma} = 2\,\mathrm{Tr}[AM] z^{1/2-\gamma} > 0$. As a consequence,

$$u(z) z^{-1/2+\gamma} = u_0 = \frac{a_0}{\nu_{1/2-\gamma}} > 0.$$

Remark 7.1 Similarly, if $\delta \in \mathbb{R}$ and $|\delta| \ll 1$, then

$$\mathcal{A}\big(z^\delta u(z)\big) > 0, \quad \forall z > 0. \tag{7.1}$$

This observation is used in the proof of the comparison result in Theorem 7.2.

In view of the explicit locally uniform upper bound (4.10) for u^ϵ from Lemma 4.6, the relaxed semi-limits u^*, u_* satisfy the growth constraint

$$0 \le u^*(z), u_*(z) \le C|z|^{1/2-\gamma}. \tag{7.2}$$

We therefore prove that the second corrector equation satisfies a comparison theorem in the class of nonnegative functions satisfying this growth condition.

Theorem 7.2 *Let $v_1, v_2 : (0,\infty) \to \mathbb{R}$ be nonnegative functions that satisfy the growth constraint (7.2). If*

$$\mathcal{A}v_1 \le a \le \mathcal{A}v_2$$

is satisfied in the viscosity sense, then

$$v_1 \le u \le v_2,$$

where $u(z) = u_0 z^{1/2-\gamma}$.

Proof We just prove that the subsolution is dominated by u; the supersolution part of the assertion follows along the same lines. Let v_1 be a subsolution to $\mathcal{A}v_1 \le a$ satisfying the growth condition (7.2). We proceed by contradiction. Suppose that $v_1(\tilde{z}) > u(\tilde{z})$ for some $\tilde{z} > 0$. We need to distinguish two cases:

Case 1: Suppose that $\gamma \ne 1/2$ and define $I_\delta(z) := v_1(z) - \phi_\delta(z)$, where we set $\phi_\delta(z) := \delta u(z)^{1+\delta} + u(z)$. Then for sufficiently small $\delta > 0$, we have $I_\delta(\tilde{z}) > 0$. The growth conditions imply that I_δ achieves a global maximum at some $z_\delta \in \mathbb{R}_+$ and $I_\delta(z_\delta) > 0$. Invoking the subsolution property of v_1 at z_δ yields $\mathcal{A}\phi_\delta(z_\delta) \le a(z_\delta)$. However, by construction of ϕ_δ and (7.1), $\mathcal{A}\phi_\delta(z_\delta) > a(z_\delta)$, which gives a contradiction.

Case 2: If $\gamma = 1/2$, then use $\phi_\delta(z) := \delta(z^{-\delta} + z^\delta) + u(z)$. The proof then follows as in the first case. □

8 Proof of the main results

We now conclude by proving the main results of the paper.

8.1 Expansion of the value function v^λ

Proof of Theorem 2.3 We have shown in Theorem 4.1 that the relaxed semi-limits u^* and u_* of (4.1) exist, are functions of wealth only by Lemma 4.7, and by Lemma 4.6 satisfy the growth condition

$$0 \le u^*(z), u_*(z) \le C|z|^{1/2-\gamma}.$$

In view of Theorems 5.1 and 6.1, $\mathcal{A}u^* \le a \le \mathcal{A}u_*$ in the viscosity sense. As a result, Theorem 7.2 gives $u^* \le u_*$. The opposite inequality evidently holds by definition; therefore,

$$u_* = u^* = u.$$

The locally uniform convergence claimed in Theorem 2.3 then follows directly from this and from the definitions of u_*, u^*. □

8.2 Almost optimal policy

With the asymptotic expansion from Theorem 2.3 at hand, we can now show that the policy from Theorem 2.4 is almost optimal for small costs. To this end, fix an initial allocation $(x, y) \in \mathrm{K}_\epsilon$ and a threshold $0 < \delta < x + y \cdot \mathbf{1}_d$. Consider the policy $\nu^\epsilon = (c^\epsilon, \tau^\epsilon, m^\epsilon) \in \Theta_\epsilon(x, y)$ from Theorem 2.4. If wealth falls below the threshold, then another strategy is pursued (see Remark 2.5). More precisely, we choose controls $\nu^* = (c^*, \tau^*, m^*) \in \Theta_\epsilon(X_\theta^{\nu^\epsilon,x}, Y_\theta^{\nu^\epsilon,y})$ such that i) $\nu^\epsilon 1_{[0,\theta)} + \nu^* 1_{[\theta,0)} \in \Theta_\epsilon(x, y)$, where $\theta = \theta^{\nu^\epsilon}$ is the first time the wealth process $Z_t = X_t + Y_t \cdot \mathbf{1}_d$ falls below the level δ, and ii) ν^* is $o(\epsilon^2)$-optimal on $[\![\theta, \infty[\![$ for each realization of $(X_\theta^{\nu^\epsilon,x}, Y_\theta^{\nu^\epsilon,y})$. The main technical concern is whether this can be done measurably, but this will follow from a construction similar to the one performed in the proof of the weak dynamic programming principle (B.2).

Let $J^{\epsilon,\delta}(x, y)$ be the corresponding expected discounted utility from consumption,

$$J^{\epsilon,\delta}(x, y) := J(\nu) = \mathbb{E}\left[\int_0^\theta e^{-\beta s} U(c_m Z_s^\epsilon)\, ds + e^{-\beta\theta} \int_0^\infty e^{-\beta s} U(c_s^*)\, ds\right].$$

Then we have the following:

Theorem 8.1 *There exists $\epsilon_\delta > 0$ such that for all $0 < \epsilon \le \epsilon_\delta$,*

$$J^{\epsilon,\delta}(x, y) \ge v(z) - \epsilon^2 u(z) + o(\epsilon^2) \quad \forall z = x + y \cdot \mathbf{1}_d \ge \delta.$$

That is, the policy from Theorem 2.4 is optimal at the leading order ϵ^2.

Proof *Step 1*: Set

$$V^\epsilon(z, \xi) = v(z) - \epsilon^2 u(z) - \epsilon^4(1 + C\epsilon^2) w(z, \xi)$$

for some sufficiently large $C > 0$ to be chosen later. Itô's formula yields

$$\begin{aligned} e^{-\beta\theta} V^\epsilon(X_\theta, Y_\theta) = V^\epsilon(x, y) &+ \int_0^\theta -e^{-\beta s}\big(\beta V^\epsilon(X_s, Y_s) - \mathscr{L} V^\epsilon(X_s, Y_s)\big)\, ds \\ &+ \int_0^\theta e^{-\beta s} \mathbf{D}_y V^\epsilon(X_s, Y_s)^\top \sigma\, dW_s \\ &+ \sum_{t \le \theta} e^{-\beta t}\big(V^\epsilon(Z_t, \xi_t) - V^\epsilon(Z_{t-}, \xi_{t-})\big). \end{aligned}$$

Step 2: We show that there are a sufficiently large $C > 0$ and a sufficiently small $\epsilon_\delta > 0$ such that, for all $\epsilon \le \epsilon_\delta$,

$$\sum_{t \le \theta} \big(V^\epsilon(Z_t, \xi_t) - V^\epsilon(Z_{t-}, \xi_{t-})\big) \ge 0.$$

Expanding a typical summand, where $z \geq \delta$ and $\hat{\xi} \in \partial \mathrm{NT}(z)$, we find that

$$
\begin{aligned}
V^\epsilon&(z-\epsilon^4, 0) - V^\epsilon(z, \hat{\xi}) \\
&= v(z-\epsilon^4) - v(z) - \epsilon^2\big(u(z-\epsilon^4) - u(z)\big) \\
&\quad - \epsilon^4(1+C\epsilon^2)\big(w(z-\epsilon^4, 0) - w(z, \hat{\xi})\big) \\
&= -\epsilon^4 v_z(z) - \epsilon^8 v_{zz}(\tilde{z_1}) + \epsilon^6 u'(z) + \epsilon^{10} u''(\tilde{z_2}) + \epsilon^4 v_z(z) + C\epsilon^6 v_z(z) \\
&\geq C\epsilon^6 v_z(z) - \epsilon^8 v_{zz}(\tilde{z_1}) + \epsilon^6 u'(z) + \epsilon^{10} u''(\tilde{z_2}) \\
&\geq C\epsilon^6 v_z(z) - K\epsilon^6\big(z^{-1/2-\gamma} + \epsilon^2 \tilde{z_1}^{-1-\gamma} - \epsilon^4 \tilde{z_2}^{-3/2-\gamma}\big) \\
&> 0
\end{aligned}
$$

can be achieved for sufficiently small $\epsilon > 0$, uniformly in $z \geq \delta$, provided that C is chosen large enough. (Here, the points $\tilde{z_1}, \tilde{z_2} \in [z-\epsilon^4, z]$ come from the Taylor remainders of v and u, respectively.)

Step 3: Next, establish that for a suitable $k^* > 0$, we have

$$
\beta V^\epsilon(z, \xi) - \mathscr{L} V^\epsilon(z, \xi) \leq U(c_m z)(1+\epsilon^3 k^*)
$$

for all $\epsilon < \epsilon_\delta$ and for all $z \geq \delta$. Expanding the elliptic operator applied to V^ϵ, we obtain

$$
\begin{aligned}
\beta V^\epsilon&(z, \xi) - \mathscr{L} V^\epsilon(z, \xi) \\
&\leq -\frac{1}{2}\epsilon^2 |\sigma^\top \xi|^2 v_{zz}(z) + U(c_m z) - \epsilon^2 \mathcal{A} u(z) + \frac{1}{2}\epsilon^2 |\mathcal{R}_u(z, \xi)| \\
&\quad + (1+C\epsilon^2)\Big(\epsilon^2 \frac{1}{2} \operatorname{Tr}[\alpha(z)\alpha(z)^\top w_{\xi\xi}(z, \xi)] + \epsilon^3 |\mathcal{R}_w(z, \xi)|\Big) \\
&\leq (1+\epsilon^3 k^*) U(c_m z)
\end{aligned}
$$

for sufficiently large $|k^*|$, where k^* is positive for $\gamma < 1$ and negative for $\gamma > 1$, thanks to the pointwise estimates on the remainder terms (see Remark A.2) and the fact that $U(c_m z)$ is proportional to $z^{1-\gamma}$. The argument for logarithmic utility is similar. The inequality therefore holds for all sufficiently small ϵ and for all $z \geq \delta$.

Step 4: We now choose an appropriate control to use after time θ. Define the set

$$
\Xi_\epsilon = \{(x', y') \in \mathrm{K}_\epsilon : \ \delta - 2\epsilon^4 \leq x' + y' \cdot \mathbf{1}_d \leq \delta + \epsilon^4\}.
$$

We also define for each $(x', y') \in \Xi_\epsilon$ the neighborhoods

$$
R(x', y') = \{(\tilde{x}, \tilde{y}) \in \Xi_\epsilon : \ \tilde{x} > x', \tilde{y} > y' \text{ with } V^\epsilon(\tilde{x}, \tilde{y}) < V^\epsilon(x', y') + \epsilon^3\}.
$$

Since V^ϵ is smooth, each $R(x', y')$ is open. By construction we have

$$
\bigcup_{(x', y') \in \Xi_\epsilon} R(x', y') \supset \mathrm{NT}^\epsilon \cap \{(\tilde{x}, \tilde{y}) \in \mathrm{K}_\epsilon : \ \delta - \epsilon^4 \leq \tilde{x} + \tilde{y} \cdot \mathbf{1}_d \leq \delta\},
$$

and by compactness, there exists a finite subcover, say $(R(\zeta_n))_{n=1}^N$, for some points $\zeta_1, \dots, \zeta_N \in \Xi_\epsilon$.

Now define a mapping $\mathcal{I} : \Xi_\epsilon \to \{1, \dots, N\}$ that assigns to each point one of the neighborhoods in the subcover to which it belongs by

$$\mathcal{I}(x', y') := \min\{n : (x', y') \in R(\zeta_n)\}$$

and set

$$\zeta(x', y') := \zeta_{\mathcal{I}(x', y')}.$$

At each ζ_n, we select a control $\nu^n \in \Theta_\epsilon(\zeta_n)$ such that

$$v^\epsilon(\zeta_n) \le \mathbb{E}\left[\int_0^\infty e^{-\beta t} U(c_t^{\nu^n})\,dt\right] + \epsilon^3.$$

Note that $\nu^n \in \Theta_\epsilon(x', y')$ for all $(x', y') \in R(\zeta_n)$. Finally, define $\nu^* \in \Theta_\epsilon(x, y)$ by

$$\nu^*(\omega \overset{\theta}{\oplus} \omega', t) := \begin{cases} \nu^\epsilon(\omega, t) & \text{if } t \in [0, \theta(\omega)], \\ \nu^{\mathcal{N}(\omega)}(\omega', t - \theta(\omega)) & \text{if } t > \theta(\omega), \end{cases}$$

with

$$\mathcal{N}(\omega) = \mathcal{I}\Big(X_{\theta(\omega)}^{\nu^\epsilon, x}, Y_{\theta(\omega)}^{\nu^\epsilon, y}\Big).$$

Step 5: Piecing together the above estimates and proceeding as in the proof of Lemma 4.6 to get rid of the local martingale term, we obtain

$$\begin{aligned}
&V^\epsilon(x, y) \\
&\quad \le \mathbb{E}\left[e^{-\beta\theta} V^\epsilon(X_\theta^{\nu^\epsilon, x}, Y_\theta^{\nu^\epsilon, y}) + \int_0^\theta e^{-\beta s}(\beta V^\epsilon - \mathscr{L} V^\epsilon)(X_s^{\nu^\epsilon, x}, Y_s^{\nu^\epsilon, y})\,ds\right] \\
&\quad \le \mathbb{E}\left[e^{-\beta\theta} V^\epsilon\big(\zeta(X_\theta^{\nu^\epsilon, x}, Y_\theta^{\nu^\epsilon, y})\big) + \int_0^\theta e^{-\beta s}(1 + \epsilon^3 k^*) U(c_m Z_s^{\nu^\epsilon, z})\,ds\right] \\
&\qquad + \epsilon^3 \\
&\quad \le \mathbb{E}\left[e^{-\beta\theta} \int_0^\infty e^{-\beta s} U(c_s^{\nu^{\mathcal{N}}})\,ds + \int_0^\theta e^{-\beta s}(1 + \epsilon^3 k^*) U(c_m Z_s^{\nu^\epsilon, z})\,ds\right] \\
&\qquad + M_\epsilon + 2\epsilon^3 \\
&\quad \le J^{\epsilon, \delta}(x, y) + M_\epsilon + \epsilon^3 k^* v(z) + 2\epsilon^3,
\end{aligned}$$

where in the last step we have used that $k^* U$ is positive for $\gamma \neq 1$,[11] and where

$$M_\epsilon := \sup_{\substack{\delta - \epsilon^4 \le x' + y' \le \delta \\ (x', y') \in \mathrm{NT}^\epsilon}} |V^\epsilon(x', y') - v^\epsilon(x', y')|.$$

[11] For logarithmic utility ($\gamma = 1$), this follows similarly by additionally exploiting the estimate (4.5).

The convergence results from Theorem 2.3 imply that $M_\epsilon/\epsilon^2 \to 0$ as $\epsilon \to 0$. Since

$$J^{\epsilon,\delta} \geq V^\epsilon(x,y) - M_\epsilon - \epsilon^3\big(2 + k^* v(z)\big) = v^\epsilon(x,y) - o(\epsilon^2),$$

the proposed policy is indeed optimal at the leading order ϵ^2. $\square$

Acknowledgements The authors thank Bruno Bouchard, Paolo Guasoni, Jan Kallsen, Ludovic Moreau, Mathieu Rosenbaum, and Peter Tankov for fruitful discussions and Martin Forde for pertinent remarks on an earlier version. Moreover, they are very grateful to two anonymous referees and the editor for numerous constructive comments.

The second author was partially supported by the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK), Project D1 (Mathematical Methods in Financial Risk Management) of the Swiss National Science Foundation (SNF), and the Swiss Finance Institute. The third author gratefully acknowledges partial support by the European Research Council under the grant 228053-FiRM, by the ETH Foundation, and the Swiss Finance Institute.

Appendix A: Pointwise estimates

Proposition A.1 *There exists $K = K(\beta,\mu,r,\sigma,\gamma) > 0$ such that for $k = 0,1,2$ and $j = 0,1,2$,*

$$|D_\xi^k \partial_z^j w(z,\xi)| \leq K z^{-j-3k/4-\gamma} \quad \textit{for all } (z,\xi) \leftrightarrow (x,y) \in \mathrm{NT}^\epsilon.$$

Proof This follows from tedious but straightforward computations since all the functions and domains involved are known explicitly (cf. [43, Sect. 4.2] for a similar calculation). $\square$

Remark A.2 Proposition A.1 yields the expansion

$$\epsilon^4\big(\beta w(z,\xi) - \mathscr{L} w(z,\xi)\big) = -\epsilon^2 \frac{1}{2}\mathrm{Tr}[\alpha(z)\alpha(z)^\top w_{\xi\xi}(z,\xi)] + \epsilon^3 \mathcal{R}_w(z,\xi),$$

where the remainder satisfies the bound

$$|\mathcal{R}_w(z,\xi)| \leq K \sum_{k=0}^{3} \epsilon^k z^{(1-k)/4-\gamma} \quad \text{for all } (z,\xi) \leftrightarrow (x,y) \in \mathrm{NT}^\epsilon.$$

In particular, over the same region, we have

$$\epsilon^4|\beta w(z,\xi) - \mathscr{L} w(z,\xi)| \leq K \sum_{k=0}^{4} \epsilon^{k+2} z^{(2-k)/4-\gamma}.$$

We can also expand

$$\epsilon^2\big(\beta u(z) - \mathscr{L} u(z)\big) = \epsilon^2\big(a(z) - c_m z u'(z)\big) + \epsilon^2 \mathcal{R}_u(z,\xi),$$

with a bound on the remainder of

$$|\mathcal{R}_u(z,\xi)| \leq K(\epsilon z^{1/4-\gamma} + \epsilon^2 z^{-\gamma}) \quad \text{for all } (z,\xi) \leftrightarrow (x,y) \in \mathrm{NT}^\epsilon.$$

Proposition A.3 *Let $\epsilon > 0$ be given and consider $S := [z_0 - r_0, z_0 + r_0] \times \mathbb{R}^d \subset \mathrm{K}_\epsilon$ for some $z_0 > r_0 > 0$. Then, given any $\Psi \in C^2(S)$ for which $D_\xi \Psi$ has a compact support, there exists $K > 0$, independent of ϵ, such that*

$$\|\Psi\|_{C^2(S)} \leq K$$

and

$$\epsilon^2 |\beta \Psi(z,\xi) - \mathscr{L}\Psi(z,\xi)| \leq K(1 + \epsilon|\xi| + \epsilon^2|\xi|^2) \quad \forall (z,\xi) \in S.$$

Proof This again follows from a tedious but straightforward calculation. □

Appendix B: Proof of Theorem 2.1

In this section, we prove that for each fixed λ, the value function v^λ is a viscosity solution of the corresponding dynamic programming equation (3.7) on the domain

$$\mathcal{O}_\lambda = \{(x, y) \in \mathrm{K}_\lambda \ : \ x + y \cdot \mathbf{1}_d > 2\lambda\}.$$

As observed by Bouchard and Touzi [8], a "weak version" of the dynamic programming principle is sufficient to derive the viscosity property. For the convenience of the reader, we present a direct proof of the weak dynamic programming principle in our specific setting using the techniques of [8]. Then we use it to prove that v^λ is indeed a viscosity solution of (3.7).

B.1 Weak dynamic programming principle for v^λ

Fixing $(x, y) \in \mathcal{O}_\lambda$ and $\delta > 0$, let $B_\delta(x, y) \subset \mathbb{R}^{d+1}$ denote the ball of radius δ centered at (x, y) and set

$$K(x, y; \delta)_\lambda := \{(x', y') \ : \ x + y \cdot \mathbf{1}_d - \delta - \lambda \leq x' + y' \cdot \mathbf{1}_d \leq x + y \cdot \mathbf{1}_d + \delta\}.$$

Take $\delta > 0$ sufficiently small so that $K(x, y; 2\delta)_\lambda \subset \mathcal{O}_\lambda$. For any investment–consumption policy ν and initial endowment $(x', y') \in B_{\delta/2}(x, y)$, define $\theta := \theta^\nu$ as the exit time of the state process $(X, Y)^{\nu, x', y'}$ from $B_{\delta/2}(x, y)$. Following the standard convention, our notation does not explicitly show the dependence of θ on ν. It is then clear that

$$(X_{\theta^-}, Y_{\theta^-}) \in \overline{B_{\delta/2}(x, y)} \quad \text{and} \quad (X_\theta, Y_\theta) \in K(x, y, \delta)_\lambda.$$

The following weak version of the DPP is introduced in [8]. Let φ be a smooth and bounded function on $K(x, y, 2\delta)_\lambda$ satisfying

$$v^\lambda \leq \varphi \quad \text{on } K(x, y, 2\delta)_\lambda.$$

Then we have

$$v^\lambda(x, y) \leq \sup_{\nu \in \Theta_\lambda(x,y)} \mathbb{E}\left[\int_0^\theta e^{-\beta t} U(c_t)\, dt + e^{-\beta\theta} \varphi\big(X_\theta, Y_\theta\big)\right]. \tag{B.1}$$

(The restriction to bounded test functions φ is possible since by (4.12) v^λ is bounded on $K(x,y;2\delta)_\lambda$.) Conversely, let φ be a smooth function bounded on $K(x,y,2\delta)_\lambda$, satisfying

$$v^\lambda \geq \varphi \quad \text{on } K(x,y,2\delta)_\lambda.$$

Then we have

$$v^\lambda(x,y) \geq \sup_{\nu\in\Theta_\lambda(x,y)} \mathbb{E}\left[\int_0^\theta e^{-\beta t} U(c_t)\,dt + e^{-\beta\theta}\varphi\big(X_\theta, Y_\theta\big)\right]. \tag{B.2}$$

For proving (B.1) and (B.2), without loss of generality, let $\Omega = C_0([0,\infty), \mathbb{R}^d)$ be the space of continuous functions starting at zero, equipped with the Wiener measure $\mathbb{P}$, a standard Brownian motion W, and the completion $(\mathcal{F}_t)_{t\geq 0}$ of the filtration generated by W. Given a control $\nu \in \Theta_\lambda(x,y)$ and the exit time $\theta := \theta^\nu$ from above, fix $\omega \in \Omega$ and define

$$\nu^{\theta,\omega}(\omega',t) := \nu\big(\omega \overset{\theta}{\oplus} \omega', t+\theta(\omega)\big) \qquad \forall \omega' \in \Omega,\ t\geq 0,$$

where

$$(\omega \overset{\theta}{\oplus} \omega')_t := \begin{cases} \omega_t & \text{if } t\in[0,\theta(\omega)), \\ \omega'_{t-\theta(\omega)} + \omega_{\theta(\omega)} & \text{if } t\geq \theta(\omega). \end{cases}$$

We start with the proof of (B.1). By construction,

$$\nu^{\theta,\omega} \in \Theta_\lambda\Big(X^{\nu,x}_{\theta(\omega)}, Y^{\nu,y}_{\theta(\omega)}\Big);$$

in particular, $\nu^{\theta,\omega}$ is a well-defined impulse control. Moreover, note that

$$(X^{\nu,x}_{\theta(\omega)}, Y^{\nu,y}_{\theta(\omega)}) \in K(x,y,\delta)_\lambda$$

lies in the set $K(x,y,2\delta)_\lambda$ on which φ dominates v^λ by definition. Therefore,

$$\begin{aligned}
&\mathbb{E}\left[\int_0^\infty e^{-\beta t} U(c^\nu_t)\,dt \,\middle|\, \mathcal{F}_\theta\right](\omega) \\
&\quad = \int_0^{\theta(\omega)} e^{-\beta t} U\big(c^\nu_t(\omega)\big)\,dt + e^{-\beta\theta(\omega)} \int_\Omega \int_0^\infty e^{-\beta t} U\big(c^{\nu^{\theta,\omega}}_t(\omega')\big)\,dt\,d\mathbb{P}(\omega') \\
&\quad \leq \int_0^{\theta(\omega)} e^{-\beta t} U\big(c^\nu_t(\omega)\big)\,dt + e^{-\beta\theta(\omega)} v^\lambda\big(X^{\nu,x}_{\theta(\omega)}, Y^{\nu,y}_{\theta(\omega)}\big) \\
&\quad \leq \int_0^{\theta(\omega)} e^{-\beta t} U\big(c^\nu_t(\omega)\big)\,dt + e^{-\beta\theta(\omega)} \varphi\big(X^{\nu,x}_{\theta(\omega)}, Y^{\nu,y}_{\theta(\omega)}\big).
\end{aligned}$$

As a result, for any $\nu \in \Theta_\lambda(x,y)$,

$$\mathbb{E}\left[\int_0^\infty e^{-\beta t} U(c^\nu_t)\,dt\right] \leq \mathbb{E}\left[\int_0^\theta e^{-\beta t} U(c^\nu_t)\,dt + e^{-\beta\theta}\varphi(X^{\nu,x}_\theta, Y^{\nu,y}_\theta)\right].$$

By taking the supremum over all policies ν we arrive at (B.1).

To prove (B.2), set $\mathbb{V}$ to be the right-hand side of (B.2), that is,

$$\mathbb{V} := \sup_{\nu \in \Theta_\lambda(x,y)} \mathbb{E}\left[\int_0^\theta e^{-\beta t} U(c_t^\nu)\, dt + e^{-\beta\theta} \varphi(X_\theta^{\nu,x}, Y_\theta^{\nu,y})\right].$$

For any $\eta > 0$, we can choose $\nu^\eta \in \Theta_\lambda(x,y)$ satisfying

$$\mathbb{V} \le \eta + \mathbb{E}\left[\int_0^\theta e^{-\beta t} U(c_t^{\nu^\eta})\, dt + e^{-\beta\theta} \varphi\big(X_\theta^{\nu^\eta,x}, Y_\theta^{\nu^\eta,y}\big)\right]. \tag{B.3}$$

We have already argued that $(X_\theta^{\nu^\eta,x}, Y_\theta^{\nu^\eta,y}) \in K(x,y,\delta)_\lambda$. The next step is to construct a countable open cover of $K(x,y,\delta)_\lambda$. For every point $\zeta = (\tilde{x}, \tilde{y})$ in $K(x,y,2\delta)_\lambda$, set

$$\begin{aligned} R(\zeta) &:= R_\eta(\tilde{x}, \tilde{y}) \\ &= \{(x',y') \in K(x,y,2\delta)_\lambda \ :\ x' > \tilde{x},\ y' > \tilde{y},\ \varphi(x',y') < \varphi(\tilde{x},\tilde{y}) + \eta\}. \end{aligned}$$

By monotonicity of the value function,

$$v^\lambda(\zeta) \le v^\lambda(x',y') \quad \forall\ (x',y') \in R(\zeta).$$

Also, since φ is smooth, each $R(\zeta)$ is open, and

$$K(x,y,\delta)_\lambda \subset \bigcup_{\zeta \in K(x,y,2\delta)_\lambda} R(\zeta).$$

Hence, by the Lindelöf covering lemma [26, Theorem 15], we can extract a countable subcover

$$K(x,y,\delta)_\lambda \subset \bigcup_{n\in\mathbb{N}} R(\zeta_n).$$

Now define a mapping $\mathcal{I} : K(x,y,\delta)_\lambda \to \mathbb{N}$ that assigns to each point one of the neighborhoods in the subcover to which it belongs by

$$\mathcal{I}(x',y') := \min\{n : (x',y') \in R(\zeta_n)\}$$

and set

$$\zeta(x',y') := \zeta_{\mathcal{I}(x',y')}.$$

By definition,these constructions imply

$$\varphi(x',y') \le \varphi\big(\zeta(x',y')\big) + \eta \quad \forall\ (x',y') \in K(x,y,\delta)_\lambda. \tag{B.4}$$

As a final step, for each positive integer n, we choose a control $\nu^n \in \Theta_\lambda(\zeta_n)$ so that

$$v^\lambda(\zeta_n) \le \mathbb{E}\left[\int_0^\infty e^{-\beta t} U(c_t^{\nu^n})\, dt\right] + \eta. \tag{B.5}$$

By monotonicity, $\nu^n \in \Theta_\lambda(x', y')$ for every $(x', y') \in R(\zeta_n)$. We now define a composite strategy ν^* that follows the policy η satisfying (B.3) until the corresponding state process $(X, Y)^{\nu^\eta, x, y}$ leaves $B_{\delta/2}(x, y)$ at time $\theta = \theta^{\nu^\eta}$. It then switches to the policy ν^n corresponding to the index n that the state process is assigned to by the mapping $\mathcal{I}$, that is,

$$\nu^*(\omega \overset{\theta}{\oplus} \omega', t) := \begin{cases} \nu^\eta(\omega, t) & \text{if } t \in [0, \theta(\omega)], \\ \nu^{\mathcal{N}(\omega)}(\omega', t - \theta(\omega)) & \text{if } t > \theta(\omega), \end{cases}$$

with $\mathcal{N}(\omega) = \mathcal{I}(X^{\nu^\eta, x}_{\theta(\omega)}, Y^{\nu^\eta, y}_{\theta(\omega)})$. This construction ensures that we have $\nu^* \in \Theta_\lambda(x, y)$. Hence, it follows from the definitions of the value function and ν^*, inequalities (B.5), $v^\lambda \geq \varphi$ (which holds on $K(x, y, 2\delta)_\lambda$ by the definition of φ), (B.4), and (B.3) that

$$\begin{aligned} v^\lambda(x, y) &\geq \mathbb{E}\left[\int_0^\infty e^{-\beta t} U(c_t^{\nu^*})\, dt\right] \\ &= \mathbb{E}\left[\int_0^\theta e^{-\beta t} U(c_t^\eta)\, dt + e^{-\beta\theta} \int_0^\infty e^{-\beta t} U(c_t^{\mathcal{N}})\, dt\right] \\ &\geq \mathbb{E}\left[\int_0^\theta e^{-\beta t} U(c_t^\eta)\, dt + e^{-\beta\theta}\Big(\varphi\Big(\zeta\big(X_\theta^{\nu^\eta, x}, Y_\theta^{\nu^\eta, y}\big)\Big) - \eta\Big)\right] \\ &\geq \mathbb{E}\left[\int_0^\theta e^{-\beta t} U(c_t^\eta)\, dt + e^{-\beta\theta}\Big(\varphi\big(X_\theta^{\nu^\eta, x}, Y_\theta^{\nu^\eta, y}\big) - 2\eta\Big)\right] \\ &\geq \mathbb{V} - 3\eta. \end{aligned}$$

Since η was arbitrary, this establishes (B.2), thereby completing the proof.

B.2 v^λ Is a viscosity solution of (3.7)

We first state and prove some facts about the intervention operator $\mathbf{M}$ from (3.8), which are needed in the subsequent proofs. Similar observations appear in [37] for the case of proportional and fixed transaction costs. The modifications below are required to extend them to the case of pure fixed costs, for which the set of attainable portfolios at a fixed wealth level is no longer compact.

Throughout, $\underline{\psi}$ and $\overline{\psi}$ will denote the lower- and upper-semicontinuous envelopes of a locally bounded function ψ, respectively.

Lemma B.1 *Suppose that $\varphi : \mathrm{K}_\lambda \to \mathbb{R}$ satisfies $\sup_{z \in K} \|\varphi(z, \cdot)\|_\infty < \infty$ for every nonempty compact set $K \subset \mathbb{R}_+$.*

(i) *If φ is lower-semicontinuous, then $\mathbf{M}\varphi$ is lower-semicontinuous. In particular, if $\varphi \geq \mathbf{M}\varphi$, then $\underline{\varphi} \geq \mathbf{M}\underline{\varphi}$.*
(ii) *Let $\varphi \in C^1(\mathrm{K}_\lambda)$. If $(z, \xi) \mapsto \mathrm{D}_\xi \varphi(z, \xi)$ is compactly supported on $C \times \mathbb{R}^d$ for any compact set $C \subset R_+$, then $\mathbf{M}\varphi$ is upper-semicontinuous.*

Proof (i) Assume to the contrary that there exist $\zeta_0 = (z_0, \xi_0) \in \mathrm{K}_\lambda$, a constant $\eta > 0$, and a sequence $\mathrm{K}_\lambda \ni \zeta_n = (z_n, \xi_n) \to \zeta_0$ for which

$$\mathbf{M}\varphi(\zeta_0) > \liminf_{n\to\infty} \mathbf{M}\varphi(\zeta_n) + 2\eta.$$

Choose $\zeta_0^* = (z_0 - \lambda, \hat{\xi}_0)$ such that $\varphi(\zeta_0^*) + \eta/2 \geq \mathbf{M}\varphi(\zeta_0)$ and $\zeta_n^* = (z_n - \lambda, \hat{\xi}_n)$ with $\varphi(\zeta_n^*) + 1/n \geq \mathbf{M}\varphi(\zeta_n)$. Then, up to choosing a subsequence,

$$\varphi(\zeta_0^*) \geq \lim_{n\to\infty} \varphi(\zeta_n^*) + \eta.$$

By the lower-semicontinuity of φ, there exist an open neighborhood O of ξ_0^* and an integer $N > 0$ such that for all $n \geq N$ and all $\zeta \in O$,

$$\varphi(\zeta) \geq \varphi(\zeta_n^*) + \eta. \tag{B.6}$$

Observe that since $z_n \to z_0$ as $n \to \infty$, we have

$$A_n := \{z_n - \lambda\} \times \mathbb{R} \cap O \neq \emptyset. \tag{B.7}$$

Combining (B.7) and (B.6) yields

$$\varphi(\zeta_n^*) + \frac{1}{n} \geq \mathbf{M}\varphi(\zeta_n) \geq \sup_{\zeta' \in A_n} \varphi(\zeta') \geq \varphi(\zeta_n^*) + \eta,$$

which is a contradiction for n large enough. Finally, observe that if φ is lower-semicontinuous and $\varphi \geq \mathbf{M}\varphi$, then $\underline{\varphi} \geq \underline{\mathbf{M}\varphi} = \mathbf{M}\underline{\varphi}$ by the previous discussion.

(ii) This follows similarly as in the proof of [37, Lemma 3.2(i)] because the requirements we place on the gradient $\mathrm{D}_\xi\varphi$ ensure the existence of optimizers and accumulation points also in our setting. □

We are now ready to tackle the proof of Theorem 2.1, which we split into two lemmas.

Lemma B.2 *The value function v^λ is a viscosity supersolution of the dynamic programming equation* (3.7) *on* $\mathcal{O}_\lambda$.

Proof Let $(x_0, y_0) \in \mathcal{O}_\lambda$, and let φ be a smooth and bounded function on $K(x_0, y_0, \delta)_\lambda$ satisfying

$$0 = (\underline{v}^\lambda - \varphi)(x_0, y_0) = \min\{(\underline{v}^\lambda - \varphi)(x', y') : (x', y') \in K(x_0, y_0, \delta)_\lambda\}.$$

Using Lemma B.1 and the inequality $\underline{v}^\lambda \geq \varphi$ on $K(x_0, y_0, \delta)_\lambda$, we obtain

$$\varphi(x_0, y_0) = \underline{v}^\lambda(x_0, y_0) \geq \mathbf{M}\underline{v}^\lambda(x_0, y_0) \geq \mathbf{M}\varphi(x_0, y_0).$$

Therefore, it remains to show that

$$\big(\beta\varphi - \widetilde{U}(\varphi_x) - \mathscr{L}\varphi\big)(x_0, y_0) \geq 0.$$

Assume to the contrary that $(\beta\varphi - U(c^*) + c^*\varphi_x - \mathscr{L}\varphi)(x_0, y_0) < 0$ for some $c^* > 0$ and set $\phi(x, y) := \varphi(x, y) - \epsilon(|x - x_0|^4 + \|y - y_0\|^4)$. Then for $\epsilon > 0$ and $r > 0$ small enough, continuity yields

$$\big(\beta\phi - U(c^*) + c^*\phi_x - \mathscr{L}\phi\big)(x, y) < 0, \quad \forall (x, y) \in \overline{B}_r(x_0, y_0) \subset K(x_0, y_0, \delta)_\lambda.$$

Select a convergent sequence of points $(x_n, y_n, v^\lambda(x_n, y_n)) \to (x_0, y_0, \underline{v}^\lambda(x_0, y_0))$ and denote by $(X_t^n, Y_t^n) := (X_t^{x_n}, Y_t^{y_n})$ the portfolio process starting at (x_n, y_n) under the consumption-only strategy $c_t \equiv c^*$. Define

$$H^n := \inf\{t \geq 0 : \ (X_t^n, Y_t^n) \notin \overline{B}_r(x_0, y_0)\}$$

and note that $\liminf_{n\to\infty} \mathbb{E}[H^n] > 0$. Hence, there exists $\delta > 0$ with $\mathbb{E}[e^{-\beta H^n}] > \delta$ for all n sufficiently large. Itô's formula gives

$$\begin{aligned} \phi(x_n, y_n) &= \mathbb{E}\bigg[e^{-\beta H^n}\phi(X_{H^n}^n, Y_{H^n}^n) + \int_0^{H^n} e^{-\beta s}(\beta\phi + c^*\phi_x - \mathscr{L}\phi)(X_s^n, Y_s^n)\, ds\bigg] \\ &\leq \mathbb{E}\bigg[e^{-\beta H^n}\phi(X_{H^n}^n, Y_{H^n}^n) + \int_0^{H^n} e^{-\beta s} U(c^*)\, ds\bigg]. \end{aligned}$$

By construction of ϕ there exists $\eta > 0$ with $\varphi \geq \phi + \eta$ on $K(x_0, y_0, \delta)\backslash\overline{B}_r(x_0, y_0)$. Hence,

$$\phi(x_n, y_n) \leq \mathbb{E}\bigg[e^{-\beta H^n}\varphi(X_{H^n}^n, Y_{H^n}^n) + \int_0^{H^n} e^{-\beta s} U(c^*)\, ds\bigg] - \delta\eta.$$

Taking into account $(v^\lambda - \phi)(x_n, y_n) \to 0$, we note that for n large enough,

$$v^\lambda(x_n, y_n) \leq \mathbb{E}\bigg[e^{-\beta H^n}\varphi(X_{H^n}^n, Y_{H^n}^n) + \int_0^{H^n} e^{-\beta s} U(c^*)\, ds\bigg] - \frac{\delta\eta}{2}.$$

This contradicts the weak dynamic programming principle (B.2) for v^λ, thereby completing the proof. □

The image of an arbitrary smooth function under $\mathbf{M}$ is upper-semicontinuous only under additional assumptions (cf. Lemma B.1(ii)). As is customary in the theory of viscosity solutions (see, e.g., Sect. 9 of [10]), the viscosity subsolution property in the following lemma is therefore formulated in terms of the lower-semicontinuous envelope of the DPE.

Lemma B.3 *The value function v^λ is a viscosity subsolution of*

$$\min\big(\beta v^\lambda - \widetilde{U}(v_x^\lambda) - \mathscr{L}v^\lambda, \underline{v^\lambda - \mathbf{M}v^\lambda}\big) = 0 \quad \text{on } \mathcal{O}_\lambda.$$

Proof Step 1. Throughout this proof, $C > 0$ denotes a generic constant that may vary from line to line. We argue by contradiction. Let $(x_0, y_0) \in \mathcal{O}_\lambda$, and let φ be a smooth and bounded function on $K(x_0, y_0, \delta)_\lambda$ satisfying

$$0 = (\overline{v}^\lambda - \varphi)(x_0, y_0) = \max\{(\overline{v}^\lambda - \varphi)(x', y') : \ (x', y') \in K(x_0, y_0, \delta)_\lambda\}.$$

Suppose that for some $\eta > 0$, we have

$$\min\big(\beta\varphi - \mathscr{L}\varphi - \tilde{U}(\varphi_x), \underline{\varphi - \mathbf{M}\varphi}\big)(x_0, y_0) > \eta.$$

By continuity, there is a small rectangular neighborhood

$$N = N(x_0, y_0, \rho) := \Big\{(x, y) \in \mathbb{R} \times \mathbb{R}^d : \max_{i=1,\dots,d} \big(|x - x_0|, |y^i - y_0^i|\big) < \rho\Big\}$$

such that

$$\min\big(\beta\varphi - \mathscr{L}\varphi + c\varphi_x - U(c), \varphi - \mathbf{M}\varphi\big)(x, y) > \eta \tag{B.8}$$

for all $c > 0$ and $(x, y) \in N$.

Step 2. Choose a sequence $N \ni (x_n, y_n) \to (x_0, y_0)$ for which $v^\lambda(x_n, y_n)$ converges to $\overline{v}^\lambda(x_0, y_0)$. At each of these points, choose a $\frac{1}{n}$-optimal control ν^n in $\Theta_\lambda(x_n, y_n)$. We denote by (c_t^n) and τ^n the consumption process and first impulse time of ν^n, respectively, and write $(X_t^n, Y_t^n) := (X_t^{\nu^n, x_n}, Y_t^{\nu^n, y_n})$ for the corresponding controlled process. Define the stopping times

$$H^n := \inf\{t \geq 0 : (X_t^n, Y_t^n) \notin N\} \wedge 1 + \infty 1_{\{\theta^n = \tau^n\}}$$

and

$$\theta^n := H^n \wedge \tau^n.$$

We can further decompose $H^n = \underline{H}^n \wedge \overline{H}^n \wedge 1$, where

$$\underline{H}^n := \inf\{t \geq 0 :\ (X_t^n, Y_t^n) \in \partial N \cap \{x_0 - \rho\} \times \mathbb{R}^d\} + \infty 1_{\{\theta^n = \tau^n\}}$$

and

$$\overline{H}^n := \inf\{t \geq 0 :\ (X_t^n, Y_t^n) \in \partial N \cap \{x : x > x_0 - \rho\} \times \mathbb{R}^d\} + \infty 1_{\{\theta^n = \tau^n\}}.$$

Then there exists $\delta > 0$ such that $\mathbb{E}[\overline{H}^n] > \delta$ for all n sufficiently large.

Step 3. Write

$$h(c, x, y) := I(c, x, y) - \sup_{\hat{c} > 0} I(\hat{c}, x, y),$$

where

$$I(c, x, y) := -\beta\varphi(x, y) + \mathscr{L}\varphi(x, y) - c\varphi_x(x, y) + U(c).$$

Note that $I(c, x, y) < 0$ for all $c \in \mathbb{R}_+$ and $(x, y) \in N$ by (B.8). If we now set $c^*(x, y) = (U')^{-1}(\varphi_x(x, y))$, then it follows that

$$h(c, x, y) = I(c, x, y) - I\big(c^*(x, y), x, y\big) \leq 0.$$

By smoothness of φ and c^* and compactness of N, there exists a constant $L_\rho > 0$ with $|I(c^*(x, y), x, y)| \leq L_\rho$ for all $(x, y) \in N$. On the other hand, there is $\alpha > 0$ such that $I(c, x, y) \leq -\alpha c$ for all $c > 0$. This leads to the upper bound

$$h(c, x, y) \leq (-\alpha c + L_\rho) \wedge 0 \qquad \text{for all } c > 0, (x, y) \in N. \tag{B.9}$$

Since we only consider times t up to θ^n, we can assume without loss of generality that $c_t^n = c^*(X_t^n, Y_t^n)$ for $t \in (\theta^n, H^n]$. Together with (B.9), we obtain

$$\begin{aligned}\mathbb{E}\left[\int_0^{\theta^n} -e^{-\beta t} h(c_t, X_t, Y_t)\, dt\right] &= \mathbb{E}\left[\int_0^{H^n} -e^{-\beta t} h(c_t, X_t, Y_t)\, dt\right] \qquad \text{(B.10)}\\ &\geq C\alpha \mathbb{E}\left[\int_0^{H^n} e^{-rt} c_t\, dt\right] - L_\rho \mathbb{E}[H^n]. \\ &\geq C\alpha \mathbb{E}\left[\int_0^{\underline{H}^n \wedge 1} e^{-rt} c_t 1_{\{\theta^n = \underline{H}^n\}}\, dt\right] - L_\rho \mathbb{E}[H^n],\end{aligned}$$

where the first inequality uses (B.9) to change the discount factor.

Step 4. Set $\zeta_t^n := (X_t^n, Y_t^n)$. Weak dynamic programming in (B.1) implies

$$\begin{aligned}v^\lambda(x_n, y_n) &\leq \frac{1}{n} + \mathbb{E}\left[\int_0^{\theta^n} e^{-\beta t} U(c_t^n)\, dt + e^{-\beta\theta^n}\varphi(\zeta_t^n)\right] \\ &\leq \frac{1}{n} + \varphi(x_n, y_n) + \mathbb{E}\left[\int_0^{\theta^n} e^{-\beta t} I(c_t^n, \zeta_t^n)\, dt\right] \\ &\quad + \mathbb{E}\left[e^{-\beta\theta^n}\left(\varphi(\zeta_{\theta^n}^n) - \varphi(\zeta_{\theta^n -}^n)\right) 1_{\{\theta^n = \tau^n\}}\right] \\ &\leq \frac{1}{n} + \varphi(x_n, y_n) + \mathbb{E}\left[\int_0^{\theta^n} e^{-\beta t} I\left(c_t^*(\zeta_t^n), \zeta_t^n\right) dt\right] \\ &\quad + \mathbb{E}\left[\int_0^{\theta^n} e^{-\beta t} h(c_t^n, \zeta_t^n)\, dt\right] - C\eta \mathbb{P}[\theta^n = \tau^n] \\ &\leq \frac{1}{n} + \varphi(x_n, y_n) - C L_\rho \eta \mathbb{E}[\theta^n] - C\eta \mathbb{P}[\theta^n = \tau^n] \\ &\quad + \mathbb{E}\left[\int_0^{\theta^n} e^{-\beta t} h(c_t^n, \zeta_t^n)\, dt\right].\end{aligned}$$

Since $v^\lambda(x_n, y_n) - \varphi(x_n, y_n) - \frac{1}{n} \to 0$ as $n \to \infty$ and since the other terms on the right-hand side are negative, they must each vanish as n tends to infinity.

Step 5. We derive a contradiction using that

$$\lim_{n\to\infty} \max\left(\mathbb{E}[\theta^n], \mathbb{P}[\theta^n = \tau^n], \mathbb{E}\left[\int_0^{\theta^n} -e^{-\beta t} h(c_t^n, \zeta_t^n)\, dt\right]\right) = 0. \qquad \text{(B.11)}$$

Observe that since the first two terms vanish, $\mathbb{E}[H^n] \to 0$ and $\mathbb{P}[H^n = \theta^n] \to 1$. Since $\mathbb{E}[\overline{H}^n] > \delta$ for all n sufficiently large, we must therefore have $\mathbb{E}[\underline{H}^n] \to 0$ and $\mathbb{P}[\underline{H}^n = \theta^n] \to 1$. As a consequence,

$$\mathbb{E}\left[\int_0^{\underline{H}^n} e^{-rt} c_t^n 1_{\{\theta^n = \underline{H}^n\}}\, dt\right] \to \rho,$$

which follows from the simple observation that for any fixed n, the term inside the expectation represents the amount of discounted consumption needed for cash in the bank account to decrease from x_n to $x_0 - \rho$. However, by (B.11) and (B.10) we must have

$$0 = \lim_{n\to\infty} \mathbb{E}\left[\int_0^{\theta^n} -e^{-\beta t} h(c_t, X_t, Y_t)\, dt\right] \geq C\alpha\rho - L_\rho \lim_{n\to\infty} \mathbb{E}[H^n] = C\alpha\rho > 0,$$

which is a contradiction. □

References

1. Alcala, J.V., Fahim, A.: Balancing small fixed and proportional transaction cost in trading strategies (2013). Available online http://arxiv.org/abs/1304.7562.pdf
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. Balduzzi, P., Lynch, A.: Transaction costs and predictability: some utility cost calculations. J. Financ. Econom. **52**, 47–78 (1999)
4. Barles, G., Perthame, B.: Discontinuous solutions of deterministic optimal stopping time problems. RAIRO Modél. Math. Anal. Numér. **21**, 557–579 (1987)
5. Bertsimas, D., Kogan, L., Lo, A.W.: When is time continuous? J. Financ. Econom. **55**, 173–204 (2000)
6. Bichuch, M.: Asymptotic analysis for optimal investment in finite time with transaction costs. SIAM J. Financ. Math. **3**, 433–458 (2012)
7. Bichuch, M., Shreve, S.E.: Utility maximization trading two futures with transaction costs. SIAM J. Financ. Math. **4**, 26–85 (2013)
8. Bouchard, B., Touzi, N.: Weak dynamic programming principle for viscosity solutions. SIAM J. Control Optim. **49**, 948–962 (2011)
9. Constantinides, G.: Capital market equilibrium with transaction costs. J. Polit. Econ. **94**, 842–862 (1986)
10. Crandall, M.G., Ishii, H., Lions, P.-L.: User's guide to viscosity solutions of second order partial differential equations. Bull. Am. Math. Soc. **27**, 1–67 (1992)
11. Davis, M.H.A., Norman, A.R.: Portfolio selection with transaction costs. Math. Oper. Res. **15**, 676–713 (1990)
12. Dumas, B., Luciano, E.: An exact solution to a dynamic portfolio choice problem under transaction costs. J. Finance **46**, 577–595 (1991)
13. Eastham, J.F., Hastings, K.J.: Optimal impulse control of portfolios. Math. Oper. Res. **13**, 588–605 (1988)
14. Evans, L.C.: The perturbed test function method for viscosity solutions of nonlinear PDE. Proc. R. Soc. Edinb. A **111**, 359–375 (1989)
15. Fleming, W.H., Soner, H.M.: Controlled Markov Processes and Viscosity Solutions, 2nd edn. Springer, New York (2006)
16. Fukasawa, M.: Asymptotically efficient discrete hedging. In: Kohatsu-Higa, A., Privault, N., Sheu, S. (eds.) Stochastic Analysis with Financial Applications. Progress in Probability, vol. 65, pp. 331–346. Birkhäuser, Basel (2011)
17. Fukasawa, M.: Efficient discretization of stochastic integrals. Finance Stoch. **18**, 175–208 (2014)
18. Gerhold, S., Guasoni, P., Muhle-Karbe, J., Schachermayer, W.: Transaction costs, trading volume, and the liquidity premium. Finance Stoch. **18**, 1–37 (2014)
19. Goodman, J., Ostrov, D.N.: Balancing small transaction costs with loss of optimal allocation in dynamic stock trading strategies. SIAM J. Appl. Math. **70**, 1977–1998 (2010)
20. Guasoni, P., Muhle-Karbe, J.: Long horizons, high risk aversion, and endogeneous spreads. Math. Finance (2013). Available online http://onlinelibrary.wiley.com/doi/10.1111/mafi.12046/abstract
21. Hayashi, T., Mykland, P.A.: Evaluating hedging errors: an asymptotic approach. Math. Finance **15**, 309–343 (2005)
22. Janeček, K., Shreve, S.E.: Asymptotic analysis for optimal investment and consumption with transaction costs. Finance Stoch. **8**, 181–206 (2004)
23. Kabanov, Y.M., Safarian, M.M.: Markets with Transaction Costs. Springer, Berlin (2009)

24. Kallsen, J., Muhle-Karbe, J.: The general structure of optimal investment and consumption with small transaction costs (2013). Available online http://arxiv.org/abs/1303.3148
25. Kallsen, J., Muhle-Karbe, J.: Option pricing and hedging with small transaction costs. Math. Finance (2013). Available online http://onlinelibrary.wiley.com/doi/10.1111/mafi.12046/abstract
26. Kelley, J.: General Topology. Van Nostrand-Reinhold, New York (1955)
27. Korn, R.: Portfolio optimisation with strictly positive transaction costs and impulse control. Finance Stoch. **2**, 85–114 (1998)
28. Liu, H.: Optimal consumption and investment with transaction costs and multiple risky assets. J. Finance **59**, 289–338 (2004)
29. Lo, A.W., Mamaysky, H., Wang, J.: Asset prices and trading volume under fixed transaction costs. J. Polit. Econ. **112**, 1054–1090 (2004)
30. Lynch, A., Balduzzi, P.: Predictability and transaction costs: the impact on rebalancing rules and behavior. J. Finance **55**, 2285–2310 (2000)
31. Lynch, A., Tan, S.: Explaining the magnitude of liquidity premia: the roles of return predictability, wealth shocks, and state dependent transaction costs. J. Finance **66**, 1329–1368 (2011)
32. Magill, M., Constantinides, G.: Portfolio selection with transaction costs. J. Econ. Theory **13**, 245–263 (1976)
33. Martin, R.: Optimal multifactor trading under proportional transaction costs (2012). Available online http://arxiv.org/abs/1204.6488
34. Morton, A.R., Pliska, S.R.: Optimal portfolio management with fixed transaction costs. Math. Finance **5**, 337–356 (1995)
35. Muthuraman, K., Kumar, S.: Multidimensional portfolio optimization with proportional transaction costs. Math. Finance **16**, 301–335 (2006)
36. Nutz, M.: Risk aversion asymptotics for power utility maximization. Probab. Theory Relat. Fields **152**, 703–749 (2012)
37. Øksendal, B., Sulem, A.: Optimal consumption and portfolio with both fixed and proportional transaction costs. SIAM J. Control Optim. **40**, 1765–1790 (2002)
38. Possamaï, D., Soner, H.M., Touzi, N.: Homogenization and asymptotics for small transaction costs: the multidimensional case (2012). Available online http://arxiv.org/abs/1212.6275
39. Rogers, L.C.G.: Why is the effect of proportional transaction costs $O(\delta^{2/3})$. In: Yin, G., Zhang, Q. (eds.) Mathematics of Finance, pp. 303–308. Am. Math. Soc., Providence (2004)
40. Rosenbaum, M., Tankov, P.: Asymptotically optimal discretization of hedging strategies with jumps. Ann. Appl. Probab. **23**, 1002–1048 (2014)
41. Schroder, M.: Optimal portfolio selection with fixed transaction costs: numerical solutions (1995). Available online https://www.msu.edu/~schrode7/numerical.pdf
42. Shreve, S.E., Soner, H.M.: Optimal investment and consumption with transaction costs. Ann. Appl. Probab. **4**, 609–692 (1994)
43. Soner, H.M., Touzi, N.: Homogenization and asymptotics for small transaction costs. SIAM J. Control Optim. **51**, 2893–2921 (2013)
44. Whalley, A.E., Wilmott, P.: An asymptotic analysis of an optimal hedging model for option pricing with transaction costs. Math. Finance **7**, 307–324 (1997)
45. Zhang, R.: Couverture Approchée des Options Européennes. Ph.D. thesis, Ecole Nationale des Ponts et Chaussées (1999). Available online https://tel.archives-ouvertes.fr/tel-00005623/document