Rebalancing with Linear and Quadratic Costs[*]

Ren Liu[†]   Johannes Muhle-Karbe[‡]   Marko H. Weber[§]

September 5, 2017

Abstract

We consider a financial market consisting of one safe and one risky asset, which offer constant investment opportunities. Taking into account both proportional transaction costs and linear price impact, we derive optimal rebalancing policies for representative investors with constant relative risk aversion and a long horizon.

1   Introduction

Proportional transaction costs are ubiquitous even in the most liquid financial markets in the form of bid-spreads. For large institutional investors, the price impact of their trades also is a key concern.[1] As a result, both frictions have generated voluminous literatures that analyze how to balance the gains and costs of portfolio rebalancing optimally. Bid-ask spreads lead to trading costs *linear* in the amounts transacted. Then, it is optimal to refrain from trading while the uncontrolled portfolio lies inside some "no-trade region" around the frictionless optimum; once its boundaries are breached, one performs the minimal amount of rebalancing necessary to remain inside.[2] In contrast, linear price impact leads to *quadratic* trading costs, which are less severe for small trades but become prohibitively expensive for larger orders. As a result, optimal policies typically prescribe rebalancing at all times but at a finite absolutely continuous rate, in contrast to the singular controls used with proportional transaction costs.[3]

[†]ETH Zürich, Departement Mathematik, Rämistrasse 101, CH-8092, Zürich, Switzerland, email: `ren.liu@math.ethz.ch`.

[‡]Carnegie Mellon University, Department of Mathematical Sciences, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA, email: `johannesmk@cmu.edu`.

[§]Columbia University, Department of Industrial Engineering and Operations Research, 535G S. W. Mudd Building, New York, NY 10027, USA email: `mhw2146@columbia.edu`.

[1]For small private investors, fixed commissions, levied on each trade regardless of its size, are also crucial (cf., e.g., [25] and the references therein). Yet, for large portfolios, their influence becomes negligible [3], and we disregard them in the present study.

[2]See [27, 8, 10, 11, 34, 9] for models with constant investment opportunities, [34, 20, 5, 14] for corresponding explicit asymptotic formulas, and [28, 36, 32, 22, 21] for recent extensions to more general settings.

[3]Price impact has been studied extensively in the optimal execution literature, see, e.g., [4, 1, 33]. Recently, increasing attention has also been devoted to its influence on dynamic portfolio choice [12, 13, 2, 19, 7, 30].

All of the extant literature studies either linear or quadratic trading costs. This paper fills this gap by analyzing the *joint* impact of proportional transaction costs and linear price impact on portfolio rebalancing.

To make the model tractable, we focus (as in Gerhold et al. [14] resp. Guasoni and Weber [19] for proportional resp. quadratic costs treated separately) on a single risky asset with constant investment opportunities, and a representative agent with constant relative risk aversion and a long horizon. The representative investor's wealth serves as a proxy for a multiple of market capitalization, and prize impact is assumed to be inversely proportional to latter. That is, a trade of a given size has a smaller impact as markets grow, compare the discussion in Guasoni and Weber [19]. The present study builds on the results of Guasoni and Weber [19] by adding a bid-ask spread. This price impact model leads to stable long-term behavior. The most popular alternative – constant price impact as in [4, 26, 1] – is well-suited to the short-horizons optimal execution problems these authors have in mind. However, it leads to degenerate results in the long run, as growing fund values eventually make rebalancing prohibitively expensive. To bridge the gap between the different models, we argue that, asymptotically for small costs, our results formally extend to rather general specifications of the price impact function.

The optimal policy in the presence of both frictions turns out to be of the following form. As with proportional transaction costs, there is a no-trade region, where it is optimal to simply hold the current portfolio. Once its boundaries are breached, price impact rules out singular controls; hence, one instead starts trading at some finite rate so as to steer the portfolio back to the no-trade region. Since this policy does not allow to keep the portfolio uniformly close to the frictionless target, trading starts earlier than in a model with only proportional costs, i.e., the width of the no-trade region is decreased by the additional price impact. On the other hand, since there are now two frictions contributing to the total rebalancing cost, the total trading rate is always lower than in a model that only takes into account price impact. We prove a rigorous verification theorem that identifies the trading boundaries, the trading rate, and the associated welfare through the solution of a nonlinear free-boundary value problem that can be solved numerically. To ease implementation, we also present asymptotics for small linear and quadratic costs, which reduce the computation of the optimal policy and welfare to finding the root of a scalar function.[4] With two competing frictions, it is particularly important to assess the quality of the asymptotic approximations, since it is not clear a priori whether the matched rescaling leads to accurate results. Our exact formulas allow to do this, and show that the small-cost approximations perform very well.

A formal extension of our asymptotic results suggests that, similarly as for proportional costs [22] and price impact [30] considered separately, the optimal policy is robust to the form of the trading cost. To wit, in Section 6, we consider general price impact functions depending on the large investor's wealth and an additional exogenous state variable (say, the wealth of other traders) in an arbitrary manner. Then, the same asymptotic trading rate as in our baseline model obtains, substituting the current value of the trading cost at each point in time.

The remainder of the paper is organized as follows. Section 2 describes the model. Our main results are presented in Section 3. Subsequently, we illustrate them with some numerical examples. Section 5 contains a heuristic derivation of our main results; our asymptotics are presented in Section 6. For better readability, all proofs are delegated to Section 7.

---

[4]This is similar in complexity to the asymptotics for small fixed and proportional costs studied by [25].

2 Model

Consider a market consisting of one safe asset normalized to one,[5] and one risky asset, whose mid price S_t follows geometric Brownian motion:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t. \tag{2.1}$$

Here, $(W_t)_{t\geq 0}$ is a standard one-dimensional Brownian motion, $\mu > 0$ is the expected excess return, and $\sigma > 0$ is the volatility. Trades are not settled at the idealized best quote S_t. Instead, sales only earn lower bid prices, whereas purchases are charged higher ask prices. Moreover, trading large positions quickly moves prices further in an adverse direction. To wit, the average execution price for trading $\Delta\theta$ shares over a time interval Δt is

$$S_t \left(1 + \varepsilon \text{sgn}(\Delta\theta) + \lambda \frac{S_t \Delta\theta}{X_t \Delta t} \right). \tag{2.2}$$

Here, the first term corresponds to a relative bid-ask spread ε, i.e., a higher ask price $(1+\varepsilon)S_t$ for purchases and a lower bid price $(1-\varepsilon)S_t$ for sales, respectively. The second term describes the additional (relative) price impact of large trades executed quickly. This price impact is proportional to the monetary trading rate $S_t\Delta\theta/\Delta t$, and inversely proportional to market capitalization, which is proxied by the representative investor's wealth X_t.[6] The constant of proportionality λ in turn quantifies the market's limited liquidity; put differently, $1/\lambda$ measures market "depth". For $\varepsilon, \lambda \to 0$, one recovers the classical frictionless case, where arbitrary amounts $\Delta\theta$ can be purchased or sold at the mid price for $S_t\Delta\theta$. Nontrivial bid-ask spreads ($\varepsilon > 0$) and finite market depth ($\lambda > 0$) lead to additional linear and quadratic trading costs, respectively. Specifically, with both frictions, the execution cost of trading $\Delta\theta$ shares over a time interval Δt is given by[7]

$$S_t \Delta\theta + \varepsilon S_t \left| \frac{\Delta\theta}{\Delta t} \right| \Delta t + \lambda \frac{S_t^2 \Delta\theta^2}{X_t \Delta t^2} \Delta t. \tag{2.3}$$

For tractability, we now pass to the continuous-time limit. Denote by θ_t the number of risky shares the investor holds at time t, and replace $\Delta\theta/\Delta t$ in (2.3) with $\dot{\theta}_t := \lim_{h\downarrow 0} \frac{\theta_{t+h}-\theta_t}{h}$. Then, the investor's cash position $C_t = X_t - S_t\theta_t$ evolves as

$$dC_t = -S_t d\theta_t - \varepsilon S_t |\dot{\theta}_t| dt - \lambda \frac{S_t^2 \dot{\theta}_t^2}{X_t} dt.$$

Write $u_t := \dot{\theta}_t S_t / X_t$ for the *wealth turnover* at time t. With this notation, a straightforward application of Itô's formula (compare [19, Lemma A.1]) shows that the corresponding wealth process $X_t := \theta_t S_t + C_t$ and risky weight $Y_t := \theta_t S_t / X_t$ have the following dynamics:

$$\frac{dX_t}{X_t} = Y_t(\mu dt + \sigma dW_t) - \varepsilon|u_t|dt - \lambda u_t^2 dt, \tag{2.4}$$

$$dY_t = (Y_t(1-Y_t)(\mu - Y_t\sigma^2) + u_t + \varepsilon|u_t|Y_t + \lambda Y_t u_t^2)dt + Y_t(1-Y_t)\sigma dW_t. \tag{2.5}$$

---

[5]That is, the interest rate r is set equal to zero.

[6]See [19] for more details on this price impact model and the related literature.

[7]Note the price impact is purely temporary in our model, in that no trade influences the subsequent ones. There is a large literature on optimal execution with persistent price impact, that only wears off gradually after the completion of each trade (cf. [31] as well as many more recent studies). Since we are working on a much longer time scale than in this literature, we abstract from this issue and instead suppose that the temporary and persistent impact costs generated by various "sub-trades" on a finer "execution time-grid" are all aggregated into our price impact cost. Indeed, suppose each infinitesimal sub-trade is executed in the setting of Obizhaeva and Wang [31]. Then, the expected execution costs are of the same linear-quadratic form as in our model.

As without proportional transaction costs [19], linear price impact implies that the risky weight is no longer a control variable that can be specified freely by the investor. Instead, it becomes a state variable, for which only the drift rate can be influenced by applying the control u. To make this precise, fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F})_{t\geq 0}, \mathbb{P})$ supporting a Brownian motion $(W_t)_{t\geq 0}$, where $\mathcal{F}_t$ is the augmentation of the filtration generated by W. We then define strategies in terms of the control variable $(u_t)_{t\geq 0}$. To rule out doubling strategies, we focus on *admissible* strategies with positive wealth process X^u:

Definition 2.1. *An* admissible *strategy is an adapted process $(u_t)_{t\geq 0}$, which is square-integrable (i.e., $\int_0^T u_t^2 dt < \infty$ a.s. for all $T > 0$) and such that (2.4) has a unique strong solution on $[0, \infty)$ for any $Y_0 \in [0, 1]$. For any such admissible strategy, the corresponding wealth process is*[8]

$$X_t^u = X_0 \exp\left(\int_0^T \left(\mu Y_t - \frac{\sigma^2}{2}Y_t^2 - \varepsilon|u_t| - \lambda u_t^2\right) dt + \int_0^T \sigma Y_t dW_t\right).$$

As in [11, 15, 16], the representative investor has constant relative risk aversion and maximizes the growth rate of her expected utility from terminal wealth over a long horizon. Put differently, she maximizes the "equivalent safe rate", for which a full safe investment yields the same utility as investing optimally in the original market:

Definition 2.2. *An admissible strategy $(u_t)_{t\geq 0}$ is called* long-run optimal, *if it maximizes the equivalent safe rate*

$$\mathrm{ESR}_\gamma(u) := \liminf_{T\to\infty} \frac{1}{T} \log \mathbb{E}\left[(X_T^u)^{1-\gamma}\right]^{\frac{1}{1-\gamma}} \tag{2.6}$$

over all admissible strategies, where $0 < \gamma \neq 1$ denotes the investor's relative risk aversion.

3   Main Results

Our main results can be summarized as follows:

Theorem 3.1. *An investor with constant relative risk aversion $0 < \gamma \neq 1$ trades to maximize the equivalent safe rate (2.6), in the presence of a nontrivial bid-ask spread ε and finite market depth $1/\lambda$. Then, if $y_* := \mu/\gamma\sigma^2 \in (0,1)$ and the trading costs ε, λ are sufficiently small:*

i) *There exist constants $\beta \in [\max\{0, \mu - \gamma\sigma^2/2\}, \mu^2/2\gamma\sigma^2]$ and $0 \leq y_- \leq y_+ \leq 1$, as well as a C^1-function $q : [0,1] \to \mathbb{R}$ which solve the ODE*

$$0 = -\beta + \mu y - \frac{\gamma\sigma^2}{2}y^2 + y(1-y)(\mu - \gamma\sigma^2 y)q + \frac{\sigma^2}{2}y^2(1-y)^2(q' + (1-\gamma)q^2)$$

$$+ \begin{cases} \frac{1}{4\lambda}\frac{(q-\varepsilon(1-yq))^2}{1-yq}, & \text{if } y \in [0, y_-], \\ 0, & \text{if } y \in [y_-, y_+], \\ \frac{1}{4\lambda}\frac{(q+\varepsilon(1-yq))^2}{1-yq}, & \text{if } y \in [y_+, 1], \end{cases} \tag{3.1}$$

---

[8]If $y_* := \mu/\gamma\sigma^2 \in (0, 1)$, i.e., the frictionless target portfolio prescribes neither leverage nor shortselling, then Lemma 7.5 shows that Y_t takes values in $[0, 1]$ almost surely for all t. In particular, $\int_0^T Y_t^2 dt < \infty$ so that the process X^u is well defined in this case.

with boundary conditions

$$q(0^+) = \varepsilon + 2\sqrt{\lambda\beta}, \tag{3.2}$$

$$q(1^-) = \frac{\lambda d - \varepsilon(1 - \varepsilon) - \sqrt{\lambda d(\lambda d - 2 + 2\varepsilon)}}{(1 - \varepsilon)^2}, \quad \text{where } d := -\gamma\sigma^2 - 2\beta + 2\mu, \tag{3.3}$$

$$q(y_-) = \frac{\varepsilon}{1 + \varepsilon y_-}, \tag{3.4}$$

$$q(y_+) = \frac{-\varepsilon}{1 - \varepsilon y_+}. \tag{3.5}$$

ii) *A long-run optimal strategy $\hat{u}$ is to remain inactive while the corresponding risky weight lies in the no-trade region $[y_-, y_+]$, and to rebalance at the following wealth turnover rate if it does not:*

$$\hat{u}(y) = \begin{cases} \frac{1}{2\lambda}\left(\frac{q(y)}{1 - yq(y)} - \varepsilon\right) \geq 0, & \text{if } y \in [0, y_-], \\ \frac{1}{2\lambda}\left(\frac{q(y)}{1 - yq(y)} + \varepsilon\right) \leq 0, & \text{if } y \in [y_+, 1]. \end{cases}$$

iii) *The maximal equivalent safe rate is β.*

The constant $y_* = \mu/\gamma\sigma^2$ is the risky weight without frictions [29]. Thus, $y_* \in (0, 1)$ means that the frictionless optimal strategy neither shorts nor levers the risky asset. As shown by Guasoni and Weber [19, Theorem 2.3], levered or short positions cannot be admissible with linear price impact for risk-averse investors, because they cannot be liquidated quickly enough to offset unfavorable diffusive price moves. This is only exacerbated by the additional linear trading cost. Hence, buy-and-hold strategies are optimal for $y_* \notin (0, 1)$, as in [19, Theorem 2.3]:

Proposition 3.2. *Under the assumptions of Theorem 3.1:*

(i) *If $\mu/\gamma\sigma^2 \leq 0$, then $Y_t = 0$ and $\hat{u}_t = 0$ for all t is long-run optimal, and $\mathrm{ESR}_\gamma(\hat{u}) = 0$.*

(ii) *If $\mu/\gamma\sigma^2 \geq 1$, then $Y_t = 1$ and $\hat{u}_t = 0$ for all t is long-run optimal, and $\mathrm{ESR}_\gamma(\hat{u}) = \mu - \gamma\sigma^2/2$.*

The objective function (2.6) uses paper wealth rather than the liquidation value of the portfolio. This seems more reasonable for long-run investments meant to run indefinitely, such as trust funds and university endowments. However, similarly as in [19, Lemma 2.4] for purely quadratic costs, it can be shown that this choice is of little consequence.[9] To wit, assuming a constant best quote (which is justified if the portfolio is sold quickly), a policy of selling at a constant turnover rate completes the liquidation within a short period of time and at a small fraction of portfolio value. Formally:

Lemma 3.3. *Let $S_t \equiv S$ be constant for $t \geq T$. Then, the liquidation time $L(u) = \inf\{t \geq 0 : \theta_{T+t} = 0\}$ of the constant selling policy $u_t \equiv u < 0$, equals*

$$L(u) = -\frac{\log\left(1 + (\varepsilon|u| + \lambda u^2)\frac{Y_T}{u}\right)}{\varepsilon|u| + \lambda u^2} \sim -\frac{Y_T}{u}.$$

The corresponding relative liquidation cost is

$$\frac{X_T - X_{T+L(u)}}{X_T} = \varepsilon Y_T - \lambda u Y_T.$$

[9]In the proof of [19, Lemma 2.4], use (2.4-2.5) instead of [19, Equation (2.4) and (2.5)]. The rest of the proof carries through unchanged.
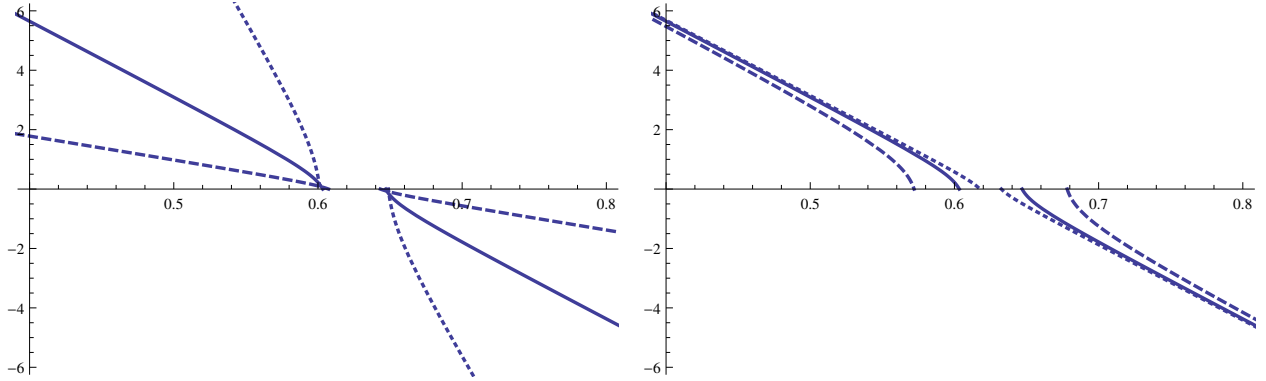
Figure 1: left panel: optimal wealth turnover $\hat{u}$ (Dotted: $\lambda = 0.001\%$, Solid: $\lambda = 0.01\%$, Dashed: $\lambda = 0.1\%$) against the risky weight Y with $\varepsilon = 0.1\%$ fixed. Right panel: wealth turnover $\hat{u}$ (Dotted: $\varepsilon = 0.01\%$, Solid $\varepsilon = 0.1\%$, Dashed: $\varepsilon = 1\%$) against the risky weight Y with $\lambda = 0.01\%$ fixed. Model parameters are $\mu = 8\%$, $\sigma = 16\%$, and $\gamma = 5$.

For instance, if[10] $u = -\lambda^{-1/2}$ the liquidation time is less than $\lambda^{1/2}$ years, since $Y_T \in [0, 1]$. Furthermore, the corresponding liquidation cost is less than $(\varepsilon + \lambda^{1/2})$ times the terminal wealth. Using the estimation interval of $[10^{-3}, 10^{-7}]$ for λ (cf. [19, Section 3.1]) and assuming a liquid stock with a bid-ask spread of 10 basis points yields a liquidation time between 0.08 and 7.91 days and a relative liquidation cost between 0.13% and 3.26%. As the horizon increases, the impact of these small costs on the equivalent safe rate vanishes. Moreover, the short liquidation time supports the constant best quote assumption.

4   Numerical Examples

In this section, we investigate the properties of the optimal rebalancing policy from Theorem 3.1 in some numerical examples. This also allows us to assess the quality of the asymptotics established in Section 6, which turns out to be excellent.

Figure 1 displays the optimal policy for various trading costs. The left panel shows how the turnover rate and the corresponding no-trade region depend on the price impact parameter λ. As the latter decreases, turnover quickly increases near the boundary of the no-trade region, converging to the singular controls ("pushing at an infinite rate") applied there with only proportional costs. For higher price impact costs, the optimal trading rate is almost linear in the deviation from the trading boundaries. Moreover, the width of the no-trade region decreases in this case, as investors start trading earlier to compensate for the slower trading rate at the boundary. However, the size of this effect is quite small, i.e., the width of the optimal no-trade region is relatively insensitive to the quadratic costs.

The right panel in Figure 1 plots the trading rate for different widths ε of the bid-ask spread. As the latter decreases, the no-trade region shrinks to zero and the optimal policy converges to the one with price impact only, i.e., rebalancing at a rate essentially proportional to the deviation from the frictionless Merton portfolio [19]. For larger spreads, the no-trade regions widens quickly, and

---

[10]As shown in [19, Equation (2.12)] the optimal turnover $u^{0,\lambda}$ without linear costs admits the following asymptotic expansion:

$$u^{0,\lambda}(y) = \sigma(\gamma/2)^{1/2}(y_* - y)\lambda^{-1/2} + o(\lambda^{-1/2}).$$
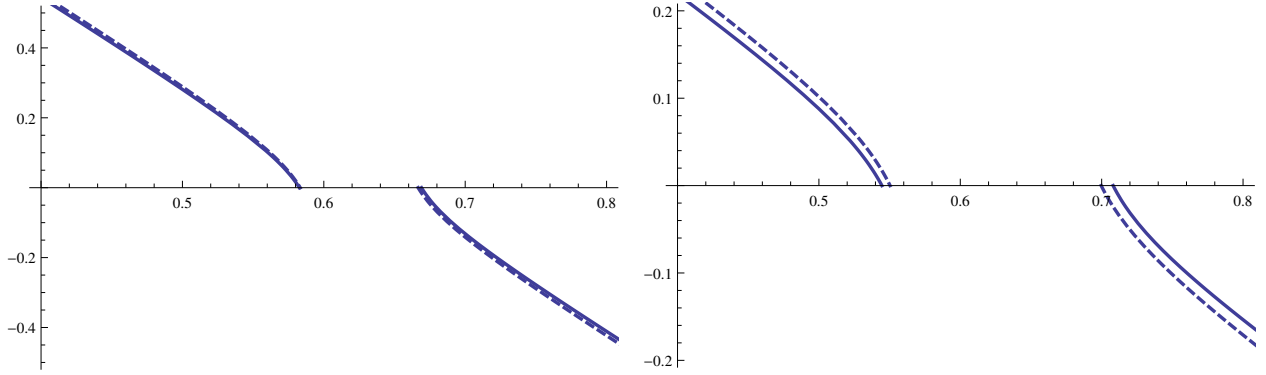
Figure 2: optimal wealth turnover $\hat{u}$ from Theorem 3.1 (solid) and its asymptotic expansion (6.15) (dashed), for $\varepsilon = \lambda = 1\%$ (left panel) and $\varepsilon = \lambda = 5\%$ (right panel). Model parameters are $\mu = 8\%$, $\sigma = 16\%$, and $\gamma = 5$.

the optimal rebalancing rate increases much faster near the trading boundaries than further away from these.[11]

In summary, the optimal policy prescribes to i) start trading earlier than with only proportional costs, and ii) rebalance slower than with only price impact. The turnover rate increases faster near the trading boundaries; further way from these, it approaches its counterpart for only quadratic costs.

As a complement, the quality of the small-cost asymptotics derived in Section 6 is assessed in Figure 2. There, we compare the optimal turnover rate to its asymptotic expansion (6.15) for two combinations of trading costs. Even for unrealistically large frictions, the approximations provide an excellent fit. Hence, the computational load can be eased to finding the root of a single scalar function with little loss in accuracy.

5 Heuristics

In this section, we use arguments from stochastic control to heuristically derive a candidate solution for the long-run problem (2.6). To this end, consider the maximization of expected power utility $U(x) = x^{1-\gamma}/(1-\gamma)$ from terminal wealth at time $T > 0$. Denote by $V(t, X_t, Y_t)$ the corresponding value function, which is assumed to depend on the current wealth X_t, the current risky weight Y_t, and time t. For any given strategy u, Itô's formula yields:

$$
\begin{aligned}
dV(t, X_t, Y_t) =& V_t dt + V_x dX_t + V_y dY_t + \frac{V_{xx}}{2} d\langle X\rangle_t + \frac{V_{yy}}{2} d\langle Y\rangle_t + V_{xy} d\langle X, Y\rangle_t \\
=& V_t dt + V_x(\mu X_t Y_t - \varepsilon X_t|u_t| - \lambda X_t u_t^2)dt + V_x X_t Y_t \sigma dW_t \\
& + V_y(Y_t(1-Y_t)(\mu - Y_t \sigma^2) + u_t + \varepsilon Y_t|u_t| + \lambda Y_t u_t^2)dt + V_y Y_t(1-Y_t)\sigma dW_t \\
& + \left(\frac{\sigma^2}{2} V_{xx} X_t^2 Y_t^2 + \frac{\sigma^2}{2} V_{yy} Y_t^2(1-Y_t)^2 + \sigma^2 V_{xy} X_t Y_t^2(1-Y_t)\right) dt.
\end{aligned}
$$

By the martingale optimality principle of stochastic control, the value function $V(t, X_t, Y_t)$ must be a supermartingale for any admissible strategy, and a martingale for the optimal one. That is,

---

[11]There, it grows according to the asymptotic formula (6.16) corresponding to a model with only quadratic costs.

the drift of $V(t, X_t, Y_t)$ cannot be positive and must become zero for the optimizer. This leads to the Hamilton-Jacobi-Bellman (henceforth HJB) equation:

$$0 = V_t + y(1-y)(\mu - \sigma^2 y)V_y + \mu xy V_x + \frac{\sigma^2 y^2}{2}(x^2 V_{xx} + (1-y)^2 V_{yy} + 2x(1-y)V_{xy})$$
$$+ \max_u \left(-\lambda xu^2 V_x - \varepsilon x|u|V_x + V_y(u + \varepsilon y|u| + \lambda yu^2)\right).$$

The homotheticity $U(x) = x^{1-\gamma}U(1)$ of the power utility function and the conjecture that – in the long run – utility should grow at a constant exponential rate motivate the following ansatz for the long-run value function:

$$V(t, x, y) = \frac{x^{1-\gamma}}{1-\gamma}e^{(1-\gamma)(\beta(T-t)+\int_p^y q(z)dz)}. \tag{5.1}$$

Note that the function q is defined up to some arbitrary p. This definition of V leads to the long-run version of the HJB-equation:

$$0 = -\beta + \mu y - \frac{\gamma \sigma^2}{2}y^2 + y(1-y)(\mu - \gamma\sigma^2 y)q + \frac{\sigma^2}{2}y^2(1-y)^2(q' + (1-\gamma)q^2)$$
$$+ \max_u(-\lambda u^2 - \varepsilon|u| + (u + \varepsilon|u|y + \lambda yu^2)q). \tag{5.2}$$

Decomposing wealth turnover u into purchase and sale turnover, i.e., $u = u^+ - u^-$, the HJB equation reduces to

$$0 = -\beta + \mu y - \frac{\gamma \sigma^2}{2}y^2 + y(1-y)(\mu - \gamma\sigma^2 y)q + \frac{\sigma^2}{2}y^2(1-y)^2(q' + (1-\gamma)q^2)$$
$$+ \max_{u^+ \geq 0}(-\lambda(u^+)^2 - \varepsilon u^+ + (u^+ + \varepsilon u^+ y + \lambda y(u^+)^2)q)$$
$$+ \max_{u^- \geq 0}(-\lambda(u^-)^2 - \varepsilon u^- + (-u^- + \varepsilon u^- y + \lambda y(u^-)^2)q). \tag{5.3}$$

Suppose the "second-order condition" $q(y)y < 1$ is satisfied (this holds for the function q constructed in Lemma 7.2). Then, the maxima are attained at

$$u^+(y) = \max\left(\frac{1}{2\lambda}\left(\frac{q(y)}{(1-yq(y))} - \varepsilon\right), 0\right), \tag{5.4}$$

$$u^-(y) = \max\left(-\frac{1}{2\lambda}\left(\frac{q(y)}{(1-yq(y))} + \varepsilon\right), 0\right). \tag{5.5}$$

The optimizer with proportional transaction costs but without price impact is characterized by a no-trade interval around the frictionless optimum $y_* = \mu/\gamma\sigma^2$. Hence, we conjecture that the no-trade region in the present setting,

$$\left\{y : u^+(y) = u^-(y) = 0\right\} = \left\{y : -\varepsilon < \frac{q(y)}{1 - yq(y)} < \varepsilon\right\},$$

is also given by some interval $[y_-, y_+]$. Substituting the optimal turnover rates (5.4-5.5) back into (5.3), the HJB equation in turn simplifies to the ODE (3.1). Imposing continuity across the boundaries y_-, y_+ of the no-trade region in turn yields (3.4-3.5). Since the differential equation (3.1) is of order one and there are four unknowns to be determined (β, y_-, y_+, and q), the value matching conditions (3.4-3.5) are not sufficient to characterize the solution. As a way out, we add two additional boundary conditions that become active when the investor's portfolio approaches full

safe ($Y_t = 0$) or full risky investment ($Y_t = 1$). The idea is that the trading rate (5.4-5.5) should remain finite in each case; moreover, it should be positive at $Y_t = 0$ and negative at $Y_t = 1$ so as to keep the risky weight in $[0, 1]$.[12] Solving the ODE (3.1) at $y \in \{0, 1\}$ leads to a quadratic equation for the boundary value of q; choosing the solution with the correct sign in turn gives

$$q(0^+) = \varepsilon + 2\sqrt{\lambda\beta}, \tag{5.6}$$

$$q(1^-) = \frac{\lambda d - \varepsilon(1 - \varepsilon) - \sqrt{\lambda d(\lambda d - 2 + 2\varepsilon)}}{(1 - \varepsilon)^2}, \quad d = -\gamma\sigma^2 - 2\beta + 2\mu. \tag{5.7}$$

Together with the value matching conditions (3.4-3.5), this yields the representation from Theorem 3.1. For $y_* \in (0, 1)$, this informal derivation indeed leads to the correct answer (cf. the rigorous verification theorem in Section 7). For $y_* \notin (0, 1)$, however, the candidate risky weight explodes with positive probability, and the massive rebalancing that comes along with this reduces the corresponding wealth to zero (cf. Lemma 7.5). Hence, our candidate strategy is not even admissible in this case, and a simple buy-and-hold strategy turns out to be optimal instead (cf. Proposition 3.2). This stresses the need for rigorous verification theorems to complement heuristic considerations, which might otherwise lead to wrong results.

6 Asymptotic Results

The differential equation (3.1) in Theorem 3.1 is of Abel type; no explicit solution is known. However, it is possible to obtain asymptotic expansions for the no-trade region $[y_-, y_+]$, the trading rate $\hat{u}$, and the corresponding equivalent safe rate β as the market frictions tend to zero. Then, the computation of the optimal policy and welfare is simplified from the solution of a nonlinear free-boundary value problem to finding the root of a nonlinear function, similarly as in Korn [25] for proportional and fixed costs.

Consider the limiting regime where both the proportional transaction cost ε *and* the price impact parameter λ tend to zero. If these frictions are considered separately, their leading-order impact on the equivalent safe rate is of order $\varepsilon^{2/3}$ and $\lambda^{1/2}$, respectively (cf. [14, Formula (2.7)] resp. [19, Formula (2.13)]). To obtain an expansion in which neither friction vanishes, we rescale them appropriately to put their asymptotic contributions on the same scale:

$$\lambda = K\varepsilon^{4/3}, \quad \text{where } K > 0. \tag{6.1}$$

With this rescaling, the joint impact of the two frictions on the equivalent safe rate is of order $\varepsilon^{2/3}$ as well:

Proposition 6.1. *Assume $\lambda = K\varepsilon^{4/3}$ and define $c(\varepsilon) := \frac{\mu^2}{2\gamma\sigma^2} - \text{ESR}_\gamma(\hat{u})$, where $\hat{u}$ is the long-run optimal strategy defined in Theorem 3.1. Then*

$$0 < \liminf_{\varepsilon \to 0} \frac{c(\varepsilon)}{\varepsilon^{2/3}} \le \limsup_{\varepsilon \to 0} \frac{c(\varepsilon)}{\varepsilon^{2/3}} < +\infty.$$

Proof. See Section 7.2. □

---

[12]Recall that this is needed to ensure solvency, since portfolios involving short or levered positions lead to bankruptcy with positive probability.

The problem of finding the asymptotic expansion splits into two regions: close to the no-trade region[13] and far away from it. To study the first regime, one can argue using homogenization techniques as in Soner and Touzi [36] (see also [32, 3, 30]). To this end, first derive the HJB equation for the value function V and postulate an appropriate asymptotic expansion. Then, substitute the expansion back into HJB equation, and collect the leading-order terms. This in turn leads to the so-called *corrector* equation. In the pure transaction cost case, the appropriate ansatz for V is

$$V(t,x,y) = V_0(t,x) - \varepsilon^{2/3}u(t,x) - \varepsilon^{4/3}w(t,x,z) + \mathcal{O}(\varepsilon), \tag{6.2}$$

with the fast variable $z = \frac{y-y_*}{\varepsilon^{1/3}}$. Here, V_0 denotes the frictionless value function, x the total wealth, and y the wealth invested in the risky asset. Note that the HJB equation for V is of second order. Hence, taking the second derivative of the function w and multiplying it with $\varepsilon^{4/3}$ produces again a term which is of order $\varepsilon^{2/3}$ justifying the power $4/3$ in the expansion.

Our long-run objective forces us to work with the reduced value function q, i.e.,

$$V(t,x,y) = \frac{x^{1-\gamma}}{1-\gamma}e^{(1-\gamma)(\beta(T-t)+\int_p^y q(w)dw)}, \quad p \in \mathbb{R}. \tag{6.3}$$

Whence, the homogenization approach has to be adapted as follows. Define

$$z := \frac{y-y_*}{\varepsilon^{1/3}}, \quad q(y) := \varepsilon \times \begin{cases} r_B(z), & y \in [0, y_-], \\ r(z), & y \in [y_-, y_+], \\ r_S(z), & y \in [y_+, 1], \end{cases} \quad \beta := \frac{\mu^2}{2\gamma\sigma^2} - \varepsilon^{2/3}l, \tag{6.4}$$

for constants $0 \le y_- \le y_+ \le 1$, $l > 0$, and functions r_B, r, r_S to be determined.

To find the boundaries of the no-trade region y_- and y_+ we need two more conditions. To this end, we consider the limit behavior of q as the risky weight y approaches the full investment levels 0 and 1. Matching the terms at the leading order $\varepsilon^{2/3}$ we define $q(y) = \varepsilon^{2/3}q^*(y)$. As $\varepsilon \downarrow 0$, the ODE (3.1) then simplifies to

$$q^*(y) = \left(2K\gamma\sigma^2\right)^{1/2}(y_* - y). \tag{6.5}$$

This equation describes the behavior of q^* (and in turn q) away from the no trade region and therefore yields the required conditions to determine y_- and y_+.[14]

As $\varepsilon \downarrow 0$, the transformation (6.4) reduces the ODE (3.1) to an inhomogeneous Riccati equation:

$$0 = -\frac{\gamma\sigma^2}{2}z^2 + l + \frac{\sigma^2}{2}y_*^2(1-y_*)^2r_B' + \frac{1}{4K}(r_B - 1)^2, \quad z \in (-\infty, z_-], \tag{6.6}$$

$$0 = -\frac{\gamma\sigma^2}{2}z^2 + l + \frac{\sigma^2}{2}y_*^2(1-y_*)^2r', \qquad\qquad z \in [z_-, z_+], \tag{6.7}$$

$$0 = -\frac{\gamma\sigma^2}{2}z^2 + l + \frac{\sigma^2}{2}y_*^2(1-y_*)^2r_S' + \frac{1}{4K}(r_S + 1)^2, \quad z \in [z_+, +\infty), \tag{6.8}$$

where the rescaled buying and selling boundaries z_- and z_+ are determined by the following value matching conditions:

$$r_B(z_-) = r(z_-) = 1, \quad r_S(z_+) = r(z_+) = -1. \tag{6.9}$$

---

[13]For small transaction costs – without market price impact – the no-trade region $[y_-, y_+]$ contains the Merton proportion y_*, its width is of order $\varepsilon^{1/3}$, and the corresponding welfare effect is of order $\varepsilon^{2/3}$, see [20]. With price impact, the no-trade region is again an interval, which contains the Merton proportion provided that ε is not too small compared to λ (cf. the discussion after Lemma 7.2).

[14]A similar "pasting" of two different expansions is used in [6] to deal with a small capital gains tax.

As $z \to -\infty$ (resp. $z \to \infty$), the function r_B (resp. r_S) must diverge with the same rate as in our first expansion in (6.5):

$$\lim_{z \to -\infty} \frac{r_B(z)}{-\sqrt{2K\gamma\sigma^2}z} = 1 \quad \text{and} \quad \lim_{z \to \infty} \frac{r_S(z)}{-\sqrt{2K\gamma\sigma^2}z} = 1. \tag{6.10}$$

In summary, this leads to the ODE (6.6-6.8) with value matching and growth conditions (6.9-6.10) – the corrector equation in the present setting.

Inserting (6.5) into (6.3) with the expansion for the equivalent safe rate in (6.4) yields that:

$$\begin{aligned}
V(t,x,y) &= V_0(t,x) \cdot e^{(1-\gamma)(-\varepsilon^{2/3}l \cdot (T-t) + \varepsilon^{2/3} \int_p^y q^*(w)dw)} \\
&= V_0(t,x) - V_0(t,x)(1-\gamma)(T-t)l \cdot \varepsilon^{2/3} \\
&\quad + V_0(t,x)(1-\gamma) \int_p^y q^*(w)dw \cdot \varepsilon^{2/3} + o(\varepsilon^{2/3}).
\end{aligned} \tag{6.11}$$

The next proposition (proved in Section 7.2) shows that the ODE (6.6-6.8) with boundary conditions (6.9-6.10) admits a unique solution. As a side product, we show that additional price impact decreases the width of the no-trade region compared to the pure linear-cost case.

Proposition 6.2. *There are unique z_-, z_+ and $l_* > 0$ such that equation (6.6-6.8) has a solution that satisfies the value matching conditions (6.9) and the growth conditions (6.10). In particular,*

$$l_* > \max\left(\sqrt{\frac{\gamma K}{2}} \sigma^3 y_*^2 (1-y_*)^2, \left(\frac{3}{4}\sqrt{\frac{\gamma}{2}} \sigma^3 y_*^2 (1-y_*)^2\right)^{2/3} \right).$$

Also, if z_-^0 and z_+^0 are the boundaries of no-trade region without price impact (see [14, Formula (2.9)]), then $z_- > z_-^0$ and $z_+ < z_+^0$.

Proof. See Section 7.2. ☐

We have shown with formal arguments that the appropriately rescaled value function $q(y)$ in Theorem 3.1 converges to the solution $r(z)$ of equation (6.6-6.8) with growth conditions (6.10). The next proposition makes these heuristic arguments rigorous.

Proposition 6.3. *Assume $\lambda = K\varepsilon^{4/3}$. Let $q^\varepsilon(y)$ be the solution to equation (3.1) as in Theorem 3.1. Then $\varepsilon^{-1}q^\varepsilon(\varepsilon^{1/3}z + y_*)$ converges to the solution of equation (6.6-6.8) defined in Proposition 6.2. In particular, $\lim_{\varepsilon \to 0} \frac{c(\varepsilon)}{\varepsilon^{2/3}} = l_*$, where $c(\varepsilon) := \frac{\mu^2}{2\gamma\sigma^2} - \mathrm{ESR}_\gamma(\hat{u})$.*

Proof. See Section 7.2. ☐

To find the explicit solution of the ODE (6.6-6.8) with boundary conditions (6.9-6.10) we proceed as follows. A simple calculation shows that the linear, inhomogeneous ODE (6.7) has the explicit solution

$$r(z) = \frac{2}{\sigma^2 y_*^2 (1-y_*)^2} \left(\frac{\gamma\sigma^2}{6} z^3 - lz \right).$$

Since r is an odd function, the boundary conditions $r(z_-) = 1$ and $r(z_+) = -1$ can be replaced by $r(z_-) = 1$ and $r(0) = 0$. Then, $z_- = -z_+$ and it suffices to determine z_- because the constant l is linked to z_- via the condition $r(z_-) = 1$:

$$l(z_-) = \frac{\gamma\sigma^2}{6} z_-^2 - \frac{\sigma^2 y_*^2 (1-y_*)^2}{2z_-}. \tag{6.12}$$

11

To find the rescaled trading boundary z_- determined by $r_B(z_-) = 1$, the Riccati equation (6.6) with boundary condition (6.10) needs to be solved first.

Equation (6.6) is – up to transformations [23, Transformations 1.25, 1.105, 2.220, and 2.273, Formula (9)] – equivalent to a Whittaker equation for which explicit solutions (in terms of the Whittaker functions) are known (cf. [35, Section 1]). For brevity, we omit the derivation of the solution and simply state the final result:

$$r_B(z, l) := -\frac{1}{z}\left(\frac{1}{2a} + \frac{c}{2a}\sqrt{\frac{1}{2K\gamma\sigma^2}}\right) + 1 + \sqrt{2K\gamma\sigma^2}z - \frac{2}{az}\frac{W\left(k+1, -1/4, a\sqrt{2K\gamma\sigma^2}z^2\right)}{W\left(k, -1/4, a\sqrt{2K\gamma\sigma^2}z^2\right)}, \quad (6.13)$$

where

$$a = \frac{1}{2K\sigma^2 y_*^2(1-y_*)^2}, \quad c = \frac{2l}{\sigma^2 y_*^2(1-y_*)^2}, \quad k = \frac{c}{4}\sqrt{\frac{1}{2K\gamma\sigma^2}},$$

and the Whittaker function W is defined via the Kummer function[15] $_1F_1(a, b, x)$ and the Tricomi function $U(\xi, \eta, x)$ cf. [35, Chapter 1]:

$$U(\xi, \eta, x) := \frac{\Gamma(1-\eta)}{\Gamma(1+\xi-\eta)}\,_1F_1(\xi, \eta, x) + \frac{\Gamma(\eta-1)}{\Gamma(\xi)}x^{1-\eta}\,_1F_1(1+\xi-\eta, 2-\eta, x),$$

$$W(k, m, x) := x^{\frac{1}{2}+m}e^{-\frac{1}{2}x}U\left(1/2 + m - k, 1 + 2m, x\right).$$

We note that the general solution of the Whittaker equation is given by a linear combination of the Whittaker functions[16] $M(k, m, x)$ and $M(k, -m, x)$ (cf. [37, Section 16]) with

$$M(k, m, x) := x^{\frac{1}{2}+m}e^{-\frac{1}{2}x}\,_1F_1\left(1/2 + m - k, 1 + 2m, x\right).$$

However, the Riccati equation needs to be solved with the initial condition (6.10). Therefore, the Whittaker function $W(k, m, x)$ is the only candidate, since it has the correct asymptotic growth [37, Section 16.31]:

$$W(k, m, x) \sim x^k e^{-\frac{1}{2}x}, \quad \text{as} \quad x \to \infty. \quad (6.14)$$

Indeed, for any $l \in \mathbb{R}$, (6.13) satisfies the boundary condition (6.10):

$$\lim_{z \to -\infty}\frac{r_B(z, l)}{-\sqrt{2K\gamma\sigma^2}z} = \lim_{z \to -\infty}\frac{1}{-\sqrt{2K\gamma\sigma^2}z}\left(-\frac{1}{z}\left(\frac{1}{2a} + \frac{c}{2a}\sqrt{\frac{1}{2K\gamma\sigma^2}}\right) + 1\right)$$

$$+ \frac{\sqrt{2K\gamma\sigma^2}}{-\sqrt{2K\gamma\sigma^2}z}\left(z - \frac{2z}{a\sqrt{2K\gamma\sigma^2}z^2}\frac{W\left(k+1, -1/4, a\sqrt{2K\gamma\sigma^2}z^2\right)}{W\left(k, -1/4, a\sqrt{2K\gamma\sigma^2}z^2\right)}\right)$$

$$= 1,$$

---

[15]For $b \neq 0, -1, -2, \cdots$, the Kummer function $_1F_1(a, b, x)$ is defined through the following absolute convergent series [35, Chapter 1]:

$$_1F_1(a, b, x) := \sum_{n=0}^{\infty}\frac{a^{(n)}x^n}{b^{(n)}n!}.$$

Here, the Pochhammer symbol $a^{(n)}$ is given by $a^{(n)} := a(a+1)(a+2)\cdots(a+n-1)$. For $a \neq 0, -1, -2, \cdots$, we can write $a^{(n)} = \Gamma(a+n)/\Gamma(a)$, where $\Gamma(x)$ denotes the Gamma function.

[16]Note that the Whittaker function W (cf. [35, Formula 1.7.1]) can be written as

$$W(k, m, x) = \frac{\pi}{\sin(2m\pi)}\left(\frac{-M(k, m, x)}{\Gamma(1/2 - m - k)\Gamma(1 + 2m)} + \frac{M(k, -m, x)}{\Gamma(1/2 + m - k)\Gamma(1 - 2m)}\right).$$

12

where we have used (6.14) in the last step.

Now, we can put everything together. z_- is determined by $r_B(z_-, l(z_-)) = 1$ with $l(z_-)$ from (6.12). Hence, the asymptotic expansions for the growth rate β and the boundaries y_-, y_+ of the no-trade region in (6.4) are given by:

$$\beta = \frac{\mu^2}{2\gamma\sigma^2} - \left(\frac{\gamma\sigma^2}{6} z_-^2 - \frac{\sigma^2 y_*^2 (1 - y_*)^2}{2z_-} \right) \varepsilon^{\frac{2}{3}} + \mathcal{O}(\varepsilon),$$

$$y_\mp = y_* \pm z_- \varepsilon^{\frac{1}{3}} + \mathcal{O}(\varepsilon^{\frac{2}{3}}),$$

where z_- is the root of the equation $r_B(z, l(z)) = 1$ with $l(z)$ from (6.12). The asymptotics for the turnover rate $\hat{u}$ can be derived similarly. Recall from Theorem 3.1 that

$$\hat{u}(y) = \begin{cases} \frac{1}{2\lambda} \left(\frac{q(y)}{1 - yq(y)} - \varepsilon \right) \geq 0, & \text{if } y \in [0, y_-], \\ 0, & \text{if } y \in [y_-, y_+], \\ \frac{1}{2\lambda} \left(\frac{q(y)}{1 - yq(y)} + \varepsilon \right) \leq 0, & \text{if } y \in [y_+, 1]. \end{cases}$$

Using the same transformations as in (6.4) for y and q, a straightforward calculation shows that

$$\hat{u}(y) = \begin{cases} \frac{1}{2K} \left(r_B(z) - 1 \right) \varepsilon^{-\frac{1}{3}} + \mathcal{O}(1), & \text{if } z = (y - y_*)\varepsilon^{-1/3} \in (-\infty, z_-], \\ 0, & \text{if } z = (y - y_*)\varepsilon^{-1/3} \in [z_-, z_+], \\ \frac{1}{2K} \left(r_S(z) + 1 \right) \varepsilon^{-\frac{1}{3}} + \mathcal{O}(1), & \text{if } z = (y - y_*)\varepsilon^{-1/3} \in [z_+, +\infty), \end{cases} \tag{6.15}$$

where we abbreviate $r_B(z) := r_B(z, l(z_-))$ and $r_S(z) := r_B(z) - 2$.

In summary, asymptotically for small trading costs, the solution of the nonlinear free-boundary problem (3.1) can be reduced to finding the root of a scalar function. The approximation (6.15) performs very well, even for relatively large values of the asymptotic parameters ε, λ, cf. Figure 2. Moreover, it also allows to say more about the structure of the optimal turnover rate near the trading boundaries and far away from these.

Far away from the no-trade region, i.e., as $z \to -\infty$ or, equivalently, $y \downarrow 0$ (resp. $z \to \infty$, i.e., $y \uparrow 1$) the boundary conditions

$$\lim_{z \to -\infty} \frac{r_B(z)}{-\sqrt{2K\gamma\sigma^2 z}} = 1, \quad \text{resp.} \quad \lim_{z \to \infty} \frac{r_S(z)}{-\sqrt{2K\gamma\sigma^2 z}} = 1,$$

imply

$$\hat{u}(y) = \frac{1}{2K} \left(-\sqrt{2K\gamma\sigma^2 z} - 1 \right) \varepsilon^{-1/3} + \mathcal{O}(1)$$

$$= \frac{1}{2K} \sqrt{2K\gamma\sigma^2} (y_* - y)\varepsilon^{-2/3} + \mathcal{O}(\varepsilon^{-1/3})$$

$$= \sigma \sqrt{\frac{\gamma}{2}} (y_* - y)\lambda^{-1/2} + \mathcal{O}(\lambda^{-1/4}). \tag{6.16}$$

Hence, for large deviations from the trading boundaries we recover the leading-order expansion of the wealth turnover without proportional transaction costs [19, Formula (2.12)].

Close to the trading boundaries, we can apply Taylor's Theorem to get a first-order approximation as well. To this end, we first compute the derivative of the corresponding Whittaker functions.
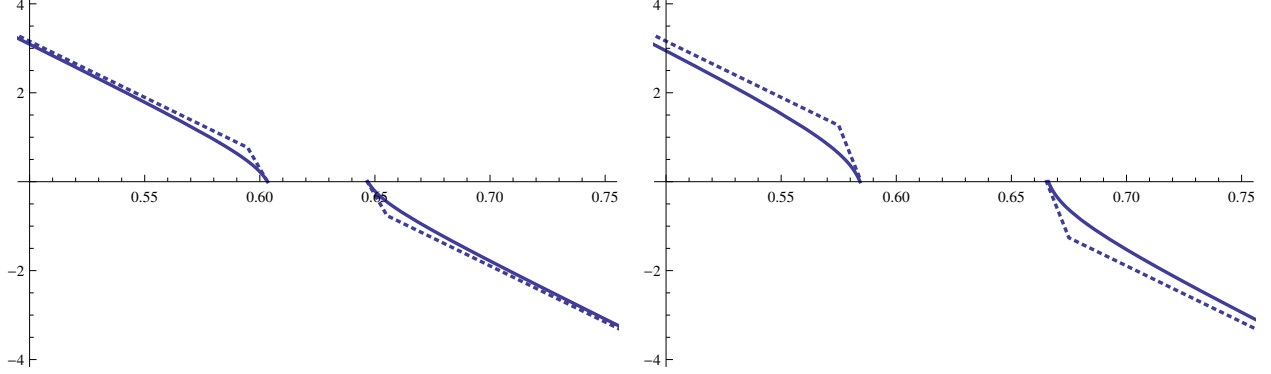
13

Figure 3: optimal wealth turnover $\hat{u}$ from Theorem 3.1 (solid), and its piecewise linear approximation (dotted) by (6.17) close to the trading boundaries and (6.16) further away from these. Model parameters are $\mu = 8\%$, $\sigma = 16\%$, $\gamma = 5$, $\lambda = 0.01\%$, and $\varepsilon = 0.1\%$ (left panel) resp. $\varepsilon = 0.5\%$ (right panel).

The differential property [35, Formula (2.4.24)] and the recurrence relation [35, Formula (2.5.11)] of $W(k, m, x)$ show that

$$
\left(\frac{W(k+1, -1/4, x)}{xW(k, -1/4, x)} \right)' = \frac{W(k+1, -1/4, x)}{xW(k, -1/4, x)} \left(\frac{W'(k+1, -1/4, x)}{W(k+1, -1/4)} - \frac{W'(k, -1/4, x)}{W(k, -1/4, x)} - \frac{1}{x} \right)
$$

$$
= \frac{W(k+1, -1/4, x)}{xW(k, -1/4, x)} \left[\frac{W(k+1, -1/4, x)}{xW(k, -1/4, x)} - \frac{2}{x} \right.
$$

$$
- \frac{1}{x^2} \left(\left(-\frac{1}{4} - (k+1) + \frac{1}{2} \right) \left(-\frac{1}{4} + (k+1) - \frac{1}{2} \right) \frac{xW(k, -1/4, x)}{W(k+1, -1/4, x)} \right)
$$

$$
\left. - (2(k+1)x - x^2) \right],
$$

where $x = a\sqrt{2K\gamma\sigma^2}z^2$. Taking into account the value matching condition $r_B(z_-) = 1$, a straightforward computation yields that the wealth turnover close to the trading boundaries $y_\pm$ is given by

$$
\hat{u}(y) = \begin{cases} \frac{\varepsilon^{-1/3}}{2K} \times F \times (z - z_-) + \mathcal{O}(1), & \text{if } z = (y - y_*)\varepsilon^{-1/3} < z_-, \\ \frac{\varepsilon^{-1/3}}{2K} \times F \times (z - z_+) + \mathcal{O}(1), & \text{if } z = (y - y_*)\varepsilon^{-1/3} > z_+, \end{cases} \tag{6.17}
$$

where

$$
x_- := a\sqrt{2K\gamma\sigma^2}z_-^2,
$$

$$
D := \frac{1}{2} \left(1 - \frac{1}{\sqrt{2K\gamma\sigma^2}z_-^2} \left(\frac{1}{2a} + \frac{c}{2a}\sqrt{\frac{1}{2K\gamma\sigma^2}} \right) \right),
$$

$$
E := D \left(D - \frac{2}{x_-} - \frac{1}{x_-^2} \left(\frac{1}{D} \left(-\frac{1}{4} - (k+1) + \frac{1}{2} \right) \left(-\frac{1}{4} + (k+1) - \frac{1}{2} \right) - (2(k+1)x_- - x_-^2) \right) \right),
$$

$$
F := \left(\frac{1}{2a} + \frac{c}{2a}\sqrt{\frac{1}{2K\gamma\sigma^2}} \right) \frac{1}{z_-^2} + \sqrt{2K\gamma\sigma^2} - 2\sqrt{2K\gamma\sigma^2}(D + 2a\sqrt{2K\gamma\sigma^2}Ez_-^2).
$$

14

In particular, for small deviations from the trading boundaries the wealth turnover – at the first order – is again linear, however with a different slope. This is illustrated in Figure 3. If the proportional costs are small compared to the price impact, then the two slopes are very similar. For larger spreads, however, turnover grows significantly faster near the trading boundaries, compare the right panel in Figure 3.

More General Asymptotics In this section we argue informally that – in the small-cost limit – the structure of the solution obtained here for the specific price-impact model (2.2) remains valid rather generally. This is consistent with results for linear and quadratic costs treated separately [22, 30][17]

To this end, consider an exogenously given diffusion process ξ with dynamics

$$d\xi_t = \mu_\xi(\xi_t)dt + \sigma_\xi(\xi_t)dW_t, \tag{6.18}$$

where μ_ξ, σ_ξ are smooth functions such that the SDE (6.18) is well-defined. This additional state variable can model the wealth of other agents trading in the market, for example. The average execution price for the large investor is in turn given by

$$S_t \left(1 + \varepsilon \cdot \mathrm{sgn}(\Delta\theta) + \lambda(\xi_t, X_t)\frac{S_t\Delta\theta}{\Delta t} \right).$$

To wit, the price impact is a general function of the large investor's wealth X and the exogenous process ξ. If $\lambda(\xi, x) = \lambda/x$ we recover the model (2.2).

An easy application of Itô's Lemma shows that in this setting, the wealth process X and the risky weight Y satisfy:

$$\frac{dX_t}{X_t} = Y_t(\mu dt + \sigma dW_t) - \varepsilon|u_t|dt - \lambda(\xi_t, X_t)X_t u_t^2 dt, \tag{6.19}$$

$$dY_t = \left(Y_t(1 - Y_t)(\mu - Y_t\sigma^2) + u_t + \varepsilon|u_t|Y_t + \lambda(\xi_t, X_t)X_t Y_t u_t^2 \right) dt + Y_t(1 - Y_t)\sigma dW_t. \tag{6.20}$$

Let $V(t, X_t, Y_t, \xi_t)$ denote the corresponding finite horizon value function, which is assumed to depend on the current wealth X_t, the current risky weight Y_t, the exogenous state ξ_t, and time t. Arguing as in Section 5, it follows that

$$V(t, x, y, \xi) = \frac{x^{1-\gamma}}{1 - \gamma} e^{(1-\gamma)(\beta(T-t)+\int_{y_*}^y q(\xi, x, u)du)}, \tag{6.21}$$

---

[17]Making these arguments rigorous would require to specify suitable integrability and boundary conditions for the additional state variables that appear in this context.

where the function $q(\xi, x, y)$ satisfies

$$
\begin{aligned}
0 = & -\beta + \mu y - \frac{\gamma \sigma^2}{2} y^2 + y(1-y)(\mu - \gamma \sigma^2 y)q + \frac{\sigma^2}{2} y^2 (1-y)^2 (q_y + (1-\gamma)q^2) \\
& + \mu y x \int_{y_*}^{y} q_x(\xi, x, u) du + \sigma^2 y^2 (1-y) \left((1-\gamma)x q(\xi, x, y) \int_{y_*}^{y} q_x(\xi, x, u) du + x q_x(\xi, x, y) \right) \\
& + \frac{\sigma^2 y^2}{2} \left[2(1-\gamma)x \int_{y_*}^{y} q_x(\xi, x, u) du + (1-\gamma)\left(x \int_{y_*}^{y} q_x(\xi, x, u) du \right)^2 \right. \\
& \left. + x^2 \int_{y_*}^{y} q_{xx}(\xi, x, u) du \right] \\
& + \mu_\xi(\xi) \int_{y_*}^{y} q_\xi(\xi, x, u) du + \frac{\sigma_\xi^2(\xi)}{2} \left(\left(\int_{y_*}^{y} q_\xi(\xi, x, u) du \right)^2 (1-\gamma) + \int_{y_*}^{y} q_{\xi\xi}(\xi, x, u) du \right) \\
& + \sigma \sigma_\xi(\xi) y(1-y) \left((1-\gamma)q(\xi, x, y) \int_{y_*}^{y} q_\xi(\xi, x, u) du + q_\xi(\xi, x, y) \right) \\
& + \sigma \sigma_\xi(\xi) y \left[(1-\gamma) \int_{y_*}^{y} q_\xi(\xi, x, u) du + (1-\gamma)x \int_{y_*}^{y} q_x(\xi, x, u) du \int_{y_*}^{y} q_\xi(\xi, x, u) du \right. \\
& \left. + x \int_{y_*}^{y} q_{\xi x}(\xi, x, u) du \right] \\
& + \begin{cases} \frac{1}{4\lambda(\xi, x)x} \frac{(q - \varepsilon(1 - yq + x \int_{y_*}^{y} q_x(\xi, x, u) du))^2}{1 - yq + x \int_{y_*}^{y} q_x(\xi, x, u) du}, & \text{if } y \in [0, y_-(\xi, x)], \\ 0, & \text{if } y \in [y_-(\xi, x), y_+(\xi, x)], \\ \frac{1}{4\lambda(\xi, x)x} \frac{(q + \varepsilon(1 - yq + x \int_{y_*}^{y} q_x(\xi, x, u) du))^2}{1 - yq + x \int_{y_*}^{y} q_x(\xi, x, u) du}, & \text{if } y \in [y_+(\xi, x), 1]. \end{cases}
\end{aligned} \tag{6.22}
$$

and

$$
q(\xi, x, y_-(\xi, x)) = \frac{\varepsilon + \varepsilon x \int_{y_*}^{y_-(\xi, x)} q_x(\xi, x, u) du}{1 + \varepsilon y_-(\xi, x)},
$$

$$
q(\xi, x, y_+(\xi, x)) = \frac{-\varepsilon - \varepsilon x \int_{y_*}^{y_+(\xi, x)} q_x(\xi, x, u) du}{1 - \varepsilon y_+(\xi, x)},
$$

for functions $0 \le y_-(\xi, x) \le y_+(\xi, x) \le 1$ to be determined. To derive the corresponding small-cost asymptotics, rescale transaction costs and price impact accordingly:

$$
\lambda(\xi, x)x = K(\xi, x)\varepsilon^{4/3},
$$

for some function K. Close to the no-trade region we again apply the homogenization approach and define

$$
z := \frac{y - y_*}{\varepsilon^{1/3}}, \quad q(\xi, x, y) := \varepsilon \times \begin{cases} r_B(\xi, x, z), & y \in [0, y_-(\xi, x)], \\ r(\xi, x, z), & y \in [y_-(\xi, x), y_+(\xi, x)], \\ r_S(\xi, x, z), & y \in [y_+(\xi, x), 1], \end{cases} \quad \beta := \frac{\mu^2}{2\gamma\sigma^2} - l\varepsilon^{2/3}, \tag{6.23}
$$

for a constant[18] $l > 0$ and functions $0 \le y_-(\xi, x) \le y_+(\xi, x) \le 1$, as well as $r_B(\xi, x, z)$, $r(\xi, x, z)$, $r_S(\xi, x, z)$ to be determined. To identify the boundaries of the no-trade region we need additional

---

[18]Note that the optimal long-run growth rate is typically constant even in factor models with an additional state variable, compare [17].

conditions. To this end, we again consider $q(\xi, x, y) = \varepsilon^{2/3} q^*(\xi, x, y)$. As $\varepsilon \downarrow 0$, a direct calculation shows that the ODE (6.22) reduces to

$$q^*(\xi, x, y) = (2K(\xi, x)\gamma\sigma^2)^{1/2}(y_* - y). \tag{6.24}$$

If the functions r_B, r, r_S are sufficiently smooth, then as $\varepsilon \downarrow 0$ transformation (6.23) reduces[19] the ODE (6.22) to

$$0 = -\frac{\gamma\sigma^2}{2}z^2 + l + \frac{\sigma^2}{2}y_*^2(1-y_*)^2 r_B' + \frac{1}{4K(\xi, x)}(r_B - 1)^2, \quad z \in (-\infty, z_-(\xi, x)], \tag{6.25}$$

$$0 = -\frac{\gamma\sigma^2}{2}z^2 + l + \frac{\sigma^2}{2}y_*^2(1-y_*)^2 r', \qquad\qquad z \in [z_-(\xi, x), z_+(\xi, x)], \tag{6.26}$$

$$0 = -\frac{\gamma\sigma^2}{2}z^2 + l + \frac{\sigma^2}{2}y_*^2(1-y_*)^2 r_S' + \frac{1}{4K(\xi, x)}(r_S + 1)^2, \quad z \in [z_+(\xi, x), +\infty), \tag{6.27}$$

where the rescaled buying and selling boundaries $z_-(\xi, x)$ and $z_+(\xi, x)$ satisfy:

$$r_B(\xi, x, z_-(\xi, x)) = r(\xi, x, z_-(\xi, x)) = 1, \quad r_S(\xi, x, z_+(\xi, x)) = r(\xi, x, z_+(\xi, x)) = -1. \tag{6.28}$$

As $z \to -\infty$ (resp. $z \to \infty$) the function $r_B(\xi, x, \cdot)$ (resp. $r_S(\xi, x, \cdot)$) must diverge with the same rate as in (6.24):

$$\lim_{z \to -\infty} \frac{r_B(\xi, x, z)}{-\sqrt{2K(\xi, x)\gamma\sigma^2}z} = 1 \quad \text{and} \quad \lim_{z \to \infty} \frac{r_S(\xi, x, z)}{-\sqrt{2K(\xi, x)\gamma\sigma^2}z} = 1. \tag{6.29}$$

As a result, the asymptotic trading boundaries and trading rate are still determined by an inhomogenous Riccati equation, but by a different one for each value of the additional state variables. To wit, the asymptotic solutions for more general cost structures are obtained by plugging in the current value of the cost into the asymptotic expansions derived in the previous section.

7 Proofs

7.1 Proof of Theorem 3.1

Assume throughout that $y_* \in (0, 1)$. The first step towards a rigorous verification theorem is to show that the differential equation (3.1) indeed admits a solution with the required properties (3.2-3.5). To this end, we first rewrite (3.1) in slope field notation:[20]

$$q' = f(y, q) = \begin{cases} f_B(y, q), & q \geq \frac{\varepsilon}{1+\varepsilon y}, \\ f_{NT}(y, q), & -\frac{\varepsilon}{1-\varepsilon y} \leq q \leq \frac{\varepsilon}{1+\varepsilon y}, \\ f_S(y, q), & q \leq -\frac{\varepsilon}{1-\varepsilon y}, \end{cases} \tag{7.1}$$

---

[19]For each $h \in \{r_B, r, r_S\}$, note that if $h_\xi(\xi, x, \cdot)$ is continuous then

$$\int_{y_*}^y q_\xi(\xi, x, u)du = \varepsilon \int_{y_*}^y h_\xi(\xi, x, (u - y_*)\varepsilon^{-1/3})du = \varepsilon^{4/3}\int_0^z h_\xi(\xi, x, s)ds = o(\varepsilon^{2/3}).$$

[20]This will be justified a posteriori by the monotonicity properties of q established in Lemma 7.2.

where

$$-\beta + \mu y - \frac{\gamma\sigma^2}{2}y^2 + y(1-y)(\mu - \gamma\sigma^2 y)q + \frac{\sigma^2}{2}y^2(1-y)^2(q' + (1-\gamma)q^2)$$

$$+ \begin{cases} \frac{1}{4\lambda}\frac{(q-\varepsilon(1-yq))^2}{1-yq} & =: f_B(y,q), \\ 0 & =: f_{NT}(y,q), \\ \frac{1}{4\lambda}\frac{(q+\varepsilon(1-yq))^2}{1-yq} & =: f_S(y,q). \end{cases}$$

Notice that $f(y,q)$ is well defined because $f_B(y,q) = f_{NT}(y,q)$ on $q = \frac{\varepsilon}{1+\varepsilon y}$ and $f_{NT}(y,q) = f_S(y,q)$ on $q = -\frac{\varepsilon}{1-\varepsilon y}$.

Remark 7.1. *Allocating the entire wealth into the riskless asset (resp. the risky asset) is an admissible strategy that does not require trading. The corresponding equivalent safe rate is 0 (resp. $\mu - \frac{\gamma\sigma^2}{2}$). This provides a natural lower bound for the optimal equivalent safe rate, namely $\beta \geq \max\{\mu - \frac{\gamma\sigma^2}{2}, 0\}$. Conversely, an upper bound is given by the frictionless equivalent safe rate $\frac{\mu^2}{2\gamma\sigma^2}$.*

Lemma 7.2. *Suppose λ and ε are sufficiently small. Then, for a suitable $\beta \in \left[\max\{\mu - \frac{\gamma\sigma^2}{2}, 0\}, \frac{\mu^2}{2\gamma\sigma^2}\right]$, there is a solution of $q' = f(y,q)$ such that*

$$q(0^+) = b_0(\varepsilon, \lambda) := \varepsilon + 2\sqrt{\lambda\beta}, \tag{7.2}$$

$$q(1^-) = b_1(\varepsilon, \lambda) := \frac{\lambda d - \varepsilon(1-\varepsilon) - \sqrt{\lambda d(\lambda d - 2 + 2\varepsilon)}}{(1-\varepsilon)^2}, \quad \text{with } d = -\gamma\sigma^2 - 2\beta + 2\mu. \tag{7.3}$$

In particular, there exist $y_-, y_+ \in [0,1]$ satisfying (3.1) and

$$\begin{cases} q(y) > \frac{\varepsilon}{1+\varepsilon y}, & \text{if } y \in [0, y_-), \\ \frac{-\varepsilon}{1-\varepsilon y} \leq q(y) \leq \frac{\varepsilon}{1+\varepsilon y}, & \text{if } y \in [y_-, y_+], \\ q(y) \leq \frac{-\varepsilon}{1-\varepsilon y}, & \text{if } y \in (y_+, 1]. \end{cases} \tag{7.4}$$

Moreover, the solution q fulfills $q(y)y < 1$ for all $y \in [0,1]$.

Proof. First, notice that for every $y \in (0,1)$ we have $\frac{1}{y} > \frac{\varepsilon}{1+\varepsilon y}$, and thus $\lim_{q \to (\frac{1}{y})^-} f(y,q) = \lim_{q \to (\frac{1}{y})^-} f_B(y,q) = -\infty$. Hence, every solution to $q' = f(y,q)$ starting below the curve $\frac{1}{y}$ must remain below this curve. The rest of the proof proceeds as follows:

(i) For every $\beta > \max\{\mu - \frac{\gamma\sigma^2}{2}, 0\}$, there is a unique solution $q_0^\beta(y)$ defined on $(0, y_0) \subset (0,1)$ that satisfies the boundary condition (7.2) at $y = 0$ and a unique solution $q_1^\beta(y)$ defined on $(y_1, 1) \subset (0,1)$ that satisfies the boundary condition (7.3) at $y = 1$.

(ii) If $\beta > \frac{\mu^2}{2\gamma\sigma^2}$, then $q_0^\beta(y) > 0$ and $0 > q_1^\beta(y)$ on their respective definition intervals.

(iii) Set $\beta = \frac{\mu^2}{2\gamma\sigma^2} - c$, with $c > 0$. If λ and ε are sufficiently small, we have $q_0^\beta(y) < q_1^\beta(y)$ on $(0, y_0) \cap (y_1, 1)$. Additionally, if $y_0 < 1$ then $\lim_{y \to y_0^-} q_0^\beta(y) = -\infty$, and if $y_1 > 0$ then $\lim_{y \to y_1^+} q_1^\beta(y) = +\infty$.

The solution to $q' = f(y,q)$ depends continuously on β. Hence, if λ and ε are sufficiently small, we therefore have $q_0^{\beta^*} \equiv q_1^{\beta^*}$ for some $\beta^* \in \left[\max\{\mu - \gamma\sigma^2/2, 0\}, \mu^2/(2\gamma\sigma^2)\right]$.

18

Proof of (i): This follows similarly as in [19, Lemma A.8(i)]. Indeed, replace $q^2/4\lambda$ by $(q-\varepsilon)^2/4\lambda$ in the first line of the proof of [19, Lemma A.8(i)]. Then, the proof proceeds analogously, leading to

$$h(0) = \varepsilon + 2\sqrt{\lambda\beta}, \quad h'(0) = -\left(\frac{\lambda}{\beta}\right)^{1/2}(\mu + (\mu + \beta)h(0)) - \varepsilon h(0) < 0.$$

This term is negative like the corresponding expression in the first displayed equation in [19, Lemma A.8(i)]. Hence, the remaining steps can be carried through unchanged.

Proof of (ii): This follows verbatim as in [19, Lemma A.8(iii)].

Proof of (iii): Here, additional work is required compared to [19]. The proof is based on the following observations:

Remark 7.3. *On* $\{(y,q) : y \in (0,1), q < \frac{1}{y}\}$, *we have* $f(y,q) \leq f_{NT}(y,q)$.

Remark 7.4. *There are* $\alpha, \zeta > 0$, *and* $\eta < 0$ *independent of* λ *and* ε *such that*

$$f(y,q) \leq f_{NT}(y,q) \leq \eta < 0, \quad \text{on } [y_* - \alpha, y_* + \alpha] \times [-\zeta, \zeta].$$

Proof of Remark 7.4. On $[y_-, y_+]$, Equation (3.1) can be rewritten as

$$y^2(1-y)^2 f_{NT}(y,q) = -c + k_1(y,q) + k_2(y,q),$$

where $\lim_{y \to y_*} k_1(y,q) = 0$ and $\lim_{q \to 0} k_2(y,q) = 0$ uniformly. In particular, there is a negative constant η such that, if y is sufficiently close to y_* and q to 0, we have $f_{NT}(y,q) \leq \eta$. $\square$

A straightforward computation shows that one can choose λ_1 and ε_1 small enough such that, for any $\lambda \leq \lambda_1$ and $\varepsilon \leq \varepsilon_1$:

- $\frac{d}{dy}[y^2(1-y)^2 f_{NT}(y, b_0(\lambda, \varepsilon))] < 0$ on $(0, y_* - \alpha)$;

- $\frac{d}{dy}[y^2(1-y)^2 f_{NT}(y, b_1(\lambda, \varepsilon))] > 0$ on $(y_* + \alpha, 1)$;

- $\max\{b_0, -b_1\} < \zeta$ and $\eta < \frac{b_1(\lambda,\varepsilon) - b_0(\lambda,\varepsilon)}{2\alpha} < 0$.

From the first point, and since $y^2(1-y)^2 f_{NT}(y, b_0(\lambda, \varepsilon)) = 0$ for $y = 0$, we get $f_{NT}(y, b_0(\lambda, \varepsilon)) < 0$ on $(0, y_* - \alpha)$. Remark 7.3 implies that $f(y, b_0(\lambda, \varepsilon)) < 0$ on $(0, y_* - \alpha)$.

The second point, with the same arguments, yields $f(y, b_1(\lambda, \varepsilon)) < 0$ on $(y_* + \alpha, 1)$.

Consider now the line $q = \frac{b_1 - b_0}{2\alpha}(y - y_* - \alpha) + b_1$. Given a solution $q(y)$, the third point and Remark 7.4 imply that if $q(y_* - \alpha) < b_0$ then $q(y) < \frac{b_1 - b_0}{2\alpha}(y - y_* - \alpha) + b_1$ on $[y_* - \alpha, y_* + \alpha]$. Define the following function:

$$g(y) = \begin{cases} b_0, & y \in (0, y_* - \alpha), \\ \frac{b_1 - b_0}{2\alpha}(y - y_* - \alpha) + b_1, & y \in [y_* - \alpha, y_* + \alpha], \\ b_1, & y \in (y_* + \alpha, 1). \end{cases} \tag{7.5}$$

We have just shown that $f(y, g(y)) < g'(y)$ on $(0,1) \setminus \{y_* - \alpha, y_* + \alpha\}$, thus $q_0^\beta(y) < g(y)$ on the definition interval of q_0. Analogously, $q_1^\beta(y) > g(y)$ and so $q_0^\beta(y) < q_1^\beta(y)$ on their common definition interval. This completes the proof of Item (iii).

We have shown that $q_0^\beta(y) > q_1^\beta(y)$ for $\beta > \frac{\mu^2}{2\gamma\sigma^2}$ and $q_1^\beta(y) > q_0^\beta(y)$ for $\beta = \frac{\mu^2}{2\gamma\sigma^2} - c$ with $c > 0$ and sufficiently small ε and λ on their common domain. This proves that there exists $\beta^* = \frac{\mu^2}{2\gamma\sigma^2} - c(\varepsilon, \lambda)$ and $\bar{y} \in (0, 1)$ such that $q_0^{\beta^*}(\bar{y}) = q_1^{\beta^*}(\bar{y})$. Since both $q_0^{\beta^*}(\cdot)$ and $q_1^{\beta^*}(\cdot)$ satisfy equation (3.1), $(q_0^{\beta^*})'(\bar{y}) = (q_1^{\beta^*})'(\bar{y})$ and hence $q_0^{\beta^*}(y) = q_1^{\beta^*}(y)$ on the whole interval $(0, 1)$. Therefore, we found a solution $q^{\beta^*}(y)$ satisfying conditions (7.2) and (7.3). Finally, we show that the set $\{y : -\varepsilon < \frac{q^{\beta^*}(y)}{1 - yq^{\beta^*}(y)} < \varepsilon\}$ is an interval $[y_-, y_+]$. In particular, it is enough to show that the solution $q^{\beta^*}(y)$ crosses the curves $q = \frac{\varepsilon}{1+\varepsilon y}$ and $q = -\frac{\varepsilon}{1-\varepsilon y}$ just once, in y_- and y_+, respectively.

The equation $f_{NT}(y, \frac{\varepsilon}{1+\varepsilon y}) = \frac{d}{dy}(\frac{\varepsilon}{1+\varepsilon y})$ has exactly two solutions $y_2 < y_3$ in $(0, 1)$. In particular, $f_{NT}(y, \frac{\varepsilon}{1+\varepsilon y}) > \frac{d}{dy}(\frac{\varepsilon}{1+\varepsilon y})$ on $(0, y_2) \cup (y_3, 1)$ and $f_{NT}(y, \frac{\varepsilon}{1+\varepsilon y}) < \frac{d}{dy}(\frac{\varepsilon}{1+\varepsilon y})$ on (y_2, y_3). Since $q^{\beta^*}(0^+) > \varepsilon$ and $q^{\beta^*}(1^-) < \frac{\varepsilon}{1+\varepsilon}$, the solution $q^{\beta^*}(y)$ crosses $q = \frac{\varepsilon}{1+\varepsilon y}$ just once in $y_- \in (y_2, y_3)$. By the same arguments, $q^{\beta^*}(y)$ crosses $q = -\frac{\varepsilon}{1-\varepsilon y}$ just once in y_+. $\qquad\square$

Unlike for small proportional transaction costs [20, 14], the frictionless Merton proportion y_* does not generally lie in the no-trade region $[y_-, y_+]$ in the present setting. To see this, recall from Guasoni and Weber [19, Remark A.13] that in their model with price impact – but without proportional transaction costs – turnover is zero at exactly one point $y_\dagger$, which is $\mathcal{O}(\lambda^{1/2})$-close but not identical to y_* for small λ. For a given small price impact λ, the trading boundaries y_-, y_+ converge to $y_\dagger$ as $\varepsilon \downarrow 0$. Hence, $y_* \notin [y_-, y_+]$ if the transaction cost is sufficiently small compared to the price impact. However, numerical evidence indicates that this effect only appears if the ratio ε/λ is extremely small. Otherwise, the Merton proportion is contained in the no-trade region, compare Figures 1 and 2.

As shown in [19, Theorem 2.3], levered or short positions in the risky asset cannot be admissible with linear price impact. This remains true in the present setting with additional proportional costs:

Lemma 7.5. *Let u be an admissible strategy. Then, for sufficiently small ε and λ, the corresponding risky weight*

$$dY_t = (Y_t(1 - Y_t)(\mu - Y_t\sigma^2) + (u(Y_t) + \varepsilon Y_t|u(Y_t)| + \lambda Y_t u(Y_t)^2))dt + Y_t(1 - Y_t)\sigma dW_t, \qquad (7.6)$$
$$Y_0 = y \in (0, 1), \qquad\qquad (7.7)$$

takes values in $[0, 1]$ a.s. for all t.

Proof. With minor modifications, the assertion follows along the lines of [19, Lemma A.2, Lemma A.3, Lemma A.4, and Theorem 2.3].[21] For the sake of completeness, we briefly recall the main ideas here. Let u be any admissible strategy. First, verify that a stochastic process with the dynamics (7.6) and initial value $y \in (1, \infty)$ resp. $y \in (-\infty, 0)$ has a finite exploding time τ with positive probability, i.e., $\mathbb{P}[\tau < \infty] > 0$. In a second step, show that the corresponding wealth process X^u satisfies, $X_\tau^u(\omega) = 0$ a.s. on $\{\tau < \infty\}$ and $\theta_\tau > 0$ a.s. This in turn implies that any admissible strategy must fulfill $Y_t \in [0, 1]$ a.s. for all t. $\qquad\square$

Next, verify that the candidate strategy $\hat{u}$ from Theorem 3.1 is admissible.

---

[21] In the proof of [19, Lemma A.2], use $-(1 - \varepsilon y)/4\lambda$ instead of $-1/4\lambda$ in the definition of $\tilde{\mu}$; then the proof can be carried through along the same lines. In the proof of Lemma 7.5, our analogue of [19, Theorem 2.3], proportional transaction costs lead to an additional term $\int_0^T |\dot{\theta}_t|dt$ in the numerator of the expression analyzed in [19, Lemma A.4]. However, the latter still converges to zero by the same arguments as in the proof of [19, Lemma A.4].

Lemma 7.6. *Define β, q, y_-, y_+ as in Lemma 7.2 and set*

$$\hat{u}(y) = \begin{cases} \frac{1}{2\lambda}\left(\frac{q(y)}{1-yq(y)} - \varepsilon\right), & \text{if } y \in [0, y_-), \\ 0, & \text{if } y \in [y_-, y_+], \\ \frac{1}{2\lambda}\left(\frac{q(y)}{1-yq(y)} + \varepsilon\right), & \text{if } y \in (y_+, 1]. \end{cases}$$

Then, for sufficiently small ε and λ, the SDE

$$dY_t^{\hat{u}} = (Y_t^{\hat{u}}(1 - Y_t^{\hat{u}})(\mu - Y_t^{\hat{u}}\sigma^2) + (\hat{u}(Y_t^{\hat{u}}) + \varepsilon Y_t^{\hat{u}}|\hat{u}(Y_t^{\hat{u}})| + \lambda Y_t^{\hat{u}}\hat{u}(Y_t^{\hat{u}})^2))dt + Y_t^{\hat{u}}(1 - Y_t^{\hat{u}})\sigma dW_t,$$
$$Y_0^{\hat{u}} = y \in (0, 1)$$

has a unique strong solution which takes values in $[0, 1]$ a.s. for all t. In particular, the strategy $\hat{u}$ is admissible.

Proof. Lemma 7.2 shows that $\hat{u}(y)$ is a bounded, continuous function on $[0, 1]$ which satisfies $\hat{u}(0) > 0$ and $\hat{u}(1) < 0$, i.e., the strategy buys at full investment and sells at zero investment. Furthermore, notice that the scale function of the process $Y^{\hat{u}}$ is given by

$$s(x) = \int_c^x \exp\left[-2\int_c^y \frac{z(1-z)(\mu - z\sigma^2) + \hat{u}(z) + \varepsilon z|\hat{u}(z)| + \lambda z\hat{u}(z)^2}{z^2(1-z)^2\sigma^2}dz\right]dy,$$

for $c \in (0, 1)$. For sufficiently small ε and λ, $\hat{u}(1) + \varepsilon|\hat{u}(1)| + \lambda\hat{u}(1)^2 < 0$. A straightforward computation shows that $s(0^+) = -\infty$ and $s(1^-) = \infty$, so that [24, Proposition 5.5.22] yields the first assertion. Finally, since $\hat{u}$ and $Y^{\hat{u}}$ are both bounded, the admissibility of the strategy follows. $\square$

With the function q, the constant β, and the boundaries y_-, y_+ of the no-trade region at hand, we can now use a variant of the verification argument[22] of Guasoni and Roberston [17, Theorem 7] to compute an upper bound for the equivalent safe rate of any admissible strategy:

Lemma 7.7. *Let $y \in (0, 1)$ be the initial risky weight, define β, q as in Lemma 7.2, and set $Q(\xi) = \int_0^\xi q(z)dz$. Then, the terminal wealth X_T^u of any given admissible strategy u satisfies:*

$$\mathbb{E}[(X_T^u)^{1-\gamma}]^{\frac{1}{1-\gamma}} \le X_0 e^{\beta T + Q(y)} \mathbb{E}^{\hat{\mathbb{P}}^u}\left[e^{-(1-\gamma)Q(Y_T^u)}\right]^{\frac{1}{1-\gamma}}, \tag{7.8}$$

where

$$\frac{d\hat{\mathbb{P}}^u|_{\mathcal{F}_T}}{d\mathbb{P}|_{\mathcal{F}_T}} = \mathcal{E}\left(\int_0^\cdot (1-\gamma)Y_s^u(1 + q(Y_s^u)(1 - Y_s^u))\sigma dW_s\right)_T. \tag{7.9}$$

Moreover, equality holds in (7.8) for the strategy $\hat{u}$ from Lemma 7.6.

Proof. Fix an admissible strategy u and omit the $u-$dependence of X, Y, and $\hat{\mathbb{P}}$ for the sake of clarity in the rest of the proof. Lemma 7.2, Lemma 7.5, and Novikov's Condition imply that the stochastic exponential on the right-hand side of (7.9) is a true martingale and therefore the density process of $\hat{\mathbb{P}}^u$ with respect to $\mathbb{P}$.

---

[22] The verification argument used in the proof was first used by Guasoni and Robertson [17, Theorem 7] in a general Markovian frictionless setting. It was in turn adapted by Guasoni and Weber [19, Lemma A.7] to a Black-Scholes model with quadratic trading costs. Here, we extend it to a setting with quadratic and linear trading costs, which leads to an additional nonlinearity in the HJB equation (7.14).

Now, one readily checks that the assertion follows from

$$\log X_T - \log X_0 - \frac{1}{1-\gamma}\log\left(\frac{d\hat{\mathbb{P}}}{d\mathbb{P}}\right) \leq \beta T - Q(Y_T) + Q(y). \tag{7.10}$$

To verify (7.10), recall the dynamics of the wealth process X and the risky weight Y from (2.4-2.5) and apply Itô's formula to $Q(Y_T)$ and $\log(X_T)$, obtaining:

$$Q(Y_T) - Q(y) = \int_0^T q(Y_t)(Y_t(1-Y_t)(\mu - Y_t\sigma^2) + u_t + \varepsilon|u_t|Y_t + \lambda Y_t u_t^2)dt$$

$$+ \int_0^T \frac{1}{2}q'(Y_t)Y_t^2(1-Y_t)^2\sigma^2 dt + \int_0^T q(Y_t)Y_t(1-Y_t)\sigma dW_t, \tag{7.11}$$

$$\log X_T - \log X_0 = \int_0^T \left(Y_t\mu - \varepsilon|u_t| - \lambda u_t^2 - \frac{1}{2}\sigma^2 Y_t^2\right)dt + \int_0^T \sigma Y_t dW_t. \tag{7.12}$$

After substituting (7.11-7.12) into (7.10), this inequality reads as

$$\int_0^T \mu Y_t - \varepsilon|u_t| - \lambda u_t^2 - \frac{1}{2}\sigma^2 Y_t^2 + \frac{1}{2}\sigma^2(1-\gamma)Y_t^2(1 + q(Y_t)(1-Y_t))^2 dt$$

$$\leq \int_0^T \left(\beta - q(Y_t)(Y_t(1-Y_t)(\mu - Y_t\sigma^2) + u_t + \varepsilon|u_t|Y_t + \lambda Y_t u_t^2) - \frac{1}{2}\sigma^2 Y_t^2(1-Y_t)^2 q'(Y_t)\right)dt.$$

Hence, it remains to verify that, for all $u \in \mathbb{R}$ and $y \in [0,1]$:

$$\mu y - \varepsilon|u| - \lambda u^2 - \frac{\sigma^2}{2}y^2 + \frac{1-\gamma}{2}\sigma^2 y^2(1 + q(y)(1-y))^2$$

$$\leq \beta - q(y)(y(1-y)(\mu - y\sigma^2) + u + \varepsilon|u|y + \lambda y u^2) - \frac{\sigma^2}{2}y^2(1-y)^2 q'(y). \tag{7.13}$$

Rearranging (7.13), it suffices to check that, for all $u \in \mathbb{R}$ and $y \in [0,1]$:

$$0 \geq -\beta + \mu y - \frac{\gamma\sigma^2}{2}y^2 + y(1-y)(\mu - \gamma\sigma^2 y)q + \frac{\sigma^2}{2}y^2(1-y)^2(q' + (1-\gamma)q^2)$$

$$- \lambda u^2 - \varepsilon|u| + (u + \varepsilon|u|y + \lambda y u^2)q. \tag{7.14}$$

Maximizing $-\lambda u^2 - \varepsilon|u| + (u + \varepsilon|u|y + \lambda y u^2)q$ over u shows that the maximum[23] is attained at

$$\tilde{u}(y) = \begin{cases} \frac{1}{2\lambda}\left(\frac{q(y)}{1-yq(y)} - \varepsilon\right), & \text{if } \frac{q(y)}{1-yq(y)} \geq \varepsilon, \\ 0, & \text{if } -\varepsilon \leq \frac{q(y)}{1-yq(y)} \leq \varepsilon, \\ \frac{1}{2\lambda}\left(\frac{q(y)}{1-yq(y)} + \varepsilon\right), & \text{if } \frac{q(y)}{1-yq(y)} \leq -\varepsilon. \end{cases}$$

In view of (7.4),

$$\tilde{u}(y) = \begin{cases} \frac{1}{2\lambda}\left(\frac{q(y)}{1-yq(y)} - \varepsilon\right), & \text{if } y \in [0, y_-), \\ 0, & \text{if } y \in [y_-, y_+], \\ \frac{1}{2\lambda}\left(\frac{q(y)}{1-yq(y)} + \varepsilon\right), & \text{if } y \in (y_+, 1]. \end{cases}$$

Now, the inequality (7.14) follows after substituting the ODE (3.1) for q and using the maximality of $\tilde{u}$. Evidently, this inequality becomes an equality for the strategy $\hat{u}$ from Lemma 7.6. $\square$

---

[23]The condition $q(y)y < 1$ on $[0,1]$ guarantees that the critical point is indeed a maximum.

To complete the proof of Theorem 3.1 we now verify that, as $T \to \infty$, the upper bound in Lemma 7.7 converges to β for any admissible strategy, and is attained for $\hat{u}$ from Lemma 7.6.

Proof of Theorem 3.1. Let β and q be defined as in Lemma 7.2 and let u be an arbitrary admissible strategy. By Lemma 7.5, we have $Y_t^u \in [0,1]$ for all t. As q is bounded on $[0,1]$ due to Lemma 7.2, the function $Q(\xi) = \int_0^\xi q(z)dz$ is also bounded on $[0,1]$. Thus, for every admissible strategy, we have

$$\lim_{T\to\infty} \frac{1}{(1-\gamma)T} \log \mathbb{E}^{\hat{\mathbb{P}}^u}[e^{-(1-\gamma)(Q(Y_T^u)-Q(y))}] = 0.$$

As $T \to \infty$, Lemma 7.7 therefore provides a strategy-*independent* upper bound for the equivalent safe rate:

$$\mathrm{ESR}_\gamma(u) = \lim_{T\to\infty} \frac{1}{(1-\gamma)T} \log \mathbb{E}[(X_T^u)^{1-\gamma}] \leq \beta.$$

This upper bound is attained for the admissible strategy $\hat{u}$ from Lemma 7.6. Hence, the latter is long-run optimal with equivalent safe rate β as claimed. $\qquad\square$

7.2 Proof of Propositions 6.1, 6.2, and 6.3

In this final section, we provide proofs for the asymptotic results from Section 6.

Lemma 7.8. *Assume $\lambda = K\varepsilon^{4/3}$. For sufficiently small ε, the solution $q(y)$ defined in Theorem 3.1 is strictly decreasing.*

Proof. Rewrite equation (3.1) as $q'(y) = f(y, q(y))$. In the proof of Theorem 3.1 we show that there exists a function $h(y)$ defined close to 0 such that $f(y, h(y)) = 0$, $h(0^+) = b_0$ and $h'(0^+) < 0$. Notice that $f(y, \varepsilon \pm 2\sqrt{K\beta}\varepsilon^{2/3}) < 0$, therefore $|h(y)-\varepsilon| \leq 2\sqrt{K\beta}\varepsilon^{2/3}$, and in particular $h(y) < \frac{1}{y}$ for small ε. The equation $f(y, h) = 0$ can then be seen as a cubic equation in h (that reduces to a quadratic equation when $h < \frac{\varepsilon}{1+\varepsilon y}$), and there exists N_0 such that eventually (when ε goes to 0) this equation has positive discriminant on $(0, y_* - N_0\sqrt{c(\varepsilon)})$, where $c(\varepsilon) := \frac{\mu^2}{2\gamma\sigma^2} - \beta$ (and discriminant equal to 0 at $y_* - N_0\sqrt{c(\varepsilon)}$). In this case there are explicit expressions for the three real roots of $f(y, h) = 0$ in terms of y, one of these roots extends the function $h(y)$ to a continuous function on the whole interval $(0, y_* - N_0\sqrt{c(\varepsilon)})$. Notice that on $(y_* - N_0\sqrt{c(\varepsilon)}, y_*)$ the cubic equation has only one real root that does not lie in (b_1, b_0), in other terms $f(y, q) < 0$ on $(y_* - N_0\sqrt{c(\varepsilon)}, y_*) \times (b_1, b_0)$. Using the fact that $|h(y)-\varepsilon| \leq 2\sqrt{K\beta}\varepsilon^{2/3}$, we see that there does not exist any $y \in (0, y_* - N_0\sqrt{c(\varepsilon)})$ such that simultaneously $\frac{d}{dy}f(y, h(y)) = 0$ and $h'(y) = 0$. This implies that $h(y)$ is monotone on its whole definition interval, in particular (since $h'(0^+) < 0$) it is a decreasing function. The solution $q(y)$ to (3.1) defined in Theorem 3.1 is such that $q(0^+) = b_0$ and $q'(0^+) = 0$, while $h(0^+) = b_0$ and $h'(0^+) < 0$, therefore close to 0 we have $q > h$. Since h is a subsolution, i.e., $0 = f(y, h(y)) > h'(y)$, we have $q(y) > h(y)$ on $(0, y_* - N_0\sqrt{c(\varepsilon)})$. This implies that $q'(y) = f(y, q(y)) < 0$ on $(0, y_* - N_0\sqrt{c(\varepsilon)})$.
A similar argument proves that $q'(y) < 0$ on $(y_* + N_1\sqrt{c(\varepsilon)}, 1)$ for some $N_1 > 0$. To conclude, it is enough to observe that $f(y, q) < 0$ on $(y_* - N_0\sqrt{c(\varepsilon)}, y_* + N_1\sqrt{c(\varepsilon)}) \times (b_1, b_0)$. $\qquad\square$

Proposition 6.1 shows that the equivalent safe loss, i.e., the positive value $c(\varepsilon) := \frac{\mu^2}{2\gamma\sigma^2} - \mathrm{ESR}_\gamma(\hat{u})$, is of order $\varepsilon^{2/3}$. The lower bound for $c(\varepsilon)$ is obtained by comparing it to the loss that would occur if the investor were facing only quadratic costs, while the upper bound is obtained by adapting arguments for a model with nonlinear trading costs of power form [18]:

Proof of Proposition 6.1. For any admissible strategy u, let $\mathrm{ESR}_\gamma^{q.c.}(u)$ be the equivalent safe rate for the model with quadratic costs only as in [19]. It is clear that for any admissible strategy u, $\mathrm{ESR}_\gamma^{q.c.}(u) \geq \mathrm{ESR}_\gamma(u)$. In [19] it is proved that the maximal equivalent safe rate in the model with only quadratic cost is of the form $\frac{\mu^2}{2\gamma\sigma^2} - \varepsilon^{2/3}l^{q.c.} + o(\varepsilon^{2/3})$ for some positive constant $l^{q.c.}$. This implies that $\liminf_{\varepsilon\to 0} \frac{c(\varepsilon)}{\varepsilon^{2/3}} \geq l^{q.c} > 0$.

We will now prove that $\limsup_{\varepsilon\to 0} \frac{c(\varepsilon)}{\varepsilon^{2/3}} < +\infty$. To this end, we assume that there exists $\varepsilon_n \downarrow 0$ such that $\lim_{n\to\infty} \frac{c(\varepsilon_n)}{\varepsilon_n^{2/3}} = +\infty$ and and show that this leads to a contradiction.

Let $q(\cdot)$ be the solution to equation (3.1) as in Theorem 3.1. Define $y_l^n := y_* - \sqrt{\frac{2}{\gamma\sigma^2}(1+\delta)c(\varepsilon_n)}$ for some fixed (i.e., independent of ε) and positive δ and let $\tilde{y}^n$ be the only point such that $q(\tilde{y}^n) = 0$. Assume that, up to a subsequence, $\tilde{y}^n \geq y_*$. (If such a subsequence does not exist, the same argument with $y_r^n := y_* + \sqrt{\frac{2}{\gamma\sigma^2}(1+\delta)c(\varepsilon_n)}$ can be used.) We divide the rest of the proof into two parts:

(i) First we prove that the assumption $\lim_{n\to\infty} \frac{c(\varepsilon_n)}{\varepsilon_n^{2/3}} = +\infty$ implies that there exists a subsequence n_k such that

$$\lim_{k\to\infty} \frac{(q(y_l^{n_k}) - \varepsilon)^2}{\varepsilon_{n_k}^{4/3} c(\varepsilon_{n_k})} = +\infty. \tag{7.15}$$

Recall that $q(0^+) = \varepsilon + 2\sqrt{K\beta}\varepsilon^{2/3} =: b_0(\varepsilon)$ and that $q(\cdot)$ is decreasing and therefore $q(y) \leq b_0$ on $[0,1]$. If there is $C_1 > 1 + 2\sqrt{K\frac{\mu^2}{2\gamma\sigma^2}}$ independent of ε such that eventually $q(y_l^n) \geq C_1 c(\varepsilon_n)$, then to conclude it is enough to note that for large n eventually $b_0(\varepsilon_n) < C_1\varepsilon_n^{2/3} < C_1 c(\varepsilon_n) \leq q(y_l^n)$, where the second inequality follows from the assumption that $\lim_{n\to\infty} \frac{c(\varepsilon_n)}{\varepsilon_n^{2/3}} = +\infty$. This contradicts the monotonicity of $q(\cdot)$, thus we can assume (up to a subsequence) that $q(y_l^n) < C_1 c(\varepsilon_n)$. With this assumption, and also using the monotonicity of $q(y)$ established in Lemma 7.8, on the interval $(y_* - \sqrt{\frac{2}{\gamma\sigma^2}(1-\delta)c(\varepsilon)}, y_*)$ we obtain from Equation (3.1) that, for some constant C_2,

$$\frac{\sigma^2}{2}y^2(1-y)^2 q'(y) < (1-\delta)c(\varepsilon) - c(\varepsilon) + 0 + C_2 c(\varepsilon)^2 + 0.$$

Thus, there is $C_3 > 0$ such that $q'(y) < -C_3 c(\varepsilon)$ on $(y_* - \sqrt{\frac{2}{\gamma\sigma^2}(1-\delta)c(\varepsilon)}, y_*)$ for sufficiently small ε. As a consequence, it follows from the definition of of y_l^n and the assumption $\tilde{y}^n \geq y_*$ that

$$q(y_l^n) = -\int_{y_l^n}^{\tilde{y}^n} q'(y)dy \geq -\int_{y_* - \sqrt{\frac{2}{\gamma\sigma^2}(1-\delta)c(\varepsilon)}}^{y_*} q'(y)dy$$

$$> \int_{y_* - \sqrt{\frac{2}{\gamma\sigma^2}(1-\delta)c(\varepsilon)}}^{y_*} C_3 c(\varepsilon)dy = C_3\sqrt{\frac{2}{\gamma\sigma^2}(1-\delta)}c(\varepsilon)^{3/2}.$$

Recall that y_- is by definition the only point such that $q(y_-) = \frac{\varepsilon}{1+\varepsilon y_-}$. Since $\frac{c(\varepsilon_n)^{3/2}}{\varepsilon_n} \uparrow \infty$ and therefore $q(y_l^n) > \frac{\varepsilon}{1+\varepsilon y_l^n}$, this implies that eventually $y_l^n < y_-$. Furthermore we get that, for some C_4 and C_5,

$$\frac{(q(y_l^n) - \varepsilon_n)^2}{\varepsilon_n^{4/3}c(\varepsilon_n)} \geq \frac{(C_4 c(\varepsilon_n)^{3/2} - \varepsilon_n)^2}{\varepsilon_n^{4/3}c(\varepsilon_n)} \geq \frac{C_5 c(\varepsilon_n)^3}{\varepsilon_n^{4/3}c(\varepsilon_n)} = C_5\left(\frac{c(\varepsilon_n)}{\varepsilon_n^{2/3}}\right)^2 \uparrow \infty,$$

which proves point (i).

(ii) We next show that the assumption that $\lim_{n\to\infty}\frac{c(\varepsilon_n)}{\varepsilon_n^{2/3}}\uparrow\infty$, together with (7.15) implies that, for some $\eta\in(0,y_l^n)$ and sufficiently small ε_n,

$$q(\eta) > q(0^+) = b_0(\varepsilon_n).$$

This contradicts the monotonicity of $q(y)$, thereby producing the desired contradiction.

Fix a large constant $M>0$. Since $\lim_{n\to\infty}\frac{(q(y_l^n)-\varepsilon_n)^2}{\varepsilon_n^{4/3}c(\varepsilon_n)}=+\infty$, for sufficientely small ε_n we have

$$\frac{(q(y_l^n)-\varepsilon_n)^2}{\varepsilon_n^{4/3}} > M\delta c(\varepsilon_n) = M\left(\frac{\gamma\sigma^2}{2}(y_l^n-y_*)^2 - c(\varepsilon_n)\right).$$

Define $y_{1,n} := \inf\{y\in(0,y_*): -\frac{\sigma^2}{2}y^2(1-y)^2(1-\gamma)q^2(y) > \frac{\gamma\sigma^2}{2}(y-y_*)^2 - c(\varepsilon_n)\} > 0$, with the usual convention $\inf\emptyset = +\infty$. Assume that eventually $y_{1,n} < y_l^n$. Define $p_n := -\frac{\sigma^2}{2}y_{1,n}^2(1-y_{1,n})^2(1-\gamma)$; we want to show that also in $y_{1,n}$ this inequality holds:

$$\frac{(q(y_{1,n})-\varepsilon_n)^2}{\varepsilon_n^{4/3}} > M\left(\frac{\gamma\sigma^2}{2}(y_{1,n}-y_*)^2 - c(\varepsilon_n)\right) = p_n M q^2(y_{1,n}),$$

where we used the definition of $y_{1,n}$ in the last equality. This inequality is trivially satisfied if $p_n < 0$ (i.e., when $\gamma < 1$), so assume p_n is positive. Recall from part (i) of the proof that $q(y_{1,n}) > q(y_l^n) \geq C_4 c(\varepsilon_n)^{3/2}$ and therefore that eventually $q(y_{1,n}) > \varepsilon$. Taking square roots (using $q(y_{1,n}) > \varepsilon > 0$) and rearranging the terms in the inequality to be proved, we obtain the equivalent inequality $q(y_{1,n}) > \frac{\varepsilon_n}{1-\sqrt{p_n M}\varepsilon_n^{2/3}}$. Since $q(y_{1,n}) > C_4 c(\varepsilon_n)^{3/2}$, it is sufficient to show that $C_4 c(\varepsilon_n)^{3/2} \geq \frac{\varepsilon_n}{1-\sqrt{p_n M}\varepsilon^{2/3}}$. This can be rewritten as $C_4\left(\frac{c(\varepsilon_n)}{\varepsilon_n^{2/3}}\right)^{3/2} \geq \frac{1}{1-\sqrt{p_n M}\varepsilon^{2/3}}$, which holds true for small ε_n since $\frac{c(\varepsilon_n)}{\varepsilon_n^{3/2}}\uparrow\infty$ and p_n is uniformly bounded in ε_n.

If eventually $y_{1,n} < y_l^n$, let $y_{2,n}$ be equal to $y_{1,n}$. Otherwise, up to a subsequence, $y_l^n < y_{1,n}$ and then define $y_{2,n} := y_l^n$. We have just shown that in both cases

$$\frac{(q(y_{2,n})-\varepsilon_n)^2}{\varepsilon_n^{4/3}} > M\left(\frac{\gamma\sigma^2}{2}(y_{2,n}-y_*)^2 - c(\varepsilon_n)\right)$$

and $y_{2,n} \leq y_{1,n}$.

Choose η such that $\eta < y_{2,n}$ for every n (this is possible because for $y\downarrow 0$, $-\frac{\sigma^2}{2}y^2(1-y)^2(1-\gamma)q^2(y)\downarrow 0$ and $q(y)$ is uniformly bounded in ε and $\frac{\gamma\sigma^2}{2}(y-y_*)^2 - c(\varepsilon_n)\uparrow\beta > 0$). On the interval $(\eta, y_{2,n})$ we have

$$\frac{\sigma^2}{2}\frac{1}{16}q'(y) \leq \frac{\sigma^2}{2}y^2(1-y)^2q'(y)$$
$$< \frac{\gamma\sigma^2}{2}(y-y_*)^2 - c(\varepsilon_n) + 0 - \frac{\sigma^2}{2}y^2(1-y)^2(1-\gamma)q^2(y) - \frac{(q(y)-\varepsilon_n)^2}{4K\varepsilon_n^{4/3}}$$
$$\leq 2\left[\frac{\gamma\sigma^2}{2}(y-y_*)^2 - c(\epsilon_n)\right] - \frac{(q(y)-\varepsilon_n)^2}{4K\varepsilon_n^{4/3}}, \quad (7.16)$$

25

where the first inequality follows from $y^2(1-y)^2 \le \frac{1}{16}$, the second from $y < y_*$ and $q > 0$, the third from $y_{2,n} \le y_{1,n}$ and the definition of $y_{1,n}$. Define $\tilde{q}(y)$ on $(\eta, y_{2,n})$ as the solution of the Cauchy problem

$$\tilde{q}'(y) = \tilde{f}(y, \tilde{q}(y))$$
$$:= \frac{32}{\sigma^2}\left(2\left[\frac{\gamma\sigma^2}{2}(y-y_*)^2 - c(\varepsilon_n)\right] - \frac{(q(y)-\varepsilon_n)^2}{4K\varepsilon_n^{4/3}}\right),$$
$$\tilde{q}(y_{2,n}) = q(y_{2,n}).$$

The inequality (7.16) implies that $q'(y) < \tilde{f}(y, q(y))$. Thus, $\tilde{q}(y) < q(y)$ on $(\eta, y_{2,n})$. Define on $(\eta, y_{2,n})$ the function $k(y) > \varepsilon_n$ by

$$\frac{(k(y)-\varepsilon_n)^2}{\varepsilon_n^{4/3}} = M\left(\frac{\gamma\sigma^2}{2}(y-y_*)^2 - c(\varepsilon_n)\right).$$

To conclude the proof it remains to show that $k(y)$ is a subsolution for $\tilde{q}(y)$ on $(\eta, y_{2,n})$, which implies that $k(\eta) < q(\eta)$. In particular, if M is chosen sufficiently large, we get the contradiction.

Thus, it remains to show that $\tilde{f}(y, k(y)) < k'(y)$ on $(\eta, y_{2,n})$, i.e.,

$$\frac{32}{\sigma^2}\left(2 - \frac{M}{4K}\right)\left[\frac{\gamma\sigma^2}{2}(y-y_*)^2 - c(\varepsilon_n)\right] < \frac{\sqrt{M}}{2}\varepsilon_n^{2/3}\left[\frac{\gamma\sigma^2}{2}(y-y_*)^2 - c(\varepsilon_n)\right]^{-1/2}\gamma\sigma^2(y-y_*).$$

This is equivalent to

$$\frac{64}{\gamma\sigma^4}\left(\frac{\sqrt{M}}{4K} - \frac{2}{\sqrt{M}}\right) > \varepsilon_n^{2/3}\left[\frac{\gamma\sigma^2}{2}(y-y_*)^2 - c(\varepsilon_n)\right]^{-3/2}(y_*-y) =: r(y). \qquad (7.17)$$

A simple calculation shows that $r'(y) > 0$ on $(0, y_l^n)$. Thus, $r(y) \le r(y_l^n)$ on $(\eta, y_{2,n})$, where

$$r(y_l^n) = \varepsilon_n^{2/3}(\delta c(\varepsilon_n))^{-3/2}\left(\frac{2}{\gamma\sigma^2}(1+\delta)c(\varepsilon_n)\right)^{1/2} = D_1\frac{\varepsilon_n^{2/3}}{c(\varepsilon)} \downarrow 0,$$

with $D_1 > 0$. If M is large enough, the left-hand side in (7.17) is strictly positive. Since $r(y)$ converges to 0 on $(\eta, y_{2,n})$ uniformly in ε, for sufficiently small ε inequality (7.17) holds true. This completes the proof. □

Next, we establish existence and uniqueness for the inhomogeneous Riccati equation that determines the small-cost asymptotics in Section 6. To this end, we first prove an auxiliary result about Riccati ODEs:

Lemma 7.9. *The Riccati equation*

$$y'(x) = f(x, y(x)) := -ax^2 + b + cy^2(x), \qquad (7.18)$$

with $a, b, c > 0$ has a unique solution such that

$$\lim_{x\to\infty}\frac{y(x)}{\sqrt{a/c}\,x} = 1. \qquad (7.19)$$

Furthermore, in equation (7.18) replace the parameter b with respectively b_1 and b_2, and consider the corresponding unique solutions $y_1(x)$ and $y_2(x)$ that satisfy (7.19). If $b_1 < b_2$, then $y_1(x) > y_2(x)$.

Proof. On $(\sqrt{b/a}, +\infty)$, define the function $h(x) := \sqrt{ax^2/c - b/c}$. Notice that by definition of $h(x)$ we have $f(x, h(x)) = 0$. For each $\bar{x} \in (\sqrt{b/a}, +\infty)$ consider the solution $y(x; \bar{x}, h(\bar{x}))$ with initial condition $(\bar{x}, h(\bar{x}))$ and define $y_*(x) := \sup\{y(x; \bar{x}, h(\bar{x})) : \bar{x} \in (\sqrt{b/a}, +\infty)\}$.

For any x_1 there is a large y_1 such that the linear function $\tilde{y}(x) = y_1 + \sqrt{a/c}(x - x_1)$ is a subsolution to (7.18) whose graph does not intersect the graph of $h(x)$. In particular, the solution $y(x; x_1, y_1)$ to (7.18) with initial condition (x_1, y_1) is (on its definition interval) strictly larger than $h(x)$. Since for any $\bar{x}$ in its definition interval $y(\bar{x}; x_1, y_1) > h(\bar{x}) = y(\bar{x}; \bar{x}, h(\bar{x}))$, we have also that $y_1 = y(x_1; x_1, y_1) > y(x_1; \bar{x}, h(\bar{x}))$ and thus $+\infty > y_1 \geq y_*(x_1)$. This argument can be repeated for any $x_1 \in (\sqrt{b/a}, +\infty)$, hence $y_*(x) < +\infty$ on $(\sqrt{b/a}, +\infty)$.

We want to prove that $y_*(x)$ is the unique solution that satisfies (7.19). By construction, $y_*(x)$ has the following properties:

i) $y_*(x) \geq h(x)$;

ii) $\left(\sqrt{\frac{b}{a}}, +\infty\right) \subset D$, where D is the domain of $y_*(x)$.

From property (i) it follows that $\liminf_{x\to\infty} \frac{y_*(x)}{\sqrt{a/cx}} \geq 1$.

Next we show that $L := \lim_{x\to\infty} \frac{y_*(x)}{\sqrt{a/cx}}$ exists. Assume $\limsup_{x\to\infty} \frac{y_*(x)}{\sqrt{a/cx}} =: M \in (1, +\infty)$ (the case $M = +\infty$ is analogous). Then, there is a sequence $(x_n)_{n\geq 0}$ such that $\lim_{n\to\infty} \frac{y_*(x_n)}{\sqrt{a/cx_n}} = M$. In particular, for any $\delta \in (0, M-1)$ there exists $N_\delta \in \mathbb{N}$ such that $\forall n \geq N_\delta$ we have $y_*(x_n) \geq (M-\delta)\sqrt{a/cx_n}$. For large x, the function $s(x) = (M-\delta)\sqrt{a/cx}$ is a subsolution to (7.18), because

$$(M-\delta)\sqrt{\frac{a}{c}} = s'(x) \leq -ax^2 + b + cs^2(x) = ax^2((M-\delta)^2 - 1) + b.$$

Thus, for every $\delta \in (0, M-1)$ and some $\bar{x}$, we have $y_*(x) \geq (M-\delta)\sqrt{a/cx}$ for $x \geq \bar{x}$. In particular, $\liminf_{x\to\infty} \frac{y_*(x)}{\sqrt{a/cx}} \geq M - \delta$ for any small δ, and $\liminf_{x\to\infty} \frac{y_*(x)}{\sqrt{a/cx}} = M = \limsup_{x\to\infty} \frac{y_*(x)}{\sqrt{a/cx}}$. In other terms, the limit L exists.

We prove next that $L = 1$. First, assume by contradiction that $1 < L < +\infty$. Since $\lim_{x\to\infty} \frac{y_*(x)}{\sqrt{a/cx}} = L < +\infty$, the function $y_*(x)$ grows linearly. On the other hand from (7.18) one gets

$$\lim_{x\to\infty} \frac{y'_*(x)}{ax^2} = L - 1 > 0,$$

which implies that $y'_*(x)$ grows quadratically, leading to a contradiction.

Assume now that $L = +\infty$. From (7.18) it follows that

$$\lim_{x\to\infty} \frac{y'_*(x)}{cy_*^2(x)} = 1.$$

For small δ and sufficiently large x, we have $(1-\delta)cy_*^2(x) \leq y'_*(x)$. This implies that $y_*(x)$ is bounded from below by a positive function of the form $\frac{1}{k-(1-\delta)cx}$ for some $k > 0$. In particular, $y_*(x)$ would have a vertical asymptote, contradicting property (ii). This proves that $L = 1$.

The next step is to prove uniqueness. Consider a sufficiently small $\delta > 0$ and $\bar{x}$ such that for any $x \geq \bar{x}$:

$$\frac{1}{\sqrt{acx^2}} \leq \delta \qquad \text{and} \qquad \frac{y_*(x)}{\sqrt{a/cx}} \geq 1 - \delta.$$

For any $d > 0$, consider the function $w(x; d) = y_*(x) + dx$. We now show that for $x \geq \bar{x}$, $w(x; d)$ is a subsolution to (7.18), i.e., $w'(x; d) \leq f(x, w(x; d))$. Since $y_*(x)$ is a solution to (7.18), this inequality is equivalent to $d \leq cd^2x^2 + 2cdy_*(x)x$, which can be rearranged to

$$\frac{1}{\sqrt{ac}x^2} \leq \sqrt{\frac{c}{a}}d + 2\sqrt{\frac{c}{a}}\frac{y_*(x)}{x}.$$

Since $x \geq \bar{x}$, this inequality follows from $\delta \leq \sqrt{\frac{c}{a}}d + 2 - 2\delta$, given that δ was chosen appropriately. Thus, $w(x; d)$ is a subsolution for any $d > 0$. In particular, let $y_2(x) > y_*(x)$ be a solution to (7.18) and choose d_* such that $y_2(\bar{x}) = y_*(\bar{x}) + d_*\bar{x}$. Then $y_2(x) \geq w(x; d_*) = y_*(x) + d_*x$ for $x \geq \bar{x}$ and $y_2(x)$ cannot satisfy (7.19). Since any solution smaller than $y_*(x)$ is also -for large x- smaller than $h(x)$ and thus eventually decreasing, this is enough to prove uniqueness.

Finally, define as before $h_1(x) = \sqrt{ax^2/c - b_1/c}$ and $h_2(x) = \sqrt{ax^2/c - b_2/c}$. Since $h_1(x) > h_2(x)$ and since any solution to equation (7.18) with coefficient b_1 is a subsolution for equation (7.18) with coefficient b_2, the solution $y_1(x, \bar{x})$ to the first equation with initial condition $(\bar{x}, h_1(\bar{x}))$ is above the solution $y_2(x, \bar{x})$ to the second equation with initial condition $(\bar{x}, h_2(\bar{x}))$ on $(-\infty, \bar{x}]$. It follows that $y_1(x) > y_2(x)$. $\qquad\square$

Proof of Proposition 6.2. For any $l > 0$, define $r_B(z; l)$ as the unique solution of (6.6) that satisfies

$$\lim_{z \to -\infty} \frac{r_B(z)}{-\sqrt{2K\gamma\sigma^2 z}} = 1. \tag{7.20}$$

Here, existence and uniqueness of this solution follow from Lemma 7.9. Let $r(z; l)$ be the unique solution to (6.7) with initial condition $r(0; l) = 0$. It is enough to prove that there exists a unique l such that for some $z_- < 0$ we get $r_B(z_-; l) = r(z_-; l) = 1$, $r_B(z_-; l) > 1$ on $(-\infty, z_-)$ and $r(z_-; l) < 1$ on $(z_-, 0]$.

Define $l_\lambda := \sqrt{\frac{\gamma K}{2}\sigma^3 y_*^2(1 - y_*)^2}$ and $l_\varepsilon := \left(\frac{3}{4}\sqrt{\frac{\gamma}{2}}\sigma^3 y_*^2(1 - y_*)^2\right)^{2/3}$. If the following values exist, define $z_B(l)$ the (only) value such that $r_B(z_B(l); l) = 1$ and $r'_B(z_B(l); l) < 0$, and $z_{NT}(l)$ the (only) value such that $r(z_{NT}(l); l) = 1$ and $r'(z_{NT}(l); l) < 0$.
If $l = l_\lambda$, the solution to ODE (6.6) with boundary condition (7.20) is $r_B(z; l_\lambda) = -\sqrt{2\gamma K}\sigma z + 1$. Thus, $z_B(l_\lambda) = 0$. From Lemma 7.9 it follows that if $l_1 < l_2$, then $r_B(z; l_1) > r_B(z; l_2)$ and thus $z_B(l_1) > z_B(l_2)$. Hence, the function $l \to z_B(l)$ is well-defined and decreasing on $[l_\lambda, +\infty)$. Furthermore, $\lim_{l \to \infty} z_B(l) = -\infty$.

Since $\max_{z \in (-\infty, 0)} r(z; l) < 1$ if and only if $l < l_\varepsilon$, the function $l \to z_{NT}(l)$ is well-defined only on $[l_\varepsilon, +\infty)$. Furthermore, it is increasing and $\lim_{l \to \infty} z_{NT}(l) = 0$.

Let $l_M := \max\{l_\lambda, l_\varepsilon\}$. If $l_\lambda \geq l_\varepsilon$, then $z_{NT}(l_\lambda) < z_B(l_\lambda) = 0$.

A brief calculation shows that $z_{NT}(l_\varepsilon) = -\sqrt{\frac{2l_\varepsilon}{\gamma\sigma^2}}$. Thus, for $(z, r) \in (-\infty, z_{NT}(l_\varepsilon)) \times \{1\}$ any solution to (6.6) has strictly positive derivative, while it is strictly negative for $(z, r) \in (z_{NT}(l_\varepsilon), 0] \times \{1\}$. In particular, since $r'_B(z_B(l); l) < 0$, if $l_\varepsilon \geq l_\lambda$, then $z_B(l_\varepsilon) \in (z_{NT}(l_\varepsilon), 0]$. In both cases, $z_{NT}(l_M) < z_B(l_M)$. Given the monotonicity of the functions $z_{NT}(l)$ and $z_B(l)$, there exists a unique $l_* \in (l_M, +\infty)$ such that $z_{NT}(l_*) = z_B(l_*)$.

Without price impact the no-trade region is given by $(z_{NT}(l_\varepsilon), -z_{NT}(l_\varepsilon))$ (see [14, Formula (2.9)]). Since $z_{NT}(l)$ is an increasing function, the no-trade region with price impact is strictly smaller: $(z_{NT}(l_*), -z_{NT}(l_*)) \subset (z_{NT}(l_\varepsilon), -z_{NT}(l_\varepsilon))$. $\qquad\square$

Finally, we rigorously prove that the value function $q^\varepsilon(y)$ converges to the solution $r(z)$ specified in Proposition 6.2.

Proof of Proposition 6.3. Rewrite equation (3.1) as $q'(y) = f^\varepsilon(y, q(y))$ and equation (6.6-6.8) as $r'(z) = g(z, r(z))$. Consider a sequence $(\varepsilon_n)_{n \geq 0}$ such that $\frac{c(\varepsilon_n)}{\varepsilon_n^{2/3}}$ has limit equal to l (for the rest of the proof we will write ε instead of ε_n for simplicity), such a limit exists by Proposition 6.1. In the proof of Theorem 3.1 we show that there exists a function $h^\varepsilon(y)$ defined close to 0 such that $f^\varepsilon(y, h^\varepsilon(y)) = 0$, $h^\varepsilon(0^+) = b_0(\varepsilon)$ and $(h^\varepsilon)'(0^+) < 0$. In the proof of Lemma 7.8 we prove that h^ε is decreasing on its whole definition interval $(0, y_* - M\varepsilon^{1/3})$.

In view of the continuity guaranteed by the implicit function theorem, for $\varepsilon \downarrow 0$, the function $\varepsilon^{-1} h^\varepsilon(\varepsilon^{1/3} z + y_*)$ converges to $k(z) := 1 + 2\sqrt{K\left(\frac{\gamma\sigma^2}{2} z^2 - l\right)}$, i.e., $k(z)$ such that $g(z, k(z)) = 0$.

In the proof of Theorem 3.1 we show that there exists a unique solution q_0^ε to equation (3.1) with limit $q^\varepsilon(0^+) = b_0(\varepsilon)$. Define $q^\varepsilon(y; \bar{y}, \bar{q})$ as the solution to equation (3.1) with initial condition $(\bar{y}, \bar{q})$ and $\tilde{q}_0^\varepsilon(y) := \sup\{q^\varepsilon(y; \bar{y}, h(\bar{y})) : \bar{y} \in (0, y_* - M\varepsilon^{1/3})\}$. Now we will prove that $q_0^\varepsilon(y) = \tilde{q}_0^\varepsilon(y)$.

First notice that since $(h^\varepsilon)'(y) < 0 = f^\varepsilon(y, h^\varepsilon(y))$, $h^\varepsilon(y)$ is a subsolution of Equation (3.1). Therefore from $(q_0^\varepsilon)'(0^+) = 0$ and $(h^\varepsilon)'(0^+) < 0$, we get that $q_0^\varepsilon(y) > h^\varepsilon(y)$ on $(0, y_* - M\varepsilon^{1/3})$. In addition, for any $\bar{y}$, $q(y; \bar{y}, h^\varepsilon(\bar{y})) < h^\varepsilon(y)$ on $(0, \bar{y})$. This shows that $q_0^\varepsilon(y) \geq \tilde{q}_0^\varepsilon(y)$. Let's assume by contradiction that $\tilde{q}_0^\varepsilon(0^+) < q_0^\varepsilon(0^+) = h^\varepsilon(0^+)$. This means that $\tilde{q}_0^\varepsilon(\bar{y}) < h^\varepsilon(\bar{y})$ for some $\bar{y} > 0$, that is $\tilde{q}_0^\varepsilon(\bar{y}) < q_0^\varepsilon(\bar{y}; \bar{y}, h^\varepsilon(\bar{y}))$, contradicting the maximality of $\tilde{q}_0^\varepsilon(y)$. This proves that $q_0^\varepsilon(y) = \sup\{q^\varepsilon(y; \bar{y}, h(\bar{y})) : \bar{y} \in (0, y_* - M\varepsilon^{1/3})\}$.

Now, let $r(z; \bar{z}, \bar{r})$ be the solution to equation (6.6-6.8) with initial condition $(\bar{z}, \bar{r})$. By continuity of the solutions to (3.1) with respect to parameters, $\varepsilon^{-1} q_0^\varepsilon(\varepsilon^{1/3} z + y_*)$ converges to $r_l(z) := \sup\{r(z; \bar{z}, k(\bar{z})) : \bar{z} \in (-\infty, -M)\}$.

In the proof of Lemma 7.9 we show that $r_l(z)$ as defined above satisfies

$$\lim_{z \to -\infty} \frac{r_l(z)}{-\sqrt{2K\gamma\sigma^2 z}} = 1.$$

Recall that q_1^ε is the unique solution to equation (3.1) with limit $q^\varepsilon(1^-) = b_1(\varepsilon)$. With the same arguments used for q_0^ε, we get that $\varepsilon^{-1} q_1^\varepsilon(\varepsilon^{1/3} z + y_*)$ converges to the unique solution $r_r(z)$ to equation (6.8) that satisfies

$$\lim_{z \to +\infty} \frac{r_r(z)}{-\sqrt{2K\gamma\sigma^2 z}} = 1.$$

In the proof of Theorem 3.1 we show that, for the optimal value of β, $q^\varepsilon(y) = q_0^\varepsilon(y) = q_1^\varepsilon(y)$. Thus, the function $\varepsilon^{-1} q^\varepsilon(\varepsilon^{1/3} z + y_*)$ converges to the unique solution of (6.6-6.8) with growth conditions (6.10). In particular, given the uniqueness of l_* in Proposition 6.2, $\liminf_{\varepsilon \to 0} \frac{c(\varepsilon)}{\varepsilon^{2/3}} = \limsup_{\varepsilon \to 0} \frac{c(\varepsilon)}{\varepsilon^{2/3}} = l = l_*$ as claimed. $\square$

References

[1] R. F. Almgren and N. Chriss. Optimal execution of portfolio transactions. *J. Risk*, 3:5–40, 2001.

[2] R. F. Almgren and T. M. Li. Option hedging with smooth market impact. *Market Microstructure Liq.*, 2(1):1650002, 2016.

[3] A. Altarovici, J. Muhle-Karbe, and H. M. Soner. Asymptotics for fixed transaction costs. *Finance Stoch.*, 19(2):702–723, 2015.

[4] D. Bertsimas and A. W. Lo. Optimal control of execution costs. *J. Financ. Markets*, 1(1):1–50, 1998.

[5] M. Bichuch. Asymptotic analysis for optimal investment in finite time with transaction costs. *SIAM J. Financ. Math.*, 3(1):433–458, 2012.

[6] X. Chen and M. Dai. Asymptotics for Merton problem with small capital gain tax and interest rate. Preprint, 2013.

[7] P. Collin-Dufresne, K. Daniel, C. Moallemi, and M. Saglam. Strategic asset allocation with predictable returns and transaction costs. Preprint, 2013.

[8] G. Constantinides. Capital market equilibrium with transaction costs. *J. Polit. Econ.*, 94(4):842–862, 1986.

[9] M. Dai and F. Yi. Finite-horizon optimal investment with transaction costs: a parabolic double obstacle problem. *J. Diff. Eq.*, 246(4):1445–1469, 2009.

[10] M. H. A. Davis and A. R. Norman. Portfolio selection with transaction costs. *Math. Oper. Res.*, 15(4):676–713, 1990.

[11] B. Dumas and E. Luciano. An exact solution to a dynamic portfolio choice problem under transaction costs. *J. Finance*, 46(2):577–595, 1991.

[12] N. Garleanu and L. H. Pedersen. Dynamic trading with predictable returns and transaction costs. *J. Finance*, 68(6):2309–2340, 2013.

[13] N. Garleanu and L. H. Pedersen. Dynamic portfolio choice with frictions. *J. Econ. Theory*, 164:487–516, 2016.

[14] S. Gerhold, P. Guasoni, J. Muhle-Karbe, and W. Schachermayer. Transaction costs, trading volume, and the liquidity premium. *Finance Stoch.*, 18(1):1–37, 2014.

[15] S. Grossman and J. Vila. Optimal dynamic trading with leverage constraints. *J. Financ. Quant. Anal.*, 27(2):151–168, 1992.

[16] S. J. Grossman and Z. Zhou. Optimal investment strategies for controlling drawdowns. *Math. Finance*, 3(3):241–276, 1993.

[17] P. Guasoni and S. Robertson. Portfolios and risk premia for the long run. *Ann. Appl. Probab.*, 22(1):239–284, 2012.

[18] P. Guasoni and M. H. Weber. Nonlinear price impact and portfolio choice. *Preprint*, 2015.

[19] P. Guasoni and M. H. Weber. Dynamic trading volume. *Math. Finance*, 27(2):313–349, 2017.

[20] K. Janeček and S. E. Shreve. Asymptotic analysis for optimal investment and consumption with transaction costs. *Finance Stoch.*, 8(2):181–206, 2004.

[21] J. Kallsen and S. Li. Portfolio optimization under small transaction costs: a convex duality approach. Preprint, 2013.

[22] J. Kallsen and J. Muhle-Karbe. The general structure of optimal investment and consumption with small transaction costs. *Math. Finance*, 27(3):659–703, 2017.

[23] E. Kamke. *Differentialgleichungen. Lösungsmethoden und Lösungen. I: Gewöhnliche Differentialgleichungen.* B. G. Teubner, Stuttgart, ninth edition, 1977.

[24] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus.* Springer, New York, second edition, 1991.

[25] R. Korn. Portfolio optimisation with strictly positive transaction costs and impulse control. *Finance Stoch.*, 2(2):85–114, 1998.

[26] A. Madhavan. Market microstructure: a survey. *Journal Financ. Markets*, 3(3):205–258, 2000.

[27] M. J. P. Magill and G. M. Constantinides. Portfolio selection with transactions costs. *J. Econ. Theory*, 13(2):245–263, 1976.

[28] R. Martin. Optimal trading under proportional transaction costs. *RISK*, August:54–59, 2014.

[29] R. C. Merton. Lifetime portfolio selection under uncertainty: the continuous-time case. *Rev. Econ. Statist.*, 51(3):247–257, 1969.

[30] L. Moreau, J. Muhle-Karbe, and H. M. Soner. Trading with small price impact. *Math. Finance*, 27(2):350–400, 2017.

[31] A. A. Obizhaeva and J. Wang. Optimal trading strategy and supply/demand dynamics. *J. Financ. Markets*, 16(1):1–32, 2013.

[32] D. Possamai, H. M. Soner, and N. Touzi. Homogenization and asymptotics for small transaction costs: the multidimensional case. *Comm. Part. Diff. Eq.*, 40(11):2005–2046, 2015.

[33] A. Schied and T. Schöneborn. Risk aversion and the dynamics of optimal liquidation strategies in illiquid markets. *Finance Stoch.*, 13(2):181–204, 2009.

[34] S. E. Shreve and H. M. Soner. Optimal investment and consumption with transaction costs. *Ann. Appl. Probab.*, 4(3):609–692, 1994.

[35] L. J. Slater. *Confluent hypergeometric functions.* Cambridge University Press, New York, 1960.

[36] H. M. Soner and N. Touzi. Homogenization and asymptotics for small transaction costs. *SIAM J. Control Optim.*, 51(4):2893–2921, 2013.

[37] E. Whittaker and G. N. Watson. *A course of modern analysis.* Cambridge University Press, 1996.