The Limits of Leverage

Paolo Guasoni† Eberhard Mayerhofer‡

October 6, 2014

Abstract

When trading incurs proportional costs, leverage can scale an asset’s return only up to a multiple, which is sensitive to its volatility and liquidity. In a model with one safe and one risky asset, with constant investment opportunities and proportional costs, we find strategies that maximize long term return given average volatility. As leverage increases, rising rebalancing costs imply declining Sharpe ratios. Beyond a critical level, even returns decline. Holding the Sharpe ratio constant, higher volatility leads to superior returns through lower costs. For funds replicating benchmark multiples, such as leveraged ETFs, our strategies optimally trade off alpha against tracking error.

JEL Classification: G11, G12.

Keywords: leverage, transaction costs, portfolio choice, performance evaluation, ETFs.

---

†Boston University, Department of Mathematics and Statistics, 111 Cummington Street, Boston, MA 02215, USA, and Dublin City University, School of Mathematical Sciences, Glasnevin, Dublin 9, Ireland, email guasoni@bu.edu. Partially supported by the ERC (278295), NSF (DMS-1412529), SFI (07/MI/008, 07/SK/M1189, 08/SRC/FMC1389), and FP7 (RG-248896).

‡Dublin City University, School of Mathematical Sciences, Glasnevin, Dublin 9, Ireland, email eberhard.mayerhofer@dcu.ie.

*For helpful comments, we thank Jakša Cvitanić, Gur Huberman, Tim Leung, Johannes Muhle-Karbe, Ronnie Sircar, Peter Tankov, Walter Schachermayer, and seminar participants at ETH Zürich, Vienna Graduate School of Finance, Quant Europe, Quant USA, and Banff Workshop on Arbitrage and Portfolio Optimization.
1 Introduction

If trading is costless, leverage can scale returns without limits. Using the words of Sharpe (2011),

“If an investor can borrow or lend as desired, any portfolio can be levered up or down. A combination with a proportion $k$ invested in a risky portfolio and $1 - k$ in the riskless asset will have an expected excess return of $k$ and a standard deviation equal to $k$ times the standard deviation of the risky portfolio. Importantly, the Sharpe Ratio of the combination will be the same as that of the risky portfolio.”

In theory, this insight implies that the efficient frontier is linear, that efficient portfolios are identified by their common maximum Sharpe ratio, and that any of them spans all the other ones. Also, if leverage can deliver any expected returns, then risk-neutral portfolio choice is meaningless, as it leads to infinite leverage.

In practice, hedge funds and high-frequency trading firms employ leverage to obtain high returns from small relative mispricing of assets. Recent financial products such as leveraged mutual funds and exchange traded funds (ETFs) closely follow the strategy described by Sharpe, rebalancing their exposure to an underlying asset, with the aim of replicating a multiple of its daily return.

This paper shows that trading costs undermine these classical properties of leverage and set sharp theoretical limits to its applications. We start by characterizing the set of portfolios that maximize long term expected returns for given average volatility, extending the familiar efficient frontier to a market with one safe and one risky asset, where both investment opportunities and relative bid-ask spreads are constant. Figure 1 plots this frontier: expectedly, trading costs decrease returns, with the exception of a full safe investment (the axes origin) or a full risky investment (the attachment point with unit coordinates), which lead to static portfolios without trading, and hence earn their frictionless return.

But trading costs do not merely reduce expected returns below their frictionless benchmarks. Unexpectedly, in the leverage regime (the right of the full-investment point) rebalancing costs rise so quickly with volatility that returns cannot increase beyond a critical factor, the leverage multiplier or, briefly, the multiplier. The multiplier depends on the relative bid-ask spread $\varepsilon$, on the expected return $\mu$ and volatility $\sigma$, and it is approximately equal to

$$0.3815 \left(\frac{\mu}{\sigma^2}\right)^{1/2} \varepsilon^{-1/2}. \quad (1.1)$$

Table 1 shows that even a modest bid-ask spread of 0.10% implies a multiplier of 23 for an asset with 10% volatility and 5% expected return, while the multiplier declines to 10 for an asset with equal Sharpe ratio, but with a volatility of 50%. Leverage opportunities are much more limited for less liquid assets, with a spread of 1%, from less than 8 for 10% volatility to less than 4 for 50% volatility. Importantly, these limits on leverage hold even allowing for continuous trading, infinite market depth (any quantity trades at the bid or ask price), and zero capital requirements.

Our results have three broad implications. First, with a positive bid-ask spread even a risk-neutral investor who seeks to maximize expected long-run returns will take finite leverage, and in fact a rather low leverage ratio in an illiquid market – risk-neutral portfolio choice is meaningful.

---

1 A famous example is Long Term Capital Management, which used leverage of up to 30 to 40 times to increase returns from convergence trades between on-the-run and off-the-run treasury bonds, see Edwards (1999).

2 As we focus on long term investments, we neglect the one-off costs of set up and liquidation, which are negligible over a long holding period.
Table 1: Leverage multiplier (maximum factor by which a risky asset’s return can be scaled) for different asset volatilities and bid-ask spreads, holding the Sharpe ratio at the constant level of 0.5. Multipliers are obtained from numerical solutions of (3.1), while their approximations from (1.1) are in brackets.

The resulting multiplier sets an endogenous level of risk that the investor chooses not to exceed regardless of risk aversion, simply to avoid reducing returns with trading costs. In this context, margin requirements based on volatility (such as value at risk and its variations) are binding only when they reduce leverage below the multiplier, and are otherwise redundant. In addition, the multiplier shows that an exogenous increase in trading costs, such as a proportional Tobin tax on financial transactions, implicitly reduces the maximum leverage that any investor who seeks return is willing to take, regardless of risk attitudes.

Second, two assets with the same Sharpe ratio do not generate the same efficient frontier with trading costs, and more volatility leads to a superior frontier. For example (Table 1) with a 1% spread the maximum leveraged return on an asset with 10% volatility and 5% return is $7.72 \times 5\% \approx 39\%$. By contrast, an asset with 50% volatility and 25% return (equivalent to the previous one from a classical viewpoint, since it has the same Sharpe ratio 0.5), leads to a maximum leveraged return of $3.66 \times 25\% \approx 92\%$. The reason is that a more volatile asset requires a lower leverage ratio (hence lower rebalancing costs) to reach a certain return. Thus, an asset with higher volatility spans an efficient frontier that achieves higher returns through lower costs.

Third, we obtain theoretical bounds on the potential returns of leveraged ETFs, and derive a testable restriction between the alpha and the tracking error of an optimally replicated ETF. In a frictionless setting, an ETF can perfectly scale returns by any factor, without any tracking error: alpha is zero and the ETF return perfectly correlated with the benchmark’s. In reality, leveraged ETFs have been introduced only since 2006, currently have leverage factors of up to three (minus three for inverse funds), and funds on less liquid assets have significant tracking error.

Under optimal replication with trading costs, we show the following relation between the intercept $\alpha$ and squared correlation $R^2$ in the regression of the ETF return (net of management fees) on the benchmark’s return

$$\alpha \approx -\frac{\sqrt{3}}{12} \sigma^2 \pi_s (1 - \pi_s)^2 \frac{\varepsilon}{\sqrt{1 - R^2}},$$

(1.2)

where $\pi_s$ is the desired scaling factor, $\sigma$ is the benchmark’s volatility, and $\varepsilon$ is its relative spread. The equation makes the optimal replication trade-off clear: a higher $R^2$ (lower tracking error) leads to a more negative alpha through higher costs, and vice versa. More importantly, the equation offers a testable relationship among observable quantities, without involving the expected return $\mu$, notoriously hard to estimate with precision.

This paper bears on the established literature on portfolio choice with frictions and on the nascent literature on leveraged ETFs. The effect of transaction costs on portfolio choice is first studied by Magill and Constantinides (1976), Constantinides (1986), and Davis and Norman (1990),
Figure 1: Efficient Frontier with trading costs, as expected return (vertical axis, in multiples of the asset’s return) against standard deviation (horizontal axis, in multiples of the asset’s volatility). The asset has expected return $\mu = 8\%$, volatility $\sigma = 16\%$, and bid-ask spread of 1%. The upper line denotes the classical efficient frontier, with no transaction costs. The maximum height of the curve corresponds to the leverage multiplier.

who identify a wide no-trade region, and derive the optimal trading boundaries through numerical procedures. While these papers focus on the maximization of expected utility from intertemporal consumption on an infinite horizon, Taksar, Klass and Assaf (1988), and Dumas and Luciano (1991) show that similar strategies are obtained in a model with terminal wealth and a long horizon – time preference has negligible effects on trading policies. This paper adopts the same approach of a long horizon, both for the sake of tractability, and because it focuses on the tradeoff between return, risk, and costs, rather than consumption.

Our asymptotic results for positive risk aversion are similar in spirit to the ones derived by Shreve and Soner (1994), Rogers (2004), Gerhold, Guasoni, Muhle-Karbe and Schachermayer (2014), and Kallsen and Muhle-Karbe (2013), whereby transaction costs imply a no-trade region with width of order $O(\varepsilon^{1/3})$ and a welfare effect of order $O(\varepsilon^{2/3})$. We also find that the trading boundaries obtained from a local mean-variance criterion are equivalent at the first order to the ones obtained from power utility. The risk-neutral expansions and the limits of leverage of order $O(\varepsilon^{-1/2})$ are new, and are qualitatively different from the risk-averse case. These results are not regular perturbations of a frictionless analogue, which is ill-posed. They are rather singular perturbations, which display the speed at which the frictionless problem becomes ill-posed as the crucial friction parameter vanishes.

Our paper also contributes to the literature on leveraged ETFs. Tang and Xu (2013) observe that leveraged funds deviate significantly from their benchmarks even after management fees, and separate tracking error into a compounding component, due to the convexity of levered returns and a rebalancing component, due to trading frictions (cf. Jarrow (2010); Lu et al. (2012); Avellaneda
and Zhang (2010); Cheng and Madhavan (2009)). Jiang and Yan (2012), Avellaneda and Dobi (2012), and Guo and Leung (2014) report that ETFs significantly underperform their benchmarks even at daily frequencies, and Wagalath (2013) derives an asymptotic expression for the slippage that results from rebalancing at fixed intervals. We incorporate trading costs explicitly in the model, and derive optimal replication policies that trade off alpha against tracking error.

Finally, this paper connects to the recent work of Frazzini and Pedersen (2012) on embedded leverage. If different investors face different leverage constraints, they find that in equilibrium assets with higher factor exposures trade at a premium, thereby earning a lower return. Frazzini and Pedersen (2014) confirm this prediction across a range of markets and asset classes, and Asness et al. (2012) use it to explain the performance risk-parity strategies. With exogenous asset prices, we find that assets with higher volatility generate a superior efficient frontier by requiring lower rebalancing costs for the same return. This observation suggests that the embedded leverage premium may be induced by rebalancing costs in addition to leverage constraints, and should be higher for more illiquid assets.

The rest of the paper is organized as follows: section 2 introduces the model and the optimization problem. Section 3 contains the main results, which characterize the efficient frontier in the risk-averse (Theorem 3.1) and risk-neutral (Theorem 3.2) cases. Section 4 discusses the implications of these results for the efficient frontier, the trading boundaries of optimal policies, the embedded leverage effect, and performance evaluation, showing that the tradeoff between alpha and tracking error is equivalent to the tradeoff between average return and volatility. The section concludes with two supporting results, which show that the risk-neutral solutions arise as limits of their risk-averse counterparts for low risk-aversion, and that the risk-neutral solutions are not constrained by the solvency condition. Section 5 offers a derivation of the main free-boundary problems from heuristic control arguments, and concluding remarks are in section 6. All proofs are in the appendix.

2 Model

The market includes one safe asset earning a constant interest rate of \( r \geq 0 \) and a risky asset with ask (buying) price \( S_t \) given by

\[
\frac{dS_t}{S_t} = (\mu + r)dt + \sigma dB_t, \quad S_0, \sigma, \mu > 0,
\]

where \( B \) is a standard Brownian motion. The risky asset’s bid (selling) price is \((1 - \varepsilon)S_t\), which implies a constant relative bid-ask spread of \( \varepsilon > 0 \), or, equivalently, constant proportional transaction costs. A self-financed trading strategy is summarized by its initial capital \( x \) and the number of shares \( \varphi_t \) of the risky asset held at time \( t \). Denote by \( w_t \) the fund’s wealth at time \( t \), which is the sum of the safe position \( x - \int_0^t S_s d\varphi_s - \varepsilon \int_0^t S_s d\varphi^+_s \) and the risky position \( S_t \varphi_t \) evaluated at the ask price\(^3\):

\[
w_t = x - \int_0^t S_s d\varphi_s - \varepsilon \int_0^t S_s d\varphi^+_s + S_t \varphi_t. \tag{2.1}
\]

We further require a strategy \( \varphi \) to be solvent, in that its corresponding wealth \( w_t \) is strictly positive at all times. (Admissible strategies are formally described in Definition A.1 below.)

\(^3\)The convention of evaluating the risky position at the ask price is inconsequential. Using the bid price instead would lead to the same results.
We focus on objectives that trade off a fund’s average return against its realized variance. The portfolio return $r_t$ over the time-interval $[t - \Delta t, t]$ is

$$r_t = \frac{w_t - w_{t-\Delta t}}{w_{t-\Delta t}},$$

while the annualized average return has the continuous-time approximation

$$\bar{r}_T = \frac{1}{n\Delta t} \sum_{t=k\Delta}^0 r_t \approx \frac{1}{T} \int_0^T \frac{dw_t}{w_t}.$$

In the familiar setting of no trading costs, $\frac{1}{T} \int_0^T \frac{dw_t}{w_t} = r + \frac{1}{T} \int_0^T \mu \pi_t dt + \frac{1}{T} \int_0^T \sigma \pi_t dB_t$, where $\pi_t = \varphi_t S_t / w_t$ is the portfolio weight of the risky asset, hence the average return equals the average risky exposure times its excess return, plus the safe rate.

Likewise, the average squared volatility on $[0, T]$ is obtained by the usual variance estimator applied to returns, and has the continuous-time approximation

$$\frac{1}{n\Delta t} \sum_{t=k\Delta}^0 (r_t - \bar{r}_T)^2 \approx \frac{1}{T} \int_0^T \frac{d(w_t)}{w_t^2} = \frac{\sigma^2}{T} \int_0^T \pi_t^2 dt.$$

(The last equality holds because the trading cost term $\varepsilon \int_0^t S_s d\varphi_s^t$ in (2.1) is increasing and continuous, hence its total sum of squares is negligible on a fine grid.) As a result, the usual risk-return trade-off is captured by maximizing

$$\max_{\varphi} \frac{1}{T} \mathbb{E} \left[ \int_0^T \frac{dw_t}{w_t} - \frac{\gamma}{2} \int_0^T \frac{d(w_t)}{w_t^2} \right],$$

where the parameter $\gamma > 0$ is interpreted as a proxy for risk-aversion. This mean-variance objective is relevant for two reasons: first, without trading costs and for a strategy with finite variance, it reduces to

$$\max_{\pi} \frac{1}{T} \mathbb{E} \left[ \int_0^T \left( \mu \pi_t - \frac{\gamma}{2} \sigma^2 \pi_t^2 \right) dt \right],$$

which leads to the optimal constant-proportion portfolio $\pi = \frac{\mu}{\gamma \sigma^2}$ dating back to Markowitz and Merton, and confirms that in a geometric Brownian motion market with costless trading, the objective considered here is equivalent to the utility-maximization with constant relative risk aversion. Second, the objective (2.3) is equivalent to the maximization of alpha for a given tracking error, or, equivalently, to the minimization of tracking error for given alpha, which is relevant for funds that aim at replicating multiples of a benchmark’s return. This equivalence is shown in Section 4.4 below.

To better understand the effect of trading costs, note that the maximand in (2.3) reduces to a more concrete expression.

**Lemma 2.1.** For any $T > 0$ and admissible trading strategy $\varphi$,

$$F_T(\varphi) := \frac{1}{T} \mathbb{E} \left[ \int_0^T \frac{dw_t}{w_t} - \frac{\gamma}{2} \int_0^T \frac{d(w_t)}{w_t^2} \right] = r + \frac{1}{T} \mathbb{E} \left[ \int_0^T \left( \mu \pi_t - \frac{\gamma \sigma^2}{2} \pi_t^2 \right) dt - \varepsilon \int_0^T \pi_t d\varphi_t^t \right].$$
The final extra term in (2.5) with trading costs hinders continuous portfolio rebalancing, making constant-proportion strategies unfeasible. The reason is that it is costly both to keep the exposure to the risky asset high enough to achieve the desired return, and to keep it low enough to limit the level of risk – trading costs reduce returns and increase risk.

To neglect the spurious, non-recurring effects of portfolio set-up and liquidation, we focus on the long run performance objective

\[ F_\infty(\varphi) := \limsup_{T \to \infty} F_T(\varphi), \quad (2.6) \]

which is akin to the one used by Dumas and Luciano (1991) in the context of utility maximization. As a consequence of the law of large numbers, the long run performance criterion entails that ex-ante and ex-post performances coincide, i.e.,

\[ F_\infty(\varphi) := \limsup_{T \to \infty} \frac{1}{T} \mathbb{E} \left[ \int_0^T \frac{dw_t}{w_t} - \frac{\gamma}{2} \int_0^T \frac{d\langle w \rangle_t}{w_t^2} \right] = \limsup_{T \to \infty} \frac{1}{T} \left( \int_0^T \frac{dw_t}{w_t} - \frac{\gamma}{2} \int_0^T \frac{d\langle w \rangle_t}{w_t^2} \right). \quad (2.7) \]

A few special cases are noteworthy. Without transaction costs, maximizing this objective is equivalent to maximizing power utility over any finite horizon \( T \). With or without transaction costs, in the risk-neutral case \( \gamma = 0 \) the objective boils down to the average annualized return over a long horizon, while for \( \gamma = 1 \) it coincides with logarithmic utility.

### 3 Main Results

The first result characterizes the optimal solution to the main objective in (2.6) in the usual case of a positive aversion to risk (\( \gamma > 0 \)). In this setting, the next theorem shows that trading costs create a no-trade region around the frictionless portfolio \( \pi_\ast = \frac{\mu}{\sigma^2} \), and states the asymptotic expansions of the resulting average return and standard deviation\(^5\), thereby extending the familiar efficient frontier to account for trading costs.

**Theorem 3.1** (Risk Aversion and Efficient Frontier). Let \( \gamma \neq \mu/\sigma^2 \).

(i) For any \( \gamma > 0 \) there exists \( \varepsilon_0 > 0 \) such that for all \( \varepsilon < \varepsilon_0 \), the free boundary problem

\[ \frac{1}{2} \sigma^2 \zeta^2 W''(\zeta) + (\sigma^2 + \mu) \zeta W'(\zeta) + \mu W(\zeta) - \frac{1}{(1 + \zeta)^2} \left( \mu - \gamma \sigma^2 \frac{\zeta}{1 + \zeta} \right) = 0, \quad (3.1) \]

\[ W'(\zeta_-) = 0, \quad (3.2) \]

\[ W'(\zeta_-) = 0, \quad (3.3) \]

\[ W'(\zeta_+) = \frac{\varepsilon}{(1 + \zeta_+)(1 + (1 - \varepsilon)\zeta_+)}, \quad (3.4) \]

\[ W'(\zeta_+) = \frac{\varepsilon((1 - \varepsilon)\zeta_+^2 - 1)}{(1 + \zeta_+)^2(1 + (1 - \varepsilon)\zeta_+)^2}, \quad (3.5) \]

has a unique solution \((W, \zeta_-, \zeta_+)\) for which \( \zeta_- < \zeta_+ \).

\(^4\)In this equation the lim sup is used merely to guarantee a good-definition a priori. A posteriori, we show that optimal strategies exist in which the lim sup is in fact a lim, hence the similar problem defined in terms of lim inf leads to the same solution.

\(^5\)The exact formulae for average return, standard deviation, and average trading costs are in Appendix C.
(ii) The trading strategy $\hat{\varphi}$, which buys at $\pi_- := \zeta_-/(1 + \zeta_-)$ and sells at $\pi_+ := \zeta_+/(1 + \zeta_+)$ as little as necessary to keep the risky weight $\pi_t = \zeta_t/(1 + \zeta_t)$ within the interval $[\pi_-, \pi_+]$, is optimal.

(iii) The maximum performance is

\[
\max_{\varphi \in \Phi} \lim_{T \to \infty} \frac{1}{T} \int_0^T dw_t - \frac{\gamma}{2} \int_0^T d(w)_t = r + \mu \pi_- - \frac{\gamma\sigma^2}{2} \pi_-^2.
\]  

where $\Phi$ is the set of admissible strategies in Definition A.1.

(iv) The trading boundaries $\pi_-$ and $\pi_+$ have the asymptotic expansions

\[
\pi_\pm = \pi^*_\pm \left( \frac{3}{4\gamma}(\pi^*_\pm)^2 (1 - \pi^*_\pm)^2 \right)^{1/3} \varepsilon^{1/3} - \frac{(\gamma - 1)(\pi^*_\pm)^2 (1 - \pi^*_\pm)}{6} \left( \frac{6}{\gamma\pi^*_\pm (1 - \pi^*_\pm)} \right)^{2/3} \varepsilon^{2/3} + O(\varepsilon).
\]

The long-run mean $\hat{m}$, standard deviation $\hat{s}$, average trading costs (ATC) and maximum performance $F(\hat{\varphi})$ have expansions (using the convention $a^{1/n} = \text{sign}(a)|a|^{1/n}$ for any $a \in \mathbb{R}$ and odd integer $n$, and $a^{2/n} = (a^2)^{1/n}$)

\[
\hat{m} := \lim_{T \to \infty} \frac{1}{T} \int_0^T \frac{dw_t}{w_t} = r + \frac{\mu^2}{2\gamma\sigma^2} + \frac{\mu(5\pi^*_\pm - 3)}{\gamma} \left( \frac{\gamma\pi^*_\pm (1 - \pi^*_\pm)}{6} \right)^{1/3} \varepsilon^{2/3} + O(\varepsilon),
\]

\[
\hat{s} := \lim_{T \to \infty} \sqrt{\frac{1}{T} \left( \int_0^T \frac{dw_t}{w_t} \right)^2} = \frac{\mu}{\gamma\sigma} + \frac{\sigma(7\pi^*_\pm - 3)}{4\gamma} \left( \frac{\gamma\pi^*_\pm (1 - \pi^*_\pm)}{6} \right)^{1/3} \varepsilon^{2/3} + O(\varepsilon),
\]

\[
\text{ATC} := \lim_{T \to \infty} \frac{1}{T} \int_0^T \frac{d\varphi^\top_t \varphi_t}{\pi_t} = \frac{3\sigma^2}{\gamma} \left( \frac{\gamma\pi^*_\pm (1 - \pi^*_\pm)}{6} \right)^{4/3} \varepsilon^{2/3} + O(\varepsilon),
\]

\[
F_\infty(\hat{\varphi}) = r + \frac{\mu^2}{2\gamma\sigma^2} - \frac{\gamma\sigma^2}{2} \left( \frac{3}{4\gamma} \pi^*_\pm^2 (1 - \pi^*_\pm)^2 \right)^{2/3} \varepsilon^{2/3} + O(\varepsilon).
\]

Proof. The proof is divided into Propositions B.1, B.4 and B.6 in Appendix B.

In contrast to the risk-averse setting, the risk-neutral objective leads to a novel solution, which does not have a frictionless analogue: for small trading costs, both the optimal policy and its performance become unbounded as the optimal leverage increases arbitrarily. The next result describes the solution to the risk-neutral problem, identifying the approximate dependence of the leverage multiplier and its performance on the asset’s risk, return, and liquidity.

**Theorem 3.2 (Risk Neutrality and Limits of Leverage).** Let $\gamma = 0$.

(i) There exists $\varepsilon_0 > 0$ such that for all $\varepsilon < \varepsilon_0$, the free boundary problem (3.1)–(3.5) has a unique solution $(W, \zeta_-, \zeta_+)$ with $\zeta_- < \zeta_+$.

(ii) The trading strategy $\hat{\varphi}$ that buys at $\pi_- := \zeta_-/(1 + \zeta_-)$ and sells at $\pi_+ := \zeta_+/(1 + \zeta_+)$ to keep the risky weight $\pi_t = \zeta_t/(1 + \zeta_t)$ within the interval $[\pi_-, \pi_+]$ is optimal.

(iii) The maximum expected return is

\[
\max_{\varphi \in \Phi} \lim_{T \to \infty} \frac{1}{T} \int_0^T \frac{dw_t}{w_t} = r + \mu \pi_-.
\]
(iv) The trading boundaries have the series expansions

\[ \begin{align*}
\pi_- &= (1 - \kappa)\kappa^{1/2} \left( \frac{\mu}{\sigma^2} \right)^{1/2} \varepsilon^{-1/2} + 1 + O(\varepsilon^{1/2}), \\
\pi_+ &= \kappa^{1/2} \left( \frac{\mu}{\sigma^2} \right)^{1/2} \varepsilon^{-1/2} + 1 + O(\varepsilon^{1/2}),
\end{align*} \]

(3.13)

(3.14)

where \( \kappa \approx 0.5828 \) is the unique solution of

\[ \frac{3}{2} \xi + \log(1 - \xi) = 0, \quad \xi \in (0, 1). \]

The next section discusses how these results modify the familiar intuition about risk, return, and performance evaluation in the context of trading costs.

4 Implications

4.1 Efficient frontier

Theorem 3.1 extends the familiar efficient frontier to account for trading costs. Compared to the linear frictionless frontier, average returns decline because of rebalancing losses. Average volatility increases because more risk becomes necessary to obtain a given return net of trading costs.

To better understand the effect of trading costs on return and volatility, consider the dynamics of the portfolio weight in the absence of trading, which is

\[ d\pi_t = \pi_t (1 - \pi_t)(\mu - \sigma^2 \pi_t)dt + \sigma \pi_t (1 - \pi_t)dB_t. \]

(4.1)

The central quantity here is the portfolio weight volatility \( \sigma \pi_t (1 - \pi_t) \), which vanishes for the single-asset portfolios \( \pi_t = 0 \) or \( \pi_t = 1 \), remains bounded above by \( \sigma/4 \) in the long-only case \( \pi_t \in [0, 1] \), and rises quickly with leverage (\( \pi_t > 1 \)). This quantity is important because it measures the extent to which a portfolio, left to itself, strays from its initial composition in response to market shocks and, by reflection, the quantity of trading that is necessary to keep it within some region. In the long-only case, the portfolio weight volatility decreases as the no-trade region widens to span \([0, 1]\), which means that a portfolio tends to spend more time near the boundaries. By contrast, with leverage portfolio weight volatility increases, which means that a wider boundary does not necessarily mitigate trading costs.

Consistent with this intuition, equations (3.8), (3.9) show that the impact of trading costs is smaller on long-only portfolios, but rises quickly with leverage. Small trading costs reduce returns and increase volatility at the order of \( \varepsilon^{2/3} \) but, crucially, as leverage increases the error of this approximation also increases, and lower values of \( \gamma \) make it precise for ever smaller values of \( \varepsilon \).

The performance (3.11) coincides at the first order with the equivalent safe rate from utility maximization with constant relative risk aversion \( \gamma \) (Gerhold et al., 2014, Equation (2.4)), supporting the interpretation of \( \gamma \) as a risk-aversion parameter, and confirming that, for asymptotically small costs, the efficient frontier captures the risk-return trade-off faced by a utility maximizer.

Figure 2 displays the effect of trading costs on the efficient frontier. As the bid-ask spread declines, the frontier increases to the linear frictionless frontier, and the asymptotic results in the theorem become more accurate. However, if the spread is held constant as leverage (hence volatility) increases, the asymptotic expansions become inaccurate, and in fact the efficient frontier ceases to increase at all after the leverage multiplier is reached.
Figure 2: Efficient Frontier with trading costs, as expected return (vertical axis, in multiples of the asset’s expected return) against standard deviation (horizontal axis, in multiples of the asset’s volatility). The asset has expected return $\mu = 8\%$, volatility $\sigma = 16\%$, and bid-ask spread of 0.1%, 0.5%, 1%, 2%. The upper line is the frictionless efficient frontier. The maximum of each curve is the leverage multiplier.

4.2 Trading Boundaries

Each point in the efficient frontier corresponds to a rebalancing strategy that is optimal for some value of the risk-aversion parameter $\gamma$. For small trading costs, equation (3.7) implies that the trading boundaries corresponding to the efficient frontier depart from the ones arising in utility maximization, which are (Gerhold et al., 2014)

$$
\pi_{\pm} = \pi_* \pm \left( \frac{3}{4\gamma} \right) \left( \frac{1 - \pi_0}{2(1 - \pi_*^2)} \right)^{1/3} \varepsilon^{1/3} + O(\varepsilon).
$$

The term of order $\varepsilon^{2/3}$ vanishes for $\gamma = 1$ because this case coincides with the maximization of logarithmic utility. For high levels of leverage ($\gamma < 1$ and $\pi_* > 1$), this term implies that the trading boundaries that generate the efficient frontier are higher than the trading boundaries that maximize utility. In Figure 3, $\gamma \to \infty$ corresponds to the safe portfolio in the origin (0,0), while $\gamma = \mu/\sigma^2$ to the risky investment (1,1), which has by definition the same volatility and return as the risky asset. As $\gamma$ declines to zero, the trading boundaries converge to the right endpoints, which correspond to the strategy that maximizes average return with no regard for risk, thereby achieving the multiplier.

Observe that (Figure 3), as leverage increases, the sell boundary rises more quickly than the buy boundary. For example, the risk-neutral portfolio tolerates leverage fluctuations from approxi-
Figure 3: Trading boundaries $\pi_{\pm}$ (vertical axis, as multiples of wealth in risky assets) against average portfolio volatility (horizontal axis, as multiples of $\sigma$). Parameter values are $\mu = 8\%$, $\sigma = 16\%$ and $\varepsilon = 1\%$. The middle curve depicts the implied Merton fraction, which equals the first median.

Importantly, these boundaries remain finite even as the frictionless Merton portfolio $\mu/(\gamma \sigma^2)$ diverges to infinity as $\gamma$ declines to zero. Thus the no-trade region is obviously not symmetric around the frictionless portfolio, in contrast to the boundaries arising from utility maximization (Gerhold et al., 2014), which are always symmetric, and hence diverge when $\gamma$ is low. The difference is that here the risk-neutral objective is to maximize the expected return of the portfolio, while a risk-neutral utility maximizer focuses on expected wealth. In a frictionless setting this distinction is irrelevant, and an investor can use a return-maximizing policy to maximize wealth instead. But trading costs drive a wedge between these two ostensibly equivalent risk-neutral criteria – maximizing expected return is not the same as maximizing expected wealth.

Theorem 3.2 (iii) describes in the risk-neutral case the optimal trading boundaries, which satisfy...
the approximate relation
\[ \frac{\pi_-}{\pi_+} \approx 0.4172 \] (4.3)
which is universal in that it holds for any asset, regardless of risk, return and liquidity. This relation means that an optimal risk-neutral rebalancing strategy should always tolerate wide variations in leverage over time, and that the maximum allowed leverage should be approximately 2.5 times the minimum. More frequent rebalancing cannot achieve the maximum return: it can be explained either by risk aversion or by elements that lie outside the model, such as jumps in asset prices.

The liquidation constraint (A.1) implies that
\[ \pi_t < \frac{1}{\varepsilon} \] (4.4)
for every admissible trading strategy. Since \( \pi_t \leq \pi_+ \) for the optimal trading strategy given by Theorem 3.1 and Theorem 3.2, the upper bound (4.4) is never binding for realistic bid-ask-spreads.

### 4.3 Embedded leverage

In frictionless markets, two perfectly correlated assets with equal Sharpe ratio generate the same efficient frontier, and in fact the same payoff space. This equivalence fails in the presence of trading costs: the more volatile asset is superior, in that it generates an efficient frontier that dominates the one generated by the less volatile asset. Figure 4 (top of the three curves) displays this phenomenon: for example, a portfolio with an average return of 50% net of trading costs is obtained from an asset with 25% return and 50% volatility at a small cost, as an average leverage factor of 2 entails moderate rebalancing.

Achieving the same 50% return from an asset with 20% volatility (and 10% return) is more onerous: trading costs require leverage higher than 5, which in turn increases trading costs. Overall, the resulting portfolio needs about 120% rather than 100% volatility to achieve the desired 50% average return (middle curve in Figure 4).

From an asset with 10% volatility (and 5% return), obtaining a 50% return net of trading costs is impossible (bottom curve in Figure 4), because the leverage multiplier is less than 8 (Table 1, top right), and therefore the return can be scaled to less than 40%. The intuition is clear: increasing leverage also increases trading costs, calling in turn for more leverage to increase return, but also further increasing costs. At some point, the marginal net return from more leverage becomes zero, and increasing it does more harm than good.

Because an asset with higher volatility is superior to another one, perfectly correlated and with equal Sharpe ratio, but with lower volatility, the model suggests that in equilibrium they cannot coexist, and that the asset with lower volatility should offer a higher return to be held by investors. Indeed, Frazzini and Pedersen (2012, 2014) document significant negative excess returns in assets with embedded leverage (higher volatility), and offer a theoretical explanation based on heterogeneous leverage constraints, which lead more constrained investors to bid up prices (and hence lower returns) of more volatile assets. This paper suggests that the same phenomenon may arise even in the absence of constraints, as a result of rebalancing costs. In contrast to constraints-based explanations, our model suggests that the premium for embedded leverage should be higher for more illiquid assets.

12
4.4 Performance Evaluation: Alpha vs. Tracking Error

Portfolio performance measures based on the regression of a fund’s return against its benchmark’s return are ubiquitous and, as a result, are closely monitored by managers who are evaluated with such performance measures.

A manager who seeks to replicate a multiple of benchmark’s return over each period faces the choice between rebalancing often to match the target return closely, thereby achieving low tracking error but high trading costs, or rebalancing infrequently to reduce costs, but at the expense of substantial tracking error. The next argument shows that this trade-off is equivalent to the one in Theorems 3.1 and 3.2, and explains the implications of these results for performance evaluation.

Denote by $r_t^F$ and $r_t^B$ the returns of a fund and its (unlevered) benchmark, respectively, over the time step $[t - \Delta t, t]$. If a fund aims at tracking a multiple $\pi_s$ of the benchmark’s return, its average excess return $\bar{\alpha}_T$ per unit of time, also known as alpha, is

$$
\bar{\alpha}_T = \frac{1}{n\Delta t} \sum_{t=k\Delta t}^{t} (r_t^F - \pi_s r_t^B) \approx \frac{1}{T} \int_0^T \left( \frac{dw_t^F}{w_t^F} - \pi_s \frac{dw_t^B}{w_t^B} \right).
$$

(4.5)
The corresponding realized tracking error, defined as realized error standard deviation per unit of time and denoted by $\bar{s}$, satisfies

$$\bar{s}^2 = \frac{1}{n\Delta t} \sum_{t=k\Delta t}^{0 \leq t \leq T} (r_t^F - \pi_* r_t^B - \bar{\alpha})^2 \approx \frac{1}{T} \left( \int_0^T \left( \frac{dw_F^T}{w_F^T} - \pi_* \frac{dw_B^T}{w_B^T} \right) \right).$$

(4.6)

Simplifying the continuous-time quantities in the right-hand sides, the objective of maximizing alpha with a penalty (or, equivalently, a constraint) for tracking error reduces to maximize

$$\frac{1}{T} \mathbb{E} \left[ \int_0^T \mu (\pi_t - \pi_*) dt - \varepsilon \int_0^T \frac{d\phi_t^\gamma}{\phi_t} - \frac{\gamma}{2} \int_0^T \sigma^2 (\pi_t - \pi_*)^2 dt \right].$$

(4.7)

The first term in this expression represents the excess return that results from the difference between the fund’s average exposure and its target exposure. The second term accounts for the losses from trading costs and, added to the first one, yields the overall fund’s alpha. The last term is the penalty for tracking error, which results from the cumulative squared difference between the fund and the target exposure. With elementary manipulations, the previous expression simplifies to

$$\frac{1}{T} \mathbb{E} \left[ \int_0^T \left( (\mu + \gamma \sigma^2 \pi_*) \pi_t - \frac{\gamma}{2} \sigma^2 \pi_t^2 \right) dt - \varepsilon \int_0^T \frac{d\phi_t^\gamma}{\phi_t} - \mu \pi_* - \frac{\gamma}{2} \sigma^2 \pi_* \right]$$

and therefore the trade-off between alpha and tracking error is equivalent to the trade-off between risk and return, up to replacing $\mu$ with $\bar{\mu} = \mu + \gamma \sigma^2 \pi_*$, where $\pi_*$ is the target leverage.

Such an equivalence is intuitive in hindsight, as in a market with excess return $\bar{\mu}$ the optimal frictionless portfolio is $\bar{\mu}/(\gamma \sigma^2) = \pi_* + \mu/(\gamma \sigma^2)$. As a result, higher risk aversion $\gamma$ leads to a portfolio that is closer to the target $\pi_*$, thereby focusing on low tracking error while disregarding the excess return. Vice versa, lower risk aversion implies a portfolio that overweighs the risky asset in order to generate alpha, while accepting higher tracking error. This intuition remains valid with transaction costs, but the risk-return trade-off is altered by rebalancing costs.

Figure 5 displays the optimal tracking error as a function of average leverage. Consistent with the intuition underlying the previous results, higher average leverage entails wider boundaries, which result in higher tracking error. In particular, high tracking error is not necessarily evidence of poor manager performance if the underlying asset is illiquid. On the contrary, a savvy management strategy must accept higher tracking error to achieve high returns, and low tracking error is consistent only with moderate leverage levels.

A scale-free indicator of tracking error that is used in practice is the $R^2$ of the time-series regression of the fund’s returns against the benchmark returns. First, note that the estimated slope, or beta, of this regression has the continuous-time approximation

$$\beta_T \approx \frac{\langle \int_0^T \frac{dw_F^T}{w_F^T}, \int_0^T \frac{dw_B^T}{w_B^T} \rangle_T}{\langle \int_0^T \frac{dw^B}{w^B} \rangle_T} = \frac{\int_0^T \pi_t \sigma^2 dt}{\sigma^2 T} = \frac{1}{T} \int_0^T \pi_t dt.$$

(4.9)

As a result, the long horizon $R^2$ of the regression, defined as the ratio between the variance of the predicted return and the variance of the realized return, is

$$R^2 = \lim_{T \to \infty} \frac{1}{T} \left( \frac{1}{T} \int_0^T \pi_t dt \right)^2.$$ 

(4.10)
Figure 5: Tracking error $\sqrt{1 - R^2}$ for different levels of leverage (horizontal axis), for spreads of $\varepsilon = 0.01\%$ (bottom), 0.1\% (center), and 1\% (top). The curves display values of $\gamma \in [0, \infty]$.

The next result offers an asymptotic approximation for this quantity, similar to the expansions in the main results.

**Proposition 4.1.** The following expansions hold

$$
\bar{\alpha} = - ATC = - \frac{3\sigma^2}{\gamma} \left( \frac{\gamma \pi_s (1 - \pi_s)}{6} \right)^{4/3} \varepsilon^{2/3} + O(\varepsilon), \quad (4.11)
$$

$$
\sqrt{1 - R^2} = \frac{\sqrt{3}}{6} \left( \frac{6(1 - \pi_s)^2}{\gamma \pi_s} \right)^{1/3} \varepsilon^{1/3} + O(\varepsilon^1), \quad (4.12)
$$

and therefore

$$
\bar{\alpha} = - \frac{\sqrt{3}}{12} \sigma^2 \pi_s (1 - \pi_s)^2 \frac{\varepsilon}{\sqrt{1 - R^2}} + O(\varepsilon^{4/3}). \quad (4.13)
$$

This result shows that the alpha of a levered portfolio, abstracting from management fees, equals minus the expected costs. The tracking error departs from zero as the spread $\varepsilon$ increases, and as the target leverage $\pi_s$ rises above the buy-and-hold level of one. Other things equal, a fund that seeks to replicate a larger multiple $\pi_s$ of a benchmark’s return has a higher tracking error and a more negative alpha. Thus, it is misleading to compare two funds with different targets, and to
conclude that one is better managed than the other merely because its $R^2$ is higher, or because its alpha is less negative. Even two funds with the same target may be optimally managed, and yet lead to different alpha and $R^2$, as one may seek lower tracking error at the expense of more negative alpha.

Equation (4.13) offers an approximate relation in terms of observable quantities only, and can be used as a measure of replication performance that controls for the effects of trading costs, volatility, target leverage, and tracking error. Thus, subtracting from the realized alpha the fund’s expense ratio and the right-hand side of (4.13) yields the amount of alpha that is unexplained by the model, and hence can be plausibly attributed to managerial skill – or lack thereof.

### 4.5 From risk aversion to risk neutrality

Theorems 3.1 and 3.2 are qualitatively different: while Theorem 3.1 with positive risk aversion leads to a regular perturbation of the Markowitz-Merton solution, Theorem 3.2 with risk-neutrality leads to a novel result with no meaningful analogue in the frictionless setting – a singular perturbation. Furthermore, a close reading of the statement of Theorem 3.1 shows that the existence of a solution to the free-boundary problem, and the asymptotic expansions, hold for $\varepsilon$ less than some threshold $\bar{\varepsilon}(\gamma)$ that depends on the risk aversion $\gamma$. In particular, if $\gamma$ approaches zero while $\varepsilon$ is held constant, Theorem 3.1 does not offer any conclusions on the convergence of the risk-averse to the risk-neutral solution. Still, if the risk-neutral result it to be accepted as a genuine phenomenon rather than an artifact, it should be clarified whether the risk averse trading policy and its performance converge to their risk neutral counterparts as risk aversion vanishes.

The next result resolves this point under some parametric restrictions.

**Theorem 4.2.** Let $\mu > \sigma^2$, $\bar{\varepsilon} > 0$, and $\bar{\gamma} > 0$, and set

$$G(\zeta) := \frac{\varepsilon}{(1 + \zeta)(1 + (1 - \bar{\varepsilon})\zeta)}, \quad h(\zeta) = \mu \zeta - \frac{\gamma \sigma^2}{2} \left(\frac{\zeta}{1 + \zeta}\right)^2.$$  

Assume that, for any $\gamma \in [0, \bar{\gamma}]$ the free boundary problem (3.1) has a unique solution $(W(;\gamma), \zeta_-(\gamma), \zeta_+(\gamma))$ and that the function

$$\hat{W}(\zeta; \gamma) := \begin{cases} 0, & \zeta < \zeta_-(\gamma) \\ W(\zeta; \gamma), & \zeta \in [\zeta_-(\gamma), \zeta_+(\gamma)] \\ G(\zeta), & \zeta \geq \zeta_+(\gamma) \end{cases}$$

satisfies the equation

$$\min \left( \frac{\sigma^2}{2} \zeta^2 \hat{W}'(\zeta; \gamma) + \mu \zeta \hat{W}(\zeta; \gamma) - h(\zeta) + h(\zeta_-(\gamma), G(\zeta) - \hat{W}(\zeta; \gamma), \hat{W}(\zeta; \gamma) \right) = 0. \quad (4.14)$$

Then, (4.14) is satisfied also for $\gamma = 0$, and for each $\gamma \in [0, \bar{\gamma}]$, the trading strategy that buys at $\pi_-(\gamma) = \frac{\zeta_-(\gamma)}{1 + \zeta_-(\gamma)}$ and sells at $\pi_+(\gamma) = \frac{\zeta_+(\gamma)}{1 + \zeta_+(\gamma)}$ to keep the risky weight $\pi_t = \zeta_t / (1 + \zeta_t)$ within the interval $[\pi_-(\gamma), \pi_+(\gamma)]$ is optimal. Furthermore, $\zeta_+(\gamma) \to \zeta_+(0)$ and $\hat{W}(\zeta; \gamma) \to W(\zeta; 0)$ as $\gamma \downarrow 0$, each $\zeta \in \mathbb{R}$.

In summary, this result confirms that, as the risk-aversion parameter $\gamma$ declines to zero, the risk-averse policy in Theorem 3.1 can only converge to the risk-neutral policy in Theorem 3.2, and that the corresponding mean-variance objective in Theorem 3.1 converges to the average return in Theorem 3.2.
5 Heuristic Solution

This section offers a heuristic derivation of the HJB equation. Consider the finite-horizon objective

\[
\max_{\varphi \in \Phi} \mathbb{E} \left[ \int_0^T \left( \mu \pi_t - \frac{\gamma \sigma^2}{2} \pi_t^2 \right) dt - \varepsilon \int_0^T \pi_t \frac{d\varphi^\downarrow_t}{\varphi_t} \right]
\]  

(5.1)

From the outset, it is clear that this objective is scale-invariant, because doubling the initial number of risky shares and safe units, and also doubling the number of shares \( \varphi_t \) held at time \( t \) has the effect of keeping the objective functional constant. Thus, we conjecture that the residual value function \( V \) depends on the calendar time \( t \) and on the variable \( \zeta_t = \pi_t/(1 - \pi_t) \), which denotes the number of shares held for each unit of the safe asset. In terms of this variable, the conditional value of the above objective at time \( t \) becomes:

\[
F^\varphi(t) = \int_0^t \left( \mu \frac{\zeta_s}{1 + \zeta_s} - \frac{\gamma \sigma^2}{2} \frac{\zeta_s^2}{(1 + \zeta_s)^2} \right) ds - \varepsilon \int_0^t \frac{\zeta_s}{1 + \zeta_s} \frac{d\varphi^\downarrow_s}{\varphi_s} + V(t, \zeta_t).
\]  

(5.2)

By Itô’s formula, the dynamics of \( F^\varphi \) is

\[
dF^\varphi(t) = \left( \mu \frac{\zeta_t}{1 + \zeta_t} - \frac{\gamma \sigma^2}{2} \frac{\zeta_t^2}{(1 + \zeta_t)^2} \right) dt - \varepsilon \frac{\zeta_t}{1 + \zeta_t} \frac{d\varphi^\downarrow_t}{\varphi_t} + V_t(t, \zeta_t)dt + V_\zeta(t, \zeta_t)d\zeta_t + \frac{1}{2} V_{\zeta\zeta}(t, \zeta_t)d(\zeta_t),
\]

where subscripts of \( V \) denote respective partial derivatives. The self-financing condition (2.1) implies that

\[
\frac{d\zeta_t}{\zeta_t} = \mu dt + \sigma dW_t + (1 + \zeta_t) \frac{d\varphi^\downarrow_t}{\varphi_t} + \varepsilon \zeta_t \frac{d\varphi^\downarrow_t}{\varphi_t},
\]  

(5.3)

which in turn allows to simplify the dynamics of \( F^\varphi \) to (henceforth the arguments of \( V \) are omitted for brevity)

\[
dF^\varphi(t) = \left( \mu \frac{\zeta_t}{1 + \zeta_t} - \frac{\gamma \sigma^2}{2} \frac{\zeta_t^2}{(1 + \zeta_t)^2} + V_t + \frac{\sigma^2}{2} \zeta_t^2 V_\zeta + \mu \zeta_t V_\zeta \right) dt
\]

(5.4)

\[ -\zeta_t \left( V_\zeta(1 + (1 - \varepsilon)\zeta_t) + \frac{\varepsilon}{1 + \zeta_t} \right) \frac{d\varphi^\downarrow_t}{\varphi_t} + \zeta_t(1 + \zeta_t)V_\zeta \frac{d\varphi^\downarrow_t}{\varphi_t} + \sigma \zeta_t V_\zeta dW_t. \]

(5.5)

Now, by the martingale principle of optimal control (Davis and Varaiya, 1973) the process \( F^\varphi(t) \) above needs to be a supermartingale for any trading policy \( \varphi \), and a martingale for the optimal policy. Since \( \varphi^\downarrow \) and \( \varphi^\downarrow \) are increasing processes, the supermartingale condition implies the inequalities

\[
-\frac{\varepsilon}{(1 + \zeta)(1 + (1 - \varepsilon)\zeta)} \leq V_\zeta \leq 0,
\]  

(5.6)

and the martingale condition prescribes that the left (respectively, right) inequality becomes an equality at the points of increase of \( \varphi^\downarrow \) (resp. \( \varphi^\downarrow \)). Likewise, it follows that

\[
\mu \frac{\zeta}{1 + \zeta} - \frac{\gamma \sigma^2}{2} \frac{\zeta^2}{(1 + \zeta)^2} + V_t + \frac{\sigma^2}{2} \zeta^2 V_\zeta + \mu \zeta V_\zeta \leq 0
\]  

(5.7)

with the inequality holding as an equality whenever both inequalities in (5.6) are strict. To achieve a stationary (that is, time-homogeneous) system, suppose that the residual value function is of the
form \( V(t, \zeta) = \lambda(T - t) - \int^t W(z)dz \) for some \( \lambda \) to be determined, which represents the average optimal performance over a long period of time. Replacing this parametric form of the solution, the above inequalities become

\[
0 \leq W(\zeta) \leq \frac{\epsilon}{(1 + \zeta)(1 + (1 - \epsilon)\zeta)},
\]

\[
\mu - \frac{\zeta}{1 + \zeta} - \frac{\gamma \sigma^2}{2} \frac{\zeta^2}{(1 + \zeta)^2} - \lambda - \frac{\sigma^2}{2} \zeta^2 W'(\zeta) - \mu \zeta W(\zeta) \leq 0,
\]

Assuming further that the first inequality holds over some interval \([\zeta_-, \zeta_+],\) with each inequality reducing to an equality at the respective endpoint, the optimality conditions become

\[
\frac{\sigma^2}{2} \zeta^2 W'(\zeta) + \mu \zeta W(\zeta) - \mu \frac{\zeta}{1 + \zeta} + \frac{\gamma \sigma^2}{2} \frac{\zeta^2}{(1 + \zeta)^2} + \lambda = 0 \quad \text{for } \zeta \in [\zeta_-, \zeta_+],
\]

\[
W(\zeta_-) = 0,
\]

\[
W(\zeta_+) = \frac{\epsilon}{(\zeta_+ + 1)(1 + (1 - \epsilon)\zeta_+)},
\]

which lead to a family of candidate value functions, each of them corresponding to a pair or boundaries \((\zeta_-, \zeta_+).\) The optimal boundaries are identified by the smooth-pasting conditions, formally derived by differentiating (5.11) and (5.12) with respect to their boundaries

\[
W'(\zeta_-) = 0,
\]

\[
W'(\zeta_+) = \frac{\epsilon((1 - \epsilon)\zeta_+^2 - 1)}{(1 + \zeta_+)^2(1 + (1 - \epsilon)\zeta_+)^2}.
\]

These conditions allow to identify the value function. The four unknowns are the free parameter in the general solution to the ordinary differential equation (5.10), the free boundaries \( \zeta_- \) and \( \zeta_+, \) and the optimal rate \( \lambda. \) These quantities are identified by the boundary and smooth-pasting conditions (5.11), (5.12), (5.13), (5.14).

6 Conclusion

The costs of rebalancing a levered portfolio are substantial, and detract from its ostensible frictionless return. As leverage increases, such costs rise faster than the frictionless return, making it impossible for an investor to lever an asset’s return beyond a certain multiple, net of trading costs.

In contrast to the frictionless theory, trading costs make the risk-return tradeoff nonlinear. An investor who seeks high return prefers an asset with high volatility to another one with equal Sharpe ratio but lower volatility, because higher volatility makes leverage cheaper to realize. A risk-neutral, return-maximizing investor does not take infinite leverage, but rather keeps it within a band that balances high exposure with low rebalancing costs.

These findings have broad implications in portfolio choice, asset pricing, and financial intermediation. For example, a bank that extends risky loans is akin to an investor trading in an illiquid risky asset: in constrast to the frictionless common wisdom, our results imply that such a bank will not increase its balance sheet without bounds, even if it is neutral to risk and regulatory capital requirements are absent. However, the endogenous finite leverage is sensitive to the volatility and
the liquidity of the loans, suggesting that attempts to encourage or discourage bank lending should address these factors.

A direct application of these results is in the performance evaluation of leveraged ETF. Optimal replication strategies face a tradeoff between high correlation with the benchmark and lower alpha, and we derive a testable restriction that any optimal replication policy must satisfy.

A Admissible Strategies

A strategy is admissible if it is nonanticipative and solvent, up to a small increase in the spread:

Definition A.1. Let $x > 0$ (the initial capital) and let $(\varphi^\uparrow_t)_{t \geq 0}$ and $(\varphi^\downarrow_t)_{t \geq 0}$ (the cumulative number of shares bought and sold, respectively) be continuous, increasing processes, adapted to the augmented natural filtration of $B$. $(x, \varphi_t = \varphi^\uparrow_t - \varphi^\downarrow_t)$ is an admissible strategy if its liquidation value is strictly positive at all times: There exists $\varepsilon' > \varepsilon$ such that

$$x - \int_0^t S_s d\varphi_s + S_t \varphi_t - \varepsilon' \int_0^t S_s d\varphi^\downarrow_s - \varepsilon' \varphi^\downarrow_t S_t > 0 \quad \text{a.s. for all } t \geq 0. \quad (A.1)$$

We denote the family of admissible trading strategies by $\Phi$.

The following lemma shows that, without loss of generality, it is safe to exclude trading strategies that involve short selling or violate certain integrability conditions, because each such strategy cannot be optimal.

Lemma A.2. Let $\varphi \in \Phi$ be optimal for (2.6). Then:

(i) the strategy $\hat{\varphi}_t := \varphi_t 1_{\{\varphi_t \geq 0\}}$ is also optimal; and

(ii) the following integrability conditions hold

$$\int_0^t \pi_u^2 du < \infty, \quad \int_0^t \pi_u d\|\varphi_u\|_u < \infty \quad \text{a.s. for all } t \geq 0,$$

where $\|\varphi_u\|$ denotes the total variation of $\varphi$ on $[0,t]$.\footnote{Note that $\frac{\rho_u}{\varphi_t} = \frac{S_u}{\varphi_t}$, therefore on the set $\{(\omega, t) : \varphi_t = 0\}$ the quantity $\frac{\rho_u}{\varphi_t}$ is well-defined.}

Proof. Proof of (i): It is clear that $\hat{\varphi}_t$ is an admissible trading strategy if $\varphi$ is. Furthermore $\hat{\pi}_t \geq \pi_t$ at all times $t$, and $\hat{\pi}_t = 0$ whenever $\varphi_t < 0$, whence $F(\hat{\varphi}, T) \geq F(\varphi, T)$, each $T > 0$.

Proof of (ii): The first integrability condition in (A.2) is trivially satisfied, since it is a direct consequence of (A.1), which in turn implies $\pi_t \leq 1/\varepsilon$, for all $t$, a.s.. To check the second one, suppose that some continuous, finite variation trading strategy $\varphi$ satisfying $\pi_t \leq 1/\varepsilon$ exists such that for some $t > 0$

$$L_t := \int_0^t \pi_u \frac{d|\varphi|_u}{\varphi_u} = \infty$$

with positive probability. Then on the same event one of the following integrals must be infinite, i.e.

$$L_t^1 := \int_0^t \pi_u \frac{d\varphi^\uparrow_u}{\varphi_u} = \infty, \quad \text{or} \quad L_t^2 := \int_0^t \pi_u \frac{d\varphi^\downarrow_u}{\varphi_u} = \infty.$$
The former case leads to infinite average transaction costs (the third term times $-1/\varepsilon$ in (2.5)), hence the objective functional equals $-\infty$, which is outperformed by any buy-and-hold strategy. Likewise, if $L_t^2 = \infty$, then

$$\log(\varphi_t) - \log(\varphi_0) = \int_0^t \frac{d\varphi_u}{\varphi_u} = \infty$$

which is impossible, since $\varphi_t$ must be finite at all times due to proportional transaction costs.

The following lemma describes the dynamics of the wealth process $w_t$, the risky weight $\pi_t$, and the risky/safe ratio $\zeta_t$.

**Lemma A.3.** For any admissible trading strategy $\varphi$, we have\textsuperscript{7}

$$\frac{d\zeta_t}{\zeta_t} = \mu dt + \sigma dB_t + (1 + \zeta_t) \frac{d\varphi_t}{\varphi_t} - (1 + (1 - \varepsilon) \zeta_t) \frac{d\varphi_t}{\varphi_t},$$  \hspace{1cm} (A.3)

$$\frac{dw_t}{w_t} = r dt + \pi_t (\mu dt + \sigma dB_t - \varepsilon \frac{d\varphi_t}{\varphi_t}),$$  \hspace{1cm} (A.4)

$$\frac{d\pi_t}{\pi_t} = (1 - \pi_t) (\mu dt + \sigma dB_t) - \pi_t (1 - \pi_t) \sigma^2 dt + \frac{d\varphi_t}{\varphi_t} - (1 - \varepsilon \pi_t) \frac{d\varphi_t}{\varphi_t}.$$  \hspace{1cm} (A.5)

For any such strategy, the functional

$$F(\varphi, T) := \frac{1}{T} \mathbb{E} \left[ \int_0^T \frac{dw_t}{w_t} - \frac{\gamma}{2} \int_0^T \frac{d(w)_t}{w_t^2} \right],$$  \hspace{1cm} (A.6)

equals to

$$F(\varphi, T) = r + \frac{1}{T} \mathbb{E} \left[ \int_0^T \left( \mu \pi_t - \frac{\gamma \sigma^2}{2} \pi_t^2 \right) dt - \varepsilon \int_0^T \pi_t \frac{d\varphi_t}{\varphi_t} \right].$$  \hspace{1cm} (A.7)

**Proof.** Denoting by $X_t$ and $Y_t$ the wealth in the safe and risky positions respectively, the self-financing condition boils down to

$$dX_t = rX_t dt - S_t d\varphi_t^\dagger + (1 - \varepsilon) S_t d\varphi_t^\dagger,$$  \hspace{1cm} (A.8)

$$dY_t = S_t d\varphi_t^\dagger - S_t d\varphi_t^\dagger + \varphi_t dS_t.$$  \hspace{1cm} (A.9)

and hence

$$\frac{dX_t}{X_t} = r dt - \zeta_t \frac{d\varphi_t^\dagger}{\varphi_t} + (1 - \varepsilon) \zeta_t \frac{d\varphi_t^\dagger}{\varphi_t},$$  \hspace{1cm} (A.10)

$$\frac{dY_t}{Y_t} = \frac{d\varphi_t^\dagger}{\varphi_t} - \frac{d\varphi_t^\dagger}{\varphi_t} + dS_t,$$  \hspace{1cm} (A.11)

$$\frac{d(Y_t/X_t)}{Y_t/X_t} = \frac{dY_t}{Y_t} - \frac{dX_t}{X_t} + \frac{d(X)_t}{X_t^2} - \frac{d(X,Y)_t}{X_t Y_t} = \frac{dY_t}{Y_t} - \frac{dX_t}{X_t}.$$  \hspace{1cm} (A.12)

Equation (A.3) follows from the last equation, and (A.4) holds in view of equation (A.10) and (A.11). For the derivation of equation (A.5), one uses the identity $\pi_t = 1 - \frac{1}{1+c_t}$ and (A.3). The expression in (A.7) for the objective functional follows from equation (A.4). \hspace{1cm} $\square$

\textsuperscript{7}The notation $\frac{dX_t}{X_t} = dY_t$ means $X_t = X_0 + \int_0^t X_s dY_s$, hence the SDEs are well defined even for zero $X_t$. 

20
Equation (A.7) implies an a priori upper bound on the objective function.

**Corollary A.4.** For any admissible strategy \( \varphi \) and any \( T > 0 \), the finite-horizon objective (A.6) satisfies

\[
F(\varphi, T) \leq r + \frac{\mu^2}{2\gamma \sigma^2},
\]

whence the long-run objective \( F(\varphi) \leq r + \frac{\mu^2}{2\gamma \sigma^2} \).

**Proof.** Follows directly from (A.7) and the pointwise inequality \( \mu \pi_t - \frac{\gamma \sigma^2}{2} \pi_t^2 \leq \frac{\mu^2}{2\gamma \sigma^2} \). \( \square \)

### A.1 Proof of Lemma 2.1

**Proof of Lemma 2.1.** See Lemma A.3.

### B Proof of Theorem 3.1

This section contains a series of propositions, which lead to the proof of Theorem 3.1 (i)–(iii). Part (iv) of the theorem is postponed to Appendix C. Set

\[
G(\zeta) := \frac{\varepsilon}{(1 + \zeta)(1 + (1 - \varepsilon)\zeta)} \quad \text{and} \quad h(\zeta) := \mu \left( \frac{\zeta}{1 + \zeta} \right) - \frac{\gamma \sigma^2}{2} \left( \frac{\zeta}{1 + \zeta} \right)^2.
\]

Defining \( H := h' \), the free boundary problem (3.1)–(3.5) reduces to

\[
\begin{align*}
\frac{1}{2} \sigma^2 \zeta^2 W''(\zeta) + (\sigma^2 + \mu) \zeta W'(\zeta) + \mu W(\zeta) - H(\zeta) &= 0, \\
W(\zeta_-) &= 0, \\
W'(\zeta_-) &= 0, \\
W(\zeta_+) &= G(\zeta_+), \\
W'(\zeta_+) &= G'(\zeta_+).
\end{align*}
\]

**Proposition B.1.** Let \( \gamma > 0 \). For sufficiently small \( \varepsilon \), the free boundary problem (B.2)–(B.6) has a unique solution \((W, \zeta_-, \zeta_+)\), with \( \zeta_- < \zeta_+ \). The free boundaries have the asymptotic expansion

\[
\zeta_\pm = \frac{\pi_*}{1 - \pi_*} \pm \left( \frac{3}{4\gamma} \right)^{1/3} \left( \frac{\pi_*}{(1 - \pi_*)^2} \right)^{2/3} \varepsilon^{1/3} + \frac{5 - 2\gamma - \pi_*}{2\gamma(1 - \pi_*)^2} \left( \frac{\gamma \pi_* (1 - \pi_*)}{6} \right)^{1/3} \varepsilon^{2/3} + O(\varepsilon).
\]

**Proof of Proposition B.1.** Since \( \zeta_- \notin \{-1, 0\} \), any solution of the initial value problem (B.2)–(B.4) is of the form

\[
W(\zeta_-, \zeta) = \frac{2}{(\sigma \zeta)^2} \int_{\zeta_-}^\zeta (h(y) - h(\zeta_-)) \left( \frac{y}{\zeta} \right)^{2\gamma \pi_* - 2} dy.
\]

Suppose \((W, \zeta_-, \zeta_+)\) is a solution of (B.2)–(B.6). Due to (B.8) \( W(\cdot) \equiv W(\zeta_-, \cdot) \). Let

\[
J(\zeta_-, \zeta) := \frac{\sigma^2 \varepsilon^{2\gamma \pi_*}}{2} W(\zeta_-, \zeta).
\]
By the terminal conditions (B.5)–(B.6) at $\zeta_+$, and setting $\delta = \varepsilon^{1/3}$, $(\zeta_-, \zeta_+)$ satisfy the following system of non-linear equations,

$$
\Psi_1(\zeta_-, \zeta_+) := W(\zeta_-, \zeta_+) - \frac{\delta^3}{(1 + \zeta_+)(1 + (1 - \delta^3)\zeta_+)} = 0, \quad (B.10)
$$

$$
\Psi_2(\zeta_-, \zeta_+) := \frac{2(h(\zeta_+) - h(\zeta_-))}{\sigma^2\zeta^2_+} - \frac{2\gamma \pi_* W(\zeta_-, \zeta_+)}{\zeta_+} - \frac{(1 - \delta^3)^2}{(1 + (1 - \delta^3)\zeta_+)^2} + \frac{1}{(1 + \zeta_+)^2} = 0. \quad (B.11)
$$

Conversely, if $(\zeta_-, \zeta_+)$ solve (B.10)–(B.11), then the triplet $(W(\cdot; \zeta_-), \zeta_-, \zeta_+)$ provides a solution to the free boundary problem (B.2)–(B.6). Therefore, to provide a unique solution of the free boundary problem, it suffices to provide a unique solution of (B.10)–(B.11).

To obtain a guess for the asymptotic expansions of $\zeta_\pm$, we develop $\Psi_{1,2}$ around

$$
\zeta_\mp = \zeta_* + B_{1,2}\delta + O(\delta^2), \quad \zeta_* = \frac{\pi_*}{1 - \pi_*},
$$

which yields

$$
\Psi_1(\zeta_\pm(\delta)) = -\frac{\gamma(1 - \pi_*)^6}{3\pi_*^2} \left(2B_1^3 - (2B_1^3B_2 + B_2^3 + \frac{3\pi_*^2}{\gamma(1 - \pi_*)^4})\delta^3 + O(\delta^4)\right), \quad (B.12)
$$

$$
\Psi_2(\zeta_\pm(\delta)) = \frac{(B_1 - B_2)(B_1 + B_2)\gamma(\pi_* - 1)^6 \delta^2 + O(\delta^3)}{\pi_*^2}. \quad (B.13)
$$

By equating the coefficients of the leading order terms to zero, we have the system

$$
2B_1^3 - (2B_1^3B_2 + B_2^3 + \frac{3\pi_*^2}{\gamma(1 - \pi_*)^4}) = 0, \quad (B.14)
$$

$$
B_1 + B_2 = 0, \quad (B.15)
$$

which implies $B_1 = -B_2$ solves

$$
B_1^3 = -\frac{3}{4\gamma} \frac{\pi_*^2}{(1 - \pi_*)^4} = 0,
$$

an equation with a single, real-valued solution, namely

$$
B_1 = -\left(\frac{3}{4\gamma}\right)^{1/3} \frac{\pi_*}{\left(1 - \pi_*\right)^{2/3}}. \quad (B.16)
$$

We claim that for sufficiently small $\delta$ the system (B.10)–(B.11) has indeed a unique analytic solution around

$$
\zeta_{0,\mp} := \frac{\pi_*}{1 - \pi_*} \pm \left(\frac{3}{4\gamma}\right)^{1/3} \frac{\pi_*}{\left(1 - \pi_*\right)^{2/3}} \delta.
$$

This is equivalent to solving the corresponding system of equations $\Phi = (\Phi_1, \Phi_2) = 0$ for $(\eta_-, \eta_+)$, around $(B_1, B_2)$, where

$$
\eta_\pm := \frac{\zeta_\pm - \frac{\pi_*}{1 - \pi_*}}{\delta}
$$

and

$$
\Phi_1 := \frac{\Psi_1(\zeta_-(\eta_-), \zeta_+(\eta_+))}{\delta^3}, \quad \Phi_2 := \frac{\Psi_2(\zeta_-(\eta_-), \zeta_+(\eta_+))}{\delta^2}. \quad (B.17)
$$
Thus, hence the implicit function theorem for analytic functions (Gunning and Rossi, 2009, Theorem I.B.4) ensures that for sufficiently small $\varepsilon$. Hence, also the original system $\Psi(\zeta_-, \zeta_+) = 0$ has a unique solution $(\zeta_-, \zeta_+)$ around $\frac{\pi_*}{1-\pi_*}$. As a consequence, the free boundary problem (B.2)–(B.6) has a unique solution for sufficiently small $\varepsilon$.

To derive the higher order terms of (B.7), it is useful to rewrite the integral (B.9) as

$$J(\zeta_-, \zeta_+) = \frac{h(\zeta_-)(\zeta^2 - \zeta^2 - 1)}{2\gamma \pi_* - 1} + \int_{\zeta_-}^{\zeta_+} h(y)y^{2\gamma \pi_* - 2} dy.$$

(B.18)

The derivative of $I_2$ with respect to $\delta$ equals

$$\frac{dI_2}{d\delta} = h(\zeta_+)^{2\gamma \pi_* - 2} d\zeta_+ - h(\zeta_-)^{2\gamma \pi_* - 2} d\zeta_-,$$

(B.19)

and shall be expanded as a power series in $\delta$. Integration with respect to $\delta$ then yields an asymptotic expansion of $I_2$.

To obtain these expansions, we guess a solution of equations (B.10)–(B.11) of the form

$$\zeta_\pm = \frac{\pi_*}{1-\pi_*} \pm \left(\frac{3}{4\gamma}\right)^{1/3} \left(\frac{\pi_*}{1-\pi_*}\right)^{2/3} \delta + A_\pm \delta^2 + O(\delta^3),$$

for some unknowns $A_\pm$, and substitute it into equations (B.10)–(B.11), using thereby (B.18) and (B.19). Comparing the coefficients in the asymptotic expansion of the two equations reveals that

$$A_- = A_+ = \frac{(5 - 2\gamma)\pi_*}{2\gamma(1-\pi_*)^2} \left(\frac{\gamma \pi_*(1-\pi_*)}{6}\right)^{1/3},$$

and thus we have derived (B.7).

□

**Proposition B.2.** For sufficiently small $\delta > 0$, the system $\Phi(\eta_\pm(\delta), \delta) = 0$ defined by (B.17), has a unique solution $(\eta_-(\delta), \eta_+(\delta))$ near $(B_1, B_2)$, which is analytic in $\delta$.

**Proof.** Denote by $D\Phi$ the Frechet differential of $\Phi$. We claim that

$$\det(D\Phi)(\eta_- = B_1, \eta_+ = B_2, \delta = 0) = \frac{6\gamma(1-\pi_*)^8(2\gamma \pi_* - 1)}{\pi_*^2} \neq 0,$$

(B.20)

hence the implicit function theorem for analytic functions (Gunning and Rossi, 2009, Theorem I.B.4) ensures that for sufficiently small $\delta$ there exists a unique solution $(\eta_-, \eta_+)$ of $\Phi(\eta_-, \eta_+) = 0$ around $(B_1, B_2)$ which is analytic in $\delta$.

So it remains to verify the determinant formula in (B.20). To this end, note that, by construction,

$$\Psi_2(\zeta_-, \zeta_+) := \frac{\partial \Phi_1(\zeta_-, \zeta_+)}{\partial \zeta_+},$$

whence

$$\frac{\partial \Phi_1(\eta_-, \eta_+)}{\partial \eta_+}|_{(B_1, B_2, 0)} = 0.$$

Thus

$$\det(D\Phi)(B_1, B_2, 0) = \frac{\partial \Phi_1(\eta_-, \eta_+)}{\partial \eta_-}|_{(B_1, B_2, 0)} \times \frac{\partial \Phi_2(\eta_-, \eta_+)}{\partial \eta_+}|_{(B_1, B_2, 0)}.$$
We have
\[
\frac{\partial \psi_1}{\partial \zeta_-} = -\frac{2h'(-\zeta_-)}{\sigma^2\zeta_-^{2\mu/\sigma^2}} \left( \frac{\zeta_-^{2\mu/\sigma^2-2}}{2\mu/\sigma^2 - 1} - \frac{\zeta_-^{2\mu/\sigma^2-2}}{2\mu/\sigma^2 - 1} \right)
\]
and since by the chain rule
\[
\frac{\partial \Phi_1(\eta_-, \eta_+)}{\partial \eta_-} = 1 \frac{\partial \psi_1}{\partial \zeta_-} \times \delta
\]
we obtain
\[
\frac{\partial \Phi_1(\eta_-, \eta_+)}{\partial \eta_-} \bigg|_{(B_1,B_2,0)} = \frac{6^{2/3}(1 - \pi_*)^3(\gamma \pi_*(1 - \pi_*)^{1/3}(1 - 2\gamma \pi_*))}{\pi_*}.
\]
Similarly, we obtain
\[
\frac{\partial \Phi_2(\eta_-, \eta_+)}{\partial \eta_+} \bigg|_{(B_1,B_2,0)} = -\frac{6^{1/3}(1 - \pi_*)^4(\gamma(1 - \pi_*)\pi_*)^{2/3}}{\pi_*^2},
\]
from which (B.20), and hence the assertion, follows.

In the following, \(C^2(A)\) denotes the space of twice continuously differentiable functions on an open set \(A \subset \mathbb{R}\).

**Definition B.3.** A solution of the HJB equation is a pair \((V, \lambda)\), where \(V \in C^2\) and \(\lambda \in \mathbb{R}\), which satisfies
\[
\min(AV(x) - h(x) + \lambda, G(x) - V'(x), V'(x)) = 0, \quad x \in \left(-\infty, \frac{1}{1-\varepsilon}\right) \cup (0, \infty), \quad (B.21)
\]
where \(A : C^2(\mathbb{R}) \rightarrow C^2(\mathbb{R})\) is the differential operator
\[
Af(x) := \frac{\sigma^2}{2} x^2 f''(x) + \mu x f'(x).
\]

**Proposition B.4.** Let \((W, \zeta_-, \zeta_+)\) be the solution of the free boundary problem (B.5)–(B.6) (provided by Proposition B.1) with asymptotic expansion (B.7). For sufficiently small \(\varepsilon\), the pair
\[
V(\cdot) := \int_0^\cdot \hat{W}(\zeta)d\zeta, \quad \lambda := h(\zeta_-),
\]
where
\[
\hat{W}(\zeta) := \begin{cases} 
0 & \text{for } \zeta < \zeta_-,
W(\zeta) & \text{for } \zeta \in [\zeta_-, \zeta_+],
G(\zeta) & \text{for } \zeta \geq \zeta_+,
\end{cases} \quad (B.22)
\]
is a solution of the HJB equation (B.21).

**Proof of Proposition B.4.** To check that \((V, \lambda)\) solves the HJB equation (B.21), consider separately the domains \([\zeta_-, \zeta_+]\), \(\zeta < \zeta_-\) and \(\zeta > \zeta_+\). In the following we use the decompositions
\[
G(\zeta) = \frac{1}{1 + \zeta} - \frac{1 - \varepsilon}{1 + (1 - \varepsilon)\zeta} \quad \text{and} \quad G'(\zeta) = \left(\frac{1 - \varepsilon}{1 + (1 - \varepsilon)\zeta}\right)^2 - \frac{1}{(1 + \zeta)^2}.
\]

24
First, note that on \([\zeta_-, \zeta+]\), by construction it holds that
\[
(AV(\zeta) - h(\zeta) + h(\zeta_-))' = \frac{1}{2} \sigma^2 \zeta^2 W''(\zeta) + (\sigma^2 + \mu)\zeta W'(\zeta) + \mu W(\zeta) - H(\zeta) = 0.
\]
Furthermore, in view of the initial conditions (B.3)–(B.4),
\[
(AV(\zeta) - h(\zeta) + h(\zeta_-))|_{\zeta=\zeta_-} = AV(\zeta) |_{\zeta=\zeta_-} = 0,
\]
whence
\[
AV(\zeta) - h(\zeta) + h(\zeta_-) \equiv 0, \quad \zeta \in [\zeta_-, \zeta_+].
\]
To see that \(0 \leq V' \leq G\) on all of \([\zeta_-, \zeta_+]\), observe that
\[
(h(\zeta) - h(\zeta_-))' = h'(\zeta) = H(\zeta) = \frac{1}{\pi_s(1 + \zeta)^2} \left( \pi_s - \frac{\zeta}{1 + \zeta} \right). \tag{B.23}
\]
Note that for \(\zeta_- < \zeta \leq \zeta^*\), where \(\zeta^*/(1 + \zeta^*) = \pi_s\), \(V'(\zeta) = W(\zeta) > 0\). It is shown that also \(W(\cdot) \geq 0\) on all of \([\zeta_-, \zeta_+]\). This is equivalent to showing non-negativity of
\[
w(\zeta) := 2\sigma^2 \zeta^{2\gamma} W(\zeta) = \int_\zeta^{\zeta_+} (h(x) - h(\zeta_-)) \zeta^{2\gamma} dx. \tag{B.24}
\]
Now \(w'(\zeta) = (h(\zeta) - h(\zeta_-)) \zeta^{2\gamma - 2} = 0\) if and only if \(h(\zeta_-) = h(\zeta)\). Hence, either \(\zeta = \zeta_-\) or \(\zeta = \zeta := 2\pi_s - \zeta_-\). By the first-order asymptotics of (B.7), we thus obtain \(\zeta \notin [\zeta_-, \zeta_+]\) for sufficiently small \(\varepsilon\). Therefore \(w' > 0\) on \([\zeta_-, \zeta_+]\), and by (B.24) it follows that \(V' \geq 0\) on all of \([\zeta_-, \zeta_+]\). To conclude the validity of the HJB equation on \([\zeta_-, \zeta_+]\), it only remains to show the inequality \(V' \leq G\). To this end, notice that \(\Psi_1(\zeta) = W(\zeta) - G(\zeta)\), (this is the function defined in (B.10), with fixed \(\zeta_-\)) satisfies
\[
\Psi_1(\zeta_-) = -G(\zeta_-) = -\frac{\varepsilon}{(1 + \zeta_-)(1 + (1 - \varepsilon)\zeta_-)} = -(1 - \pi_s)^2 \varepsilon + O(\varepsilon^{4/3}),
\]
hence for sufficiently small \(\varepsilon\), \(\Psi_1(\zeta) < 0\) on some interval \([\zeta_-, \zeta]\), and \(\Psi_1(\zeta) = 0\). Therefore, \(\zeta \leq \zeta_+\). Since \(\Psi_1(\zeta_+) = 0\) by construction, it suffices to show that \(\zeta = \zeta_+\) to prove non-negativity of \(V'\) on \([\zeta_-, \zeta_+]\). Assume, for a contradiction, there exists a sequence \(\delta_k \downarrow 0\) such that for each \(k \in \mathbb{N}\) \(\Psi_1(\zeta(\delta_k)) = 0\), and that \(\zeta_- (\delta_k) \leq \zeta(\delta_k) < \zeta_+(\delta_k)\). We change to the variable \(u = \zeta - \zeta_-\), and introduce the notation \(u_+ = \frac{\zeta_+ - \zeta_-}{\delta}, \quad \bar{u} = \frac{\zeta - \zeta_-}{\delta}\). By selecting, if necessary, a subsequence, we may assume without loss of generality that \(\bar{u}(\delta_k)\) converges, hence it must satisfy
\[
\lim_{k \to \infty} \bar{u}(\delta_k) =: B_0 \in [B_1, B_2],
\]
where \(B_1\) is defined in (B.16), and \(B_2 = -B_1\). By the calculations leading to (B.16), we therefore conclude that \(B_0\) must satisfy (B.14) in place of \(B_2\), i.e.
\[
2B_1^3 - 3B_1^2 B_0 + B_0^3 + \frac{3\pi_s^2}{\gamma(1 - \pi_s)^4} = 0. \tag{B.25}
\]
With \(B_1\) from (B.16) and after the change of variable \(\xi = -B_0/B_1\), we obtain the equation
\[
2 - 3\xi + \xi^3 = 0
\]
which has the only solutions 1 and −2. Therefore, (B.25) has the only relevant solution

\[ B_0 = -B_1 = B_2. \]

By intertwining \( u_+ (\delta) \) and \( \bar{u} (\delta_k) \), we introduce

\[ \bar{u}^* (\delta) = \begin{cases} \bar{u} (\delta_k), & k \in \mathbb{N} \\ u_+ (\delta), & \text{otherwise} \end{cases} \]

Hence \((u_- (\delta), u^* (\delta))\) satisfies \( \Phi (u_- , u_+ ) = 0 \) near \((B_1 , B_2 )\), for sufficiently small \( \delta \). By Proposition B.2, \( u^* (\delta) = u_+ (\delta) \), which contradicts our assumption \( \tilde{\zeta} \neq \zeta_+ \).

Consider now \( \zeta \leq \zeta_- \). \( V \) solves the HJB equation, if

\[ AV - h(\zeta) + h(\zeta_-) = h(\zeta_-) - h(\zeta) \geq 0, \quad G(\zeta) \geq 0. \]

Since \( h(\zeta) - h(\zeta_-) \neq 0 \) for \( \zeta = \zeta_- \), to obtain the first inequality it suffices to show that (B.23) is non-negative. Now for small \( \varepsilon \) clearly \( \pi_- < \pi_+ \), hence for \( \zeta = \zeta_- \) (B.23) is indeed strictly positive. To settle the second inequality, recall that either \( \zeta < -1/(1 - \varepsilon) \) or \( \zeta > 0 \). On these domains, \( G \) is clearly a strictly positive function. Hence it is proved that \( V \) satisfies the HJB equation for \( \zeta \leq \zeta_- \).

Finally, consider \( \zeta \geq \zeta_+ \). Since \( G = W \), it suffices to show

\[ L(\zeta) := AV(\zeta) - h(\zeta) + h(\zeta_-) \geq 0, \quad G(\zeta) \geq 0. \]  \hfill (B.26)

The second estimate is straightforward: Let \( \zeta_+ > -1 \), then we have \( \zeta > -1 \), and for sufficiently small \( \varepsilon \), \( (1 - \varepsilon) \zeta > -1 \), hence \( G(\zeta) > 0 \). The case \( \zeta_+ < -1 \) can be dealt with similarly. For the first inequality in (B.26), note that

\[ L(\zeta) = \frac{\sigma^2 \zeta^2}{2} G'(\zeta) + \mu \zeta G(\zeta) - h(\zeta) + h(\zeta_-) \]

\[ = h(\zeta_-) - h_1 ((1 - \varepsilon) \zeta) + \frac{(\gamma - 1) \sigma^2}{2} \left( \frac{\zeta}{1 + \zeta} \right)^2. \]

Therefore, it suffices to show that

\[ \kappa(\zeta) := h(\zeta_-) - h_1 ((1 - \varepsilon) \zeta) - (1 - \gamma) \frac{\sigma^2}{2} \left( \frac{\zeta}{1 + \zeta} \right)^2 \]

has no zeros on \([\zeta_+, -1/(1 - \varepsilon)]\), besides \( \zeta_+ \).

The case \( \gamma = 1 \) is simple as \( \kappa(\zeta) = 0 \) can be reduced to solving a quadratic equation (see also Taksar, Klass and Assaf (1988)). All other cases require investigating a fourth-order polynomial, as seen below. However, to demonstrate the strength and clarity of the asymptotic approach of this paper, we first discuss the case \( \gamma = 1 \). The transformation \( z = \frac{\sqrt{\zeta}}{1 + \zeta} \) leads to

\[ \frac{(1 - \varepsilon) \zeta}{1 + (1 - \varepsilon) \zeta} = \frac{(1 - \varepsilon) z}{1 - \varepsilon z} \]

and thus one can rewrite (B.27) in terms of \( z \), denoting it by

\[ F(z, \varepsilon) = \mu \pi_- - \frac{\sigma^2}{2} \pi_-^2 - \mu \left( (1 - \varepsilon) \frac{z}{1 - \varepsilon z} \right) + \frac{\sigma^2}{2} \left( (1 - \varepsilon) \frac{z}{1 - \varepsilon z} \right)^2. \]
It is proved next that \( F \) has no zeros on \((\pi_+, 1/\epsilon)\). Since \( F(\pi_+) = 0 \), polynomial division by \((z - \pi_+)\) yields

\[
F(z, \epsilon) = \frac{(z - \pi_+)}{(1 - \epsilon z)^2} g(z),
\]

(B.28)

\( g(z) \) is a linear factor, and the following asymptotic expansions hold

\[
g(\pi_+) = \sigma^2 \left( \frac{3}{4\gamma} \left( \frac{\mu}{\sigma^2} \right) \left( 1 - \left( \frac{\mu}{\sigma^2} \right) \right)^2 \right)^{1/3} \epsilon^{1/3} + O(\epsilon^{2/3}),
\]

\[
g(1/\epsilon) = \frac{\sigma^2}{2\epsilon} + O(1).
\]

It follows that \( g \) has no zeros on \([\pi_+, 1/\epsilon]\), for sufficiently small \( \epsilon \). Hence \( F(z) > 0 \) for \( z \in (\pi_+, 1/\epsilon) \).

For the remainder of the proof, suppose \( \gamma \neq 1 \). Using the transformation \( z = \frac{\epsilon}{1+\epsilon} \), one can rewrite, similarly as in the \( \gamma = 1 \) case, function (B.27) in terms of \( z \),

\[
F(z, \epsilon) = \mu \pi_- - \frac{\gamma \sigma^2}{2} \pi_-^2 - \mu \left( \frac{(1 - \epsilon)z}{1 - \epsilon z} \right) + \frac{\sigma^2}{2} \left( \frac{(1 - \epsilon)z}{1 - \epsilon z} \right)^2 - (1 - \gamma) \frac{\sigma^2}{2} z^2.
\]

It is proved next that \( F \) has no zeros on \((\pi_+, 1/\epsilon)\).

Since \( F(\pi_+) = 0 \), polynomial division by \((z - \pi_+)\) yields (B.28), where the third order polynomial \( g \) has derivative

\[
g' = a_0 + a_1 z + a_2 z^2,
\]

with

\[
a_0 = \pi_- \mu \epsilon^2 - \frac{1}{2} \epsilon^2 - \frac{\pi_-^2}{2} \sigma^2 \epsilon^2 - \frac{\pi_-^2}{2} \sigma^2 (1 - \gamma) \epsilon^2
\]

\[
+ \pi_+ \sigma^2 (1 - \gamma) \epsilon - \mu \epsilon^2 + \frac{\sigma^2 (\gamma + \epsilon^2)}{2} + \mu \epsilon - \sigma^2 \epsilon,
\]

\[
a_1 = 2 \left( \sigma^2 (1 - \gamma) \epsilon - \frac{1}{2} \pi_+ \sigma^2 (1 - \gamma) \epsilon^2 \right),
\]

\[
a_2 = -\frac{3}{2} \sigma^2 (1 - \gamma) \epsilon^2.
\]

In view of (B.28), it is enough to show that \( g \) has no zeros on \([\pi_+, 1/\epsilon]\). First, note the following asymptotic expansions,

\[
g(\pi_+) = \gamma \sigma^2 \left( \frac{3}{4\gamma} \pi_-^2 (1 - \pi_+)^2 \right)^{1/3} \epsilon^{1/3} + O(\epsilon^{2/3}),
\]

(B.30)

\[
g(1/\epsilon) = \frac{\sigma^2}{2\epsilon} + O(1).
\]

(B.31)

Therefore, for sufficiently small \( \epsilon \), \( g > 0 \) on both endpoints of \([\pi_+, 1/\epsilon]\). It remains to show that any local minimum of \( g \) in \([\pi_+, 1/\epsilon]\) is non-negative. In searching for local extrema, one obtains complex numbers \( z_\pm \) where \( g'(z_\pm) = 0 \). The asymptotic expansions of \( z_\pm \) are given by

\[
z_\pm = \frac{2}{3\epsilon} \pm \frac{1}{3\epsilon} \sqrt{\frac{\gamma - 4}{\gamma - 1}} + O(1).
\]
Obviously, there are no local extrema in $[\pi_+, 1/\varepsilon]$ whenever $\gamma \in [1, 4)$. Therefore $g > 0$ on all of $[\pi_+, 1/\varepsilon]$, and thus $F(z) \geq 0$ on $[\pi_+, 1/\varepsilon]$. The non-trivial case $\gamma \notin [1, 4)$ remains:

For $0 < \gamma < 1$ it holds that $\frac{4\gamma}{\sqrt{1+4\gamma^2}} > 2$, hence $z_\pm \notin [\pi_+, 1/\varepsilon]$. It follows that $g'$ has no zeros in this interval and thus $g > 0$ on $[\pi_+, 1/\varepsilon]$.

Next, consider $\gamma \geq 4$: The local minimum $z_-$ of a third order polynomial with negative leading coefficient satisfies $z_- < z_+$ and $g(z_-) < g(z_+)$. In view of (B.30) and (B.31), it remains to show $g(z_-) > 0$. It holds that

$$
g(z_-) = \frac{(-\sqrt{\gamma^2 - 5\gamma + 4} + 2\gamma - 2) \left(\sqrt{\gamma^2 - 5\gamma + 4} + 2\gamma + 2\right) \sigma^2}{27(\gamma - 1)\varepsilon} + O(1)
$$

whence $g(z_-) > 0$ for sufficiently small $\varepsilon$. We have shown that $g > 0$ on $[\pi_+, 1/\varepsilon]$.

Summarizing, $\kappa(\zeta) \geq 0$ on $\zeta \geq \zeta_+$, which proves that the HJB equation (B.21) holds.

\[\square\]

**Lemma B.5.** Let $\eta_- < \eta_+$ be such that either $\eta_+ < -1/(1 - \varepsilon)$ or $\eta_- > 0$. Then there exists an admissible trading strategy $\hat{\phi}$ such that the risky/safe ratio $\eta_t$ satisfies SDE (A.3). Moreover, $(\eta_t, \hat{\phi}_t, \hat{\phi}_t^\dagger)$ is a reflected diffusion on the interval $[\eta_-, \eta_+]$. In particular, $\eta_t$ has stationary density equals

$$
\nu(\eta) := \frac{2\mu}{\sqrt{2\mu - 1}} \frac{1}{\eta_t^{2\mu - 1} - \eta_t^{2\mu - 2}} \eta_t^{2\mu - 2}, \quad \eta \in [\eta_-, \eta_+],
$$

(B.32)

when $\eta_- > 0$, and otherwise equals

$$
\nu(\eta) := \frac{2\mu}{\sqrt{2\mu - 1}} \frac{1}{|\eta_+|^{2\mu - 1} - |\eta_-|^{2\mu - 1}} \eta_t^{2\mu - 2}, \quad \eta \in [\eta_-, \eta_+].
$$

(B.33)

**Proof.** By the solution of the Skorohod problem for two reflecting boundaries Kruk et al. (2007), there exists a well-defined reflected diffusion $(\eta_t, L_t, U_t)$ satisfying

$$
\frac{d\eta_t}{\eta_t} = \mu dt + \sigma dB_t + dL_t - dU_t,
$$

where $W$ is a standard Brownian motion, and $L$ (resp. $U$) is a non-decreasing processes which increases only on the set $\{\eta = \eta_-\}$ (resp. $\{\eta = \eta_+\}$). Also, $\eta_- > 0$ or $\eta_+ < -1/(1 - \varepsilon)$ implies that $\eta_t > 0$ or $\eta_t < -1/(1 - \varepsilon)$ for all $t$, almost surely. Hence for each $t > 0$,

$$
(1 + (1 - \varepsilon)\eta_t), \quad (1 + \eta_t)
$$

are invertible, almost surely. Define the increasing processes $(\hat{\phi}_t^\dagger, \hat{\phi}_t)$ by

$$
\frac{d\hat{\phi}_t^\dagger}{\hat{\phi}_t} = (1 + \eta_t)^{-1} dL_t
$$

and

$$
\frac{d\hat{\phi}_t}{\hat{\phi}_t} = (1 + (1 - \varepsilon)\eta_t)^{-1} dU_t.
$$

28
By construction, the associated measures $d\hat{\varphi}^\uparrow, d\hat{\varphi}^\downarrow$ are supported on $\eta_l = \eta_-$ and $\eta_r = \eta_+$, respectively. Hence $\hat{\varphi}$ is a trading strategy, which by Lemma A.3 yields a risky/safe satisfying precisely the stochastic differential equation (A.3).

The admissibility of the trading strategy is clear, as $\hat{\varphi}$ is a continuous, finite variation trading strategy, and since it satisfies $\pi_+ < 1/\epsilon$, which implies that there exists $\epsilon' > \epsilon$ such that $\pi_t < 1/\epsilon'$, for all $t > 0$, a.s.. Finally, the form of the stationary density $\nu(\eta)$, follows from the stationary Fokker-Planck equation: The infinitesimal generator of $\zeta_t$ is given by

$$A^f(\zeta) = \frac{\sigma^2}{2} \zeta^2 f''(\zeta) + \mu \zeta f'(\zeta) =: a(\zeta)/2\varphi''(\zeta) + b(\zeta)\varphi'(\zeta).$$

The invariant density $\nu$ solves the adjoint differential equation

$$A^*\nu(\eta) = (a(\eta)\nu(\eta))' - 2b(\eta)\nu(\eta) = 0$$

and therefore equals

$$\nu(\eta) = \frac{c}{a(\eta)} \exp \left( \int \frac{2b(\eta)}{a(\eta)} d\eta \right), \quad (B.34)$$

where the constant $c > 0$ depends on the boundaries $\zeta_-, \zeta_+$. By integration, and distinguishing the cases $\eta_+ < 0$ or $\eta_- > 0$, we obtain the explicity expressions (B.32) and (B.33).

The following constitutes the verification of optimality of the trading strategy of Lemma B.5 with the trading boundaries given by Proposition B.1:

**Proposition B.6.** Let $\zeta_\pm$ be the free boundaries as derived in Proposition B.1, and denote by $\hat{\varphi}$ the trading strategy of Lemma B.5 associated with these free boundaries. Set

$$\pi_\pm := \zeta_\pm/(1 + \zeta_\pm).$$

Then for all $t > 0$, the fraction of wealth $\pi_t$ invested in the risky asset lies in the interval $[\pi_- , \pi_+]$, almost surely, entails no trading whenever $\pi \in (\pi_- , \pi_+)$ (the no-trade region) and engages in trading only at the boundaries $\pi_\pm$. For sufficiently small $\epsilon$, $\hat{\varphi}$ is optimal, and the value function is given by

$$F_\infty(\hat{\varphi}) = \max_{\varphi \in \Phi} \lim_{T \to \infty} \frac{1}{T} \mathbb{E}[R_T^2 - \gamma/2(R_0^\varphi)_T] = r + h(\zeta_-) = r + \mu \pi_- - \frac{\gamma \sigma^2}{2} \pi_-^2. \quad (B.35)$$

**Proof of Proposition B.6.** Recall from Proposition B.4 that $\lambda = h(\zeta_-)$ and $(V, \lambda)$, defined from the unique solution of the free boundary problem, is a solution of the HJB equation (B.21). For the verification, we use the proportion $\pi_t$ of wealth in the risky asset instead of the risky/safe ratio $\zeta_t$. The change of variables

$$\zeta = -1 + \frac{1}{1 - \pi}$$

amounts to a compactification of the real line, such that the two intervals $[-\infty, -1/(1 - \epsilon)]$ and $(0, \infty]$ are mapped onto the connected interval $[0, 1/\epsilon)$. Denote by $\mathcal{L}$ the differential operator

$$\mathcal{L}f(\pi) := \frac{\sigma^2}{2} f''(\pi)\pi^2(1 - \pi)^2 + f'(\pi)(\mu - \sigma^2\pi)\pi(1 - \pi).$$
The function $\mathcal{V}(\pi) := V(\zeta(\pi))$ satisfies the HJB equation

$$\min(L\mathcal{V}(\pi) - \dot{h}(\pi) + \lambda, \dot{\mathcal{V}}(\pi) - \frac{\varepsilon}{1 - \varepsilon^{\pi}}) = 0,$$

(B.36)

for $0 \leq \pi < 1/\varepsilon$, where $\dot{h}(\pi) = h(\zeta(\pi)) = \mu\pi - \frac{\gamma\sigma^{2}}{2} \pi^{2}$.

First, it is shown that $F_{\infty}(\varphi) \leq \lambda + r$, for any admissible trading strategy $\varphi$. By Lemma A.2 we may without loss of generality assume $\pi_{t} \geq 0$, almost surely, for all $t \geq 0$. An application of Itô’s formula to the stochastic process $\mathcal{V}(\pi_{t})$, where $\mathcal{V}$ is the solution of the HJB equation (B.36), yields

$$\mathcal{V}(\pi_{T}) - \mathcal{V}(\pi_{0}) = \int_{0}^{T} \dot{\mathcal{V}}'(\pi_{t})d\pi_{t} + \frac{1}{2} \mathcal{V}''(\pi_{t})d(\pi_{t})$$

(B.37)

$$= \int_{0}^{T} (\mathcal{L}\mathcal{V}(\pi) - \dot{h}(\pi_{t}) + \lambda) dt + \int_{0}^{T} (\dot{h}(\pi_{t}) - \lambda)dt$$

(B.38)

$$+ \int_{0}^{T} \dot{\mathcal{V}}'(\pi_{t})\pi_{t}(1 - \pi_{t})\sigma dB_{t}$$

(B.39)

$$- \int_{0}^{T} \dot{\mathcal{V}}'(\pi_{t})(1 - \varepsilon \pi_{t})\pi_{t}\frac{d\varphi_{t}^{\uparrow}}{\varphi_{t}}$$

(B.40)

$$+ \int_{0}^{T} \dot{\mathcal{V}}'(\pi_{t})\pi_{t}\frac{d\varphi_{t}^{\uparrow}}{\varphi_{t}}.$$

(B.41)

The first term in line (B.38) is non-negative, due to (B.36). Furthermore, (A.1) implies the existence of $\varepsilon' > \varepsilon$ such that $\pi_{t} < 1/\varepsilon' < 1/\varepsilon$, for all $t$, a.s., and using (B.36) we have

$$\dot{\mathcal{V}}'(\pi_{t}) \leq \frac{\varepsilon\varepsilon'}{\varepsilon' - \varepsilon}, \quad \text{a.s. for all } t \geq 0.$$  

(B.42)

Hence (B.39) is a martingale with zero expectation. Again by (B.36) one has that

$$\dot{\mathcal{V}}'(\pi_{t})\pi_{t}(1 - \varepsilon \pi_{t}) \leq \varepsilon \pi_{t},$$

which implies that for (B.40) we have

$$- \int_{0}^{T} \dot{\mathcal{V}}'(\pi_{t})(1 - \varepsilon \pi_{t})\pi_{t}\frac{d\varphi_{t}^{\uparrow}}{\varphi_{t}} \geq -\varepsilon \int_{0}^{T} \pi_{t}\frac{d\varphi_{t}^{\uparrow}}{\varphi_{t}}.$$

Finally, (B.41) is non-negative, because $\dot{\mathcal{V}}' \geq 0$ due to (B.36). Thus, taking the expectation of (B.37) we obtain the estimate,

$$\frac{1}{T}E[\mathcal{V}(\pi_{T}) - \mathcal{V}(\pi_{0})] \geq -\lambda + \frac{1}{T}E[\int_{0}^{T} \dot{h}(\pi_{t})dt] - \varepsilon \frac{1}{T} \int_{0}^{T} \pi_{t}\frac{d\varphi_{t}^{\uparrow}}{\varphi_{t}}.$$

(B.43)

By (B.42)

$$|\mathcal{V}(\pi_{t}) - \mathcal{V}(\pi_{0})| \leq |\pi_{T} - \pi_{0}| \sup_{0 < u \leq 1/\varepsilon'} |\dot{\mathcal{V}}'(u)| \leq \frac{\varepsilon}{\varepsilon' \varepsilon},$$

30
therefore
\[
\lim_{T \to \infty} \frac{1}{T} \mathbb{E}[\bar{V}(\pi_T) - \hat{V}(\pi_0)] = 0.
\]

Hence letting \( T \to \infty \) in (B.43) reveals that for any admissible strategy
\[
F_\infty(\varphi) \leq \lambda + r. \tag{B.44}
\]

Finally, it is shown that the bound \( \lambda + r \) is attained by the admissible trading strategy \( \hat{\varphi} \) defined by Lemma (B.5) in terms of the free boundaries \((\zeta_-, \zeta_+)\). Let \( \zeta_t \) be the corresponding risky/safe ratio. Using Itô’s formula, one has
\[
dV(\zeta_t) = V'(\zeta_t)\zeta_t \sigma dB_t + 0 - \varepsilon \pi_t \frac{d\varphi_t}{\varphi_t} + (h(\zeta_t) - \lambda)dt.
\]

Integration with respect to \( t \) and division by \( T \) yields, in view of (A.7),
\[
\frac{1}{T} \mathbb{E}[R_T^\hat{\varphi} - \frac{\gamma}{2} (R_T^\hat{\varphi})^2] = r + \lambda + \frac{1}{T} \mathbb{E}[\bar{V}(\pi_T) - \hat{V}(\pi_0)].
\]

Letting \( T \to \infty \) we have \( F_\infty(\hat{\varphi}) = \lambda + r. \) Due to (B.44), \( \hat{\varphi} \) is an optimal trading strategy.

\[\square\]

### B.1 Proof of Theorem 3.1 (i)–(iii)

Theorem 3.1 (i) is proved in Proposition B.1, and Theorem 3.1 (ii) & (iii) are proved in Proposition B.6.

### C Ergodic results

In this section, ergodic theorems are applied to derive closed-form expressions for average trading costs (ATC) and long-run mean and long-run variance of the optimal trading strategy. These formulas are then used to prove the asymptotic expansions of Theorem 3.1 (iv).

Let \( \zeta_-, \zeta_+ \) be the free boundaries obtained in Proposition B.1. Without loss of generality, assume that either \( \zeta_- < \zeta_+ < -1 \) (levered case) or \( \zeta_- > \zeta_+ > 0 \) throughout (non-levered case), and define the integral
\[
I := \frac{1}{c} \int_{\zeta_-}^{\zeta_+} h(\zeta) |\zeta|^{2\gamma \pi_s - 2} d\zeta, \tag{C.1}
\]
where the normalizing constant is given by
\[
c := \int_{\zeta_-}^{\zeta_+} |\zeta|^{2\gamma \pi_s - 2} d\zeta = \text{sgn}(\zeta_-) \frac{|\zeta_+|^{2\gamma \pi_s - 1} - |\zeta_-|^{2\gamma \pi_s - 1}}{2\gamma \pi_s - 1}.
\]

The objective functional

**Lemma C.1.**
\[
I = h(\zeta_-) + \frac{\sigma^2 (2\gamma \pi_s - 1)}{2} \left( \frac{G(\zeta_+) \zeta_+}{1 - \left( \frac{\zeta_-}{\zeta_+} \right)^{2\gamma \pi_s - 1}} \right). \tag{C.2}
\]
Proof. From equations (B.8) and (B.10) it follows that
\[
\int_{\zeta_-}^{\zeta_+} h(\zeta) |\zeta|^{2\gamma \pi_s - 2} d\zeta = h(\zeta_-) \frac{|\zeta_+|^{2\gamma \pi_s - 1} - |\zeta_-|^{2\gamma \pi_s - 1}}{2\gamma \pi_s - 1} + \frac{\sigma^2 |\zeta_+|^{2\gamma \pi_s}}{2} G(\zeta_+).
\]
By normalizing, (C.2) follows. \qed

Let now \( \zeta \) be a geometric Brownian motion with parameters \((\mu, \sigma)\), reflected at \( \zeta_-, \zeta_+ \) respectively, as in Lemma B.5. Recall the following ergodic result, of (Gerhold et al., 2014, Lemma C.1):

**Lemma C.2.** Let \( \eta_t \) be a diffusion on an interval \([l, u]\), reflected at the boundaries, i.e.
\[
d\eta_t = b(\eta_t)dt + a(\eta_t)^{1/2}dB_t + dL_t - dU_t,
\]
where the mappings \( a(\eta) > 0 \) and \( b(\eta) \) are both continuous, and the continuous, non-decreasing processes \( L_t \) and \( U_t \) satisfy \( L_0 = U_0 = 0 \) and increase only on \( \{L_t = l\} \) and \( \{U_t = u\} \), respectively. Denoting by \( \nu(\eta) \) the invariant density of \( \eta_t \), the following almost sure limits hold:
\[
\lim_{T \to \infty} \frac{L_T}{T} = \frac{a(l)\nu(l)}{2}, \quad \lim_{T \to \infty} \frac{U_T}{T} = \frac{a(u)\nu(u)}{2}.
\]

The next formula evaluates trading costs.

**Lemma C.3.** The average trading costs for the optimal trading policy are
\[
ATC := \lim_{T \to \infty} \frac{1}{T} \int_0^T \pi_t d\varphi_t^\perp = \frac{\sigma^2 (2\gamma \pi_s - 1)}{2} \left( \frac{G(\zeta_+)\zeta_+}{1 - \left(\frac{\zeta_-}{\zeta_+}\right)^{2\gamma \pi_s - 1}} \right). \tag{C.3}
\]

Proof. Apply Lemma C.2 to the stochastic process \( \eta := \zeta, u = \zeta_+ \). Realizing that
\[
\varepsilon \int_0^T \pi_t d\varphi_t^\perp = G(\zeta_+) \frac{U_T}{T}
\]
and using the stationary density of \( \zeta_t \), Lemma B.5, which equals
\[
\nu(\zeta) := \text{sgn}(\zeta_-) \frac{2\gamma \pi_s - 1}{|\zeta_+|^{2\gamma \pi_s - 1} - |\zeta_-|^{2\gamma \pi_s - 1}} |\zeta|^{2\gamma \pi_s - 2}, \quad \zeta \in [\zeta_-, \zeta_+],
\]
and applying Lemma C.2, we obtain (C.3). \qed

**Remark C.4.** An alternative proof provides a consistency check for the theory provided so far: By Lemma 2.1 one can rewrite the objective functional as
\[
F(\varphi) = r + \lim_{T \to \infty} \frac{1}{T} \int_0^T h(\zeta) dt - ATC.
\]
Now by the ergodic theorem (Borodin and Salminen, 2002, II.35 and II.36),
\[
\lim_{T \to \infty} \frac{1}{T} \int_0^T h(\zeta) dt = I,
\]
hence using Lemma C.1 it follows that
\[
F(\varphi) = r + h(\zeta_-) + ATC - ATC = r + h(\zeta_-)
\]
which is in agreement with the formula in Proposition B.6.
C.1 Long-run mean and variance

Define

\[ I_{\mu} := \int_{\zeta_-}^{\zeta_+} \left( \frac{\zeta}{1 + \zeta} \right) |\zeta|^{2\gamma\pi_s - 2} d\zeta, \quad I_{s^2} := \int_{\zeta_-}^{\zeta_+} \left( \frac{\zeta}{1 + \zeta} \right)^2 |\zeta|^{2\gamma\pi_s - 2} d\zeta. \]

Since the long-run mean and long-run variance are given by

\[ \hat{m} := \lim_{T \to \infty} \frac{1}{T} E\left[ R_T \right] = r + \mu \lim_{T \to \infty} \frac{1}{T} E\left[ \int_0^T \pi_t dt \right] - ATC \]

\[ = r + \frac{\mu}{c} I_{\mu} - ATC, \]

\[ \hat{s}^2 := \lim_{T \to \infty} \frac{1}{T} E\left[ (R)_T \right] = \sigma^2 \lim_{T \to \infty} \frac{1}{T} E\left[ \int_0^T \pi_t^2 dt \right] \]

\[ = \frac{\sigma^2}{c} I_{s^2}, \]

the following decomposition holds in view of the ergodic theorem (Borodin and Salminen, 2002, II.35 and II.36):

\[ I = \frac{1}{c} \left( \mu I_{\mu} - \frac{\gamma \sigma^2}{2} I_{s^2} \right) = (\hat{m} - r + ATC) - \frac{\gamma \sigma^2}{2} \]

\[ = h(\zeta_-) + ATC. \quad (C.4) \]

Integration by parts yields

\[ I_{\mu} = \int_{\zeta_-}^{\zeta_+} \frac{\zeta}{1 + \zeta} |\zeta|^{2\gamma\pi_s - 2} d\zeta = \frac{|\zeta|^{2\gamma\pi_s}}{2\gamma\pi_s(1 + \zeta_+)} - \frac{|\zeta_-|^{2\gamma\pi_s}}{2\gamma\pi_s(1 + \zeta_-)} + I_{s^2} \frac{2\gamma\pi_s}{2\gamma\pi_s - 2}. \quad (C.5) \]

An application of this identity to (C.4) yields

\[ I = \frac{\sigma^2}{2c} \left( \frac{|\zeta_+|^{2\gamma\pi_s}}{1 + \zeta_+} - \frac{|\zeta_-|^{2\gamma\pi_s}}{1 + \zeta_-} + (1 - \gamma) I_{s^2} \right). \]

Whenever \( \gamma \neq 1 \), one may extract \( I_{s^2} \), and thus (C.5) and (C.3) yield a formula for \( \hat{\sigma}^2 \). Therefore, the right side of equation (C.4) gives a formula for \( \hat{m} \) in terms of \( \hat{s} \):

**Lemma C.5.** When \( \gamma \neq 1 \), the following identities hold:

\[ \hat{s}^2 = \frac{2}{1 - \gamma} \left( h(\zeta_-) + ATC - \frac{\sigma^2}{2c} \left( \frac{|\zeta_+|^{2\gamma\pi_s}}{1 + \zeta_+} - \frac{|\zeta_-|^{2\gamma\pi_s}}{1 + \zeta_-} \right) \right), \quad (C.6) \]

\[ \hat{m} = r + \frac{\gamma}{2} \hat{s}^2 + h(\zeta_-). \quad (C.7) \]

C.2 Proof of Theorem 3.1 (iv)

**Proof.** The asymptotic expansion (3.7) for the trading boundaries \( \pi_{\pm} \) can be derived by developing \( \frac{\zeta_{\pm}}{1 + \zeta_{\pm}} \) into a power series, thereby using the asymptotic expansions (B.7) of \( \zeta_{\pm} \).

Long-run mean \( \hat{m} \) and long-run variance \( \hat{s}^2 \), as well as average trading costs ATC and the value function \( \lambda \) have closed form expressions in terms of the free boundaries \( \zeta_- \), \( \zeta_+ \) (see equations (C.7), (C.6), and equations (C.3) and (B.35)). Using these formulas in combination with the asymptotic expansions (B.7) of the free boundaries, the assertion follows. \( \square \)
D Proof of Theorem 3.2

In this section the free boundary problem (3.1)–(3.5) for \( \gamma = 0 \) is solved for sufficiently small \( \varepsilon \), it is shown that the solution \((W, \zeta_-, \zeta_+)\) allows to construct a solution of the corresponding HJB equation, and similarly to the case \( \gamma > 0 \), a verification argument reveals an optimal trading strategy.

Numerical experiments using \( \gamma > 0 \) indicate that the trading boundaries \( \pi_\pm \) (hence the leverage multiplier) satisfy

\[
\lim_{\varepsilon \to 0} \varepsilon^{1/2} \pi_\pm = 1/A_\pm
\]

for two constants \( A_- > A_+ > 0 \). Thus, we expect that the free boundaries in terms of the risky/safe ratio obey for sufficiently small \( \varepsilon \) the approximation

\[
\zeta_\pm \approx -1 - A_\pm \varepsilon^{1/2}.
\]

This insight lets us conjecture that \( \zeta_\pm \) are analytic in \( \delta := \varepsilon^{1/2} \).

Using the free boundary problem, the system (B.10)–(B.11) is rewritten with \( \delta := \varepsilon^{1/2} \) and the second equation is multiplied by \( \delta \).

\[
W(\zeta_-, \zeta_+) - \frac{\delta^2}{(1 + \zeta_+)(1 + (1 - \delta^2)\zeta_+)} = 0, \quad (D.1)
\]

\[
delta \left( \frac{2(h(\zeta_+ - h(\zeta_-)) - 2\mu/\sigma^2}{\sigma^2 \zeta_+^2} W(\zeta_-, \zeta_+) - \frac{(1 - \delta^2)^2}{(1 + (1 - \delta^2)\zeta_+)^2} + \frac{1}{(1 + \zeta_+)^2} \right) = 0. \quad (D.2)
\]

Using the transformation \( u = \frac{-1 - \xi}{\delta} \) and noting that \( |\zeta| = 1 + \delta u \), one obtains

\[
\Xi(u_-, u) := W(-1 - u_- \delta, -1 - u_- \delta) = \frac{2\mu}{\sigma^2 (1 + u_\delta)^2} \int_{u_-}^{u} \left( \frac{1}{u_- - \xi} \right) \left( \frac{1 + \xi \delta}{1 + u_\delta} \right)^{2\mu/\sigma^2 - 2} d\xi.
\]

Accordingly, the system (D.1)–(D.2) transforms into

\[
\Xi(u_-, u_+) - \frac{1}{u_+ ((1 - \delta^2)u_+ - \delta)} = 0, \quad (D.3)
\]

\[
\frac{2\mu}{\sigma^2} \left( \frac{1}{u_+} - \frac{1}{u_-} + \frac{\delta}{1 + u_\delta} \Xi(u_-, u_+) \right) - \frac{2}{u_+^2} \frac{(1 - \delta^2)u_+ - \delta}{u_+^2 (\delta + (\delta^2 - 1)u_+)^2} = 0. \quad (D.4)
\]

Letting \( \delta \to 0 \) in (D.3)–(D.4), we get an equation for, say \((A_-, A_+)\),

\[
\frac{2\mu}{\sigma^2} \left( \log(A_-/A_+) - \frac{A_- - A_+}{A_-} \right) - \frac{1}{A_+^2} = 0, \quad (D.5)
\]

\[
\frac{\mu}{\sigma^2} \left( \frac{1}{A_+} - \frac{1}{A_-} \right) - \frac{1}{A_+^2} = 0. \quad (D.6)
\]

Lemma D.1. There is a unique solution \((A_-, A_+)\) of system (D.5)–(D.6), given by

\[
A_- = \kappa^{-1/2} \sqrt{\frac{\sigma^2}{\mu}}, \quad A_+ = \kappa^{1/2} \sqrt{\frac{\sigma^2}{\mu}},
\]

\[
34
\]
where \( \kappa \approx 0.5828 \) is the unique solution of

\[
f(\xi) := \frac{3}{2} \xi + \log(1 - \xi) = 0, \quad \xi \in (0, 1).
\]

**Proof.** Equation (D.6) gives

\[
A_+ = \frac{\mu A_+^3}{\mu A_+^2 - \sigma^2}.
\]

Hence substituting (D.9) into (D.5) gives the well-posed transcendental equation

\[
-\frac{3}{A_+^2} + \frac{2\mu \log \left( \frac{\mu A_+^2}{\mu A_+^2 - \sigma^2} \right)}{\sigma^2} = 0, \quad A_+ > 0.
\]

Therefore it is enough to establish that (D.10) has a unique solution given by the second formula in (D.7); the formula for \( A_- \) then follows from (D.9). To this end, substitute

\[
\xi := \frac{\sigma^2}{\mu A_+^2}
\]

into (D.10) to obtain equation (D.8). Note that \( f(0) = 0, f' > 0 \) on \((0, 1/3)\) and \( f' < 0 \) on \((1/3, 1)\), while \( f(\xi) \downarrow -\infty \) as \( \xi \to 1 \). This implies that \( f \) has a single zero \( \kappa \) on \((1/3, 1)\).

**Proposition D.2.** For sufficiently small \( \delta \), there exists a unique solution \((u_+, u_-)\) of (D.3)–(D.4) near \((A_-, A_+)\). This solution is analytic in \( \delta \) and satisfies the asymptotic expansion \( u_\pm = A_\pm + O(\delta) \), where \( A_\pm \) are given by (D.7).

**Proof.** Denote the left sides of (D.3)–(D.4), by \( F_i((u_-, u_+), \delta), i = 1, 2 \) and \( F = (F_1, F_2) \). By Lemma D.1, \( F((A_-, A_+), 0) = 0 \). Since

\[
\frac{\partial F_1}{\partial u_-}((A_-, A_+), 0) = \frac{2\mu}{\sigma^2} \left( \frac{A_- - A_+}{A_-^2} \right), \quad \frac{\partial F_2}{\partial u_+}((A_-, A_+), 0) = \frac{2\mu}{\sigma^2} \left( \frac{A_+ - A_-}{A_- A_+} \right),
\]

one obtains

\[
\frac{\partial F_1}{\partial u_-}((A_-, A_+), 0) = \frac{2\mu}{\sigma^2} \left( \frac{A_- - A_+}{A_-^2} \right),
\]

\[
\frac{\partial F_1}{\partial u_+}((A_-, A_+), 0) = \frac{2\mu}{\sigma^2} \left( \frac{A_+ - A_-}{A_- A_+} \right) = 0,
\]

\[
\frac{\partial F_2}{\partial u_+}((A_-, A_+), 0) = \frac{6}{A_+} - \frac{2\mu}{\sigma^2} \left( \frac{1}{A_+^2} \right),
\]

where the second line vanishes due to (D.6), and therefore, the Jacobian \( DF \) of \( F \) satisfies

\[
\det(DF)((A_-, A_+), 0) = \frac{\partial F_1}{\partial u_-}((A_-, A_+), 0) \times \frac{\partial F_2}{\partial u_+}((A_-, A_+), 0)
\]

\[
= -4(\mu/\sigma^2)^{7/2}(\kappa - 1)\kappa^{5/2}(3\kappa - 1) \neq 0,
\]

because \( \kappa \in (1/3, 1) \). Hence by the implicit function theorem for analytic functions (Gunning and Rossi, 2009, Theorem I.B.4) the assertion follows. \( \square \)
Lemma D.3. Let \( \kappa \) be the solution of (D.8), and \( \theta \in [0,1] \). If
\[
f(\theta) = \log(1 - \kappa(1 - \theta)) + (1 - \theta)\kappa + \frac{1}{2} \frac{\kappa(1 - \kappa)^2}{(1 - \kappa(1 - \theta))^2} = 0 \tag{D.11}
\]
then \( \theta = 0 \).

Proof. Clearly \( f(0) = 0 \) and also \( f(1) = 1/2\kappa(1 - \kappa)^2 > 0 \). There is a single local extremum of \( f \), in \((0,1)\), namely,
\[
\theta_1 = \frac{0.5 \left( 3\kappa^2 + \sqrt{4\kappa^3 - 3\kappa^4 - 2\kappa} \right)}{\kappa^2} \approx 0.7669,
\]
but since \( f'(0) = 0 \), and
\[
f''(0) = \frac{\kappa^2 (\kappa (3\kappa^2 - 7\kappa + 5) - 1)}{(1 - \kappa)^4} > 0
\]
\( \theta_1 \) must be the global maximum. Hence \( f > 0 \) on \((0,1]\). \( \square \)

Lemma D.4. Let \( A_- \) be as in (D.7). The only solution of
\[
\frac{2\mu}{\sigma^2} \left( \log(A_-/\xi) - \frac{A_- - \xi}{A_-} \right) - \frac{1}{\xi^2} = 0 \tag{D.12}
\]
on \([A_+, A_-]\) is \( \xi = A_+ \).

Proof. Let \( \xi \) be a solution of (D.12). There exists \( \theta \in [0,1] \) such that
\[
\xi = \theta A_- + (1 - \theta)A_+ = A_+ \left( \frac{1 + \kappa(1 - \theta)}{1 - \kappa} \right).
\]
Hence \( A_+^* / A_- = 1 + \kappa(\theta - 1) \), and therefore (D.12) can be rewritten as (D.11). An application of Lemma D.3 yields \( \xi = A_+ \). \( \square \)

Proof of Theorem 3.2

Proof. Arguing similarly as in the Proof of Proposition B.1 for the case \( \gamma > 0 \), the solvability of the free boundary problem (3.1)–(3.5) for \( \gamma = 0 \) is equivalent to solvability of the non-linear system (D.1)–(D.2). This, in turn, is equivalent to solving (D.3)–(D.4) for \( (u_+(\delta), u_-(\delta)) \). A unique solutions of the transformed system (D.3)–(D.4) near \((A_+, A_-)\) is provided by Proposition D.2, and one has \( \zeta_\pm = -1 - u_\pm \delta \). In particular, one obtains
\[
\zeta_\pm = -1 - A_\pm \varepsilon^{1/2} + O(1). \tag{D.13}
\]
The solution of (3.1)–(3.5) is given by
\[
W(\zeta) := \frac{2\mu}{\sigma^2|\zeta|^2} \int_{\zeta_-}^\zeta \left( \frac{y}{1 + y} - \frac{\zeta_-}{1 + \zeta_-} \right) |y|^{2\mu/\sigma^2 - 1}dy. \tag{D.14}
\]
One defines exactly as in (B.22) a candidate \((V, \lambda)\) for the HJB equation (B.21). Next it is shown that \((V, \lambda)\) solves the HJB equation (B.21) (for the intervals \([\zeta_-, \zeta_+], (-\infty, \zeta_-] \) and finally for \([\zeta_+, \infty)\)). In fact, the interval \([-1/(1 - \varepsilon), 0]\) is excluded.

36
On $[\zeta_-,\zeta_+]$, we have
\[
(AV(\zeta) - h(\zeta) + h(\zeta_-))' = \frac{1}{2} \sigma^2 \zeta^2 W''(\zeta) + (\sigma^2 + \mu) \zeta W''(\zeta) + \mu W(\zeta) - \frac{\mu}{(1 + \zeta)^2} = 0
\]
by construction. Furthermore, because of the initial conditions (3.2)–(3.3)
\[
(AV(\zeta) - h(\zeta) + h(\zeta_-)) |_{\zeta = \zeta_-} = AV(\zeta) |_{\zeta = \zeta_-} = 0
\]
and therefore
\[
AV(\zeta) - h(\zeta) + h(\zeta_-) \equiv 0, \quad \zeta \in [\zeta_-, \zeta_+].
\]
Next it is shown that $0 \leq V' \leq G$ on all of $[\zeta_-, \zeta_+]$. Since
\[
(h(\zeta) - h(\zeta_-))' = h'(\zeta) = \frac{\mu}{(1 + \zeta)^2}
\]
is strictly positive, $h(\zeta) - h(\zeta_-) > 0$ for $\zeta \in (\zeta_-, \zeta_+).$. From the explicit formula (D.14) one therefore may conclude that $V' = W \geq 0$ for $\zeta \in [\zeta_-, \zeta_+].$. It remains to show $V' \leq G$. Since $V'(\zeta_+) - G(\zeta_+) = 0$, and since $V'(\zeta_-) - G(\zeta_-) = -G(\zeta_-) < 0$, it suffices to rule out any zero $\zeta^+_k$ of $V'(\zeta) - G(\zeta)$ on $(\zeta_-, \zeta_+)$, for sufficiently small $\varepsilon$. This is equivalent to ruling out any zeros of
\[
\kappa(u, \delta) := V'(\zeta(u)) - G(\zeta(u)), \quad u \in (u_+(\delta), u_-(\delta)),
\]
where $\zeta(u) = -1 - u\delta$, for sufficiently small $\delta$. Recall that $u_+(\delta)$ is implicitly defined by $\zeta_+ = -1 - u_+(\delta)\delta$, and we have $\lim_{\delta \to 0} u_+(\delta) = A_+$. Assume, for a contradiction, there exists $\delta_k \downarrow 0$ and a sequence $u_+(\delta_k)$ satisfying $u_-(\delta_k) < u_+(\delta_k) < u_+(\delta_k)$ which is a solution of $\kappa(u_+(\delta_k), \delta_k) = 0$ for each $k \in \mathbb{N}$. By taking a subsequence, if necessary, we may without loss of generality assume $u_+(\delta_k) \to A^*_+ \in [A_+, A_-]$ as $k \to \infty$. We distinguish two cases now.

- $A^*_+ = A_+$: We define the map $\delta \mapsto u^*_+(\delta)$ by intertwining $u_+$ and $u^*_+$ as follows:
\[
u^*_+(\delta) = \begin{cases} 
u^*_+(\delta_k), & k \in \mathbb{N} \\ u_+(\delta), & \delta \neq \delta_k \end{cases}.
\]
Then for sufficiently small $\delta$, the pair $(u_-(\delta), u^*_+(\delta))$ solves (D.3)–(D.4) near $(A_-, A_+)$, hence by Proposition D.2, $u^*_+ = u_+$, a contradiction to our previous assumption $\zeta^*_+ \in (\zeta_-, \zeta_+)$.

- $A^*_+ \in (A_+, A_-)$: By equation (D.3) we have
\[
\frac{2\mu}{\sigma^2} \left( \log(A_-/A^*_+) - \frac{A_- - A^*_+}{A_-} \right) - \frac{1}{(A^*_+)^2} = 0.
\]
Lemma D.4 states $A^*_+ = A_+$, which is impossible.

Therefore $V$ solves the HJB equation on $[\zeta_-, \zeta_+]$.
Consider now $\zeta \leq \zeta_-$. $V$ solves the HJB equation, if
\[
AV - h(\zeta) + h(\zeta_-) = h(\zeta_-) - h(\zeta) \geq 0, \quad G(\zeta) \geq 0.
\]
The first inequality is clearly fulfilled. Also, since $\zeta < -1/(1 - \varepsilon)$ or $\zeta > 0$, $G$ is a strictly positive function on $[-\infty, \zeta_-]$, which finishes the proof for $\zeta \leq \zeta_-$. 37
Finally, consider $\zeta \geq \zeta_+$. Since $G = W$, it suffices to show that

$$L(\zeta) := AV(\zeta) - h(\zeta) + h(\zeta_-) \geq 0, \quad G(\zeta) \geq 0. \tag{D.16}$$

The second inequality has just been proved. So only the first inequality in (D.16) needs to be shown. Denoting $h_1(\zeta)$ the function $h$ for $\gamma = 1$, one obtains

$$L(\zeta) = \frac{\sigma^2 \zeta^2}{2} G'(\zeta) + \mu \zeta G(\zeta) - h(\zeta) + h(\zeta_-)$$

$$= h(\zeta_-) - h_1((1 - \varepsilon)\zeta) + \frac{(\gamma - 1)\sigma^2}{2} \left( \frac{\zeta}{1 + \zeta} \right)^2.$$

Therefore, by the boundary conditions at $\zeta_+$,

$$L(\zeta_+) = \frac{\sigma^2 \zeta^2}{2} W'(\zeta_+) + \mu \zeta W(\zeta_+) + h(\zeta_-) - h(\zeta_+).$$

The last equality follows from our knowledge concerning the HJB equation on $[\zeta_-, \zeta_+]$.

To show that $L(\zeta) \geq 0$ for all $\zeta$, it suffices to show that there are no solutions of the equation

$$\kappa(\zeta := h(\zeta_-) - h_1((1 - \varepsilon)\zeta) - \frac{\sigma^2}{2} \left( \frac{\zeta}{1 + \zeta} \right)^2 = 0 \tag{D.17}$$

on $\zeta \geq \zeta_+$ except $\zeta_+$. The transformation $z = \frac{\zeta}{1 + \zeta}$ yields

$$\frac{(1 - \varepsilon)\zeta}{1 + (1 - \varepsilon)\zeta} = \frac{(1 - \varepsilon)z}{1 - \varepsilon z}$$

and thus one can rewrite (D.17) in terms of $z$, denoting it by

$$F(z, \varepsilon) = \mu - \mu \left( \frac{(1 - \varepsilon)z}{1 - \varepsilon z} \right) + \frac{\sigma^2}{2} \left( \frac{(1 - \varepsilon)z}{1 - \varepsilon z} \right)^2 - \frac{\sigma^2 z^2}{2}.$$

It is proved next that $F$ has no zeros on $(\pi_+, 1/\varepsilon)$: Since $F(\pi_+) = 0$, polynomial division by $(z - \pi_+)$ yields

$$F(z, \varepsilon) = \frac{(z - \pi_+)}{(1 - \varepsilon z)^2} g(z), \tag{D.18}$$

where the third order polynomial $g$ has derivative

$$g' = a_0 + a_1 z + a_2 z^2,$$

with coefficients given by (B.29) (setting $\gamma = 0$). By the second formula of (D.7)

$$g(\pi_+) = -\mu + \frac{3\sigma^2}{A^2_+} + O(\varepsilon^{1/2}) \tag{D.19}$$

is strictly positive for sufficiently small $\varepsilon$, since $\kappa > 1/3$. The solutions $z_\pm$ of the equation

$$g'(z) = 0$$
are
\[ z_- = -\frac{1}{2A_+\varepsilon^{1/2}} + O(1) \]
which is negative for sufficiently small \( \varepsilon \), hence irrelevant, and
\[ z_+ = \frac{4}{3\varepsilon} + O(1), \]
which is larger than \( 1/\varepsilon \), hence also irrelevant. Since
\[ g'(1/\varepsilon) = \sigma^2/2 + O(\varepsilon^{1/2}) \]
it follows that \( g'(z) > 0 \) on all of \( [\pi_+, 1/\varepsilon] \). Together with (D.19) it follows that \( g > 0 \) on \( [\pi_+, 1/\varepsilon] \). Hence \( F(z) > 0 \) for all \( z > \pi_+ \) which admit the conclusion that \( (V, \lambda) \) solves the HJB equation (B.21).

Using the proof of Proposition B.6, one can obtain assertion (ii) and (iii). Finally, the expansions of the trading boundaries claimed in (iv) follow from the asymptotic expansions of the free boundaries \( \zeta_- , \zeta_+ \) in (D.13).

**D.1 Proof of Proposition 4.1**

*Proof of Proposition 4.1.* While the first formula holds in view of (4.9), the second one is a result of the asymptotic expansions provided by Theorem 3.1 (iv).

**E Convergence**

**Lemma E.1.** Let \( \mu > \sigma^2 \). There exists \( \delta_0 > 0 \) such that for all \( 0 \leq \gamma < \gamma_0 := \frac{\mu}{\sigma^2}, \delta \leq \delta_0 \) the objective functional for a trading strategy \( \varphi \) which only engages in buying at \( \pi_- = 1 + \delta \) and selling at \( \pi_+ = (1 - \delta)/\varepsilon \) outperforms a buy and hold strategy. More precisely, for all \( \gamma < \gamma_0, \delta \leq \delta_0 \)
\[ F_\infty(\varphi) \geq r + \mu - \frac{\gamma\sigma^2}{2} + \left( \frac{\mu - \gamma\sigma^2}{2} \right) \delta > r + \mu - \frac{\gamma\sigma^2}{2}. \]

*Proof.* Using the stationary density \( \rho(d\pi) \) of \( \pi_t \) on \( [\pi_-, \pi_+] \) (which can be derived from Lemma B.5), one obtains
\[ F(\varphi) = r + \int_{\pi_-}^{\pi_+} \left( \mu\pi - \frac{\gamma\sigma^2}{2}\pi^2 \right) \rho(d\pi) - ATC \]
\[ \geq r + \mu(1 + \delta) - \frac{\gamma\sigma^2}{2}(1 + \delta)^2 - \frac{(\delta + 1)(2\varepsilon - 1)^3(2\mu - \sigma^2)}{4\varepsilon \left( \delta \left( \frac{-2(\delta+1)\varepsilon+\delta+1}{\delta} \right) \frac{2\mu}{\sigma^2} + (\delta + 1)(2\varepsilon - 1) \right)} \]
\[ \geq r + \mu - \frac{\gamma\sigma^2}{2} + (\mu - \gamma\sigma^2)\delta - O(\delta^{\min(2, \frac{2\mu}{\sigma^2} - 1)}). \] (E.1)
where Lemma C.3 has been invoked to calculate and estimate the average trading costs ATC. The asymptotic expansion holds for sufficiently small \( \delta \). Since \( \mu > \gamma\sigma^2 \), the claim follows. \( \square \)
E.1 Proof of Theorem 4.2

Proof. By equation (4.4), the curves \((0, \gamma) \to \mathbb{R} : \gamma \mapsto \pi_\pm(\gamma)\) range in a relatively compact set, namely \([1, \frac{1}{\varepsilon}]\). Consider therefore a sequence \(\gamma_k, k = 1, 2, \ldots\) which satisfies

\[
1 \leq \pi_0 := \lim_{i \to \infty} \pi_-(\gamma_k) \leq \lim_{i \to \infty} \pi_+(\gamma_k) =: \pi_+ = 1/\varepsilon.
\]

Define accordingly \(\zeta_0 \). Then \(-\infty \leq \zeta_0^0 \leq \zeta_0^+ \leq \frac{1}{1-\varepsilon}\). First, three elementary facts are proved:

(i) \(\pi_0^+ > 1\), which is equivalent to \(\zeta_0^+ > -\infty\). Assume, for a contradiction, \(\pi_0^+ = 1\). Then \(\pi_-^k(\gamma_k) \to 1\) and thus \(\lambda(\gamma_k) \to r + \mu, \) as \(k \to \infty\). Hence, the objective functional eventually minorizes the uniform bound provided by Lemma E.1, a mere impossibility. We conclude that \(\pi_0^- > 1\).

(ii) \(\pi_0^- < \pi_0^+\): This holds due to the fact that, by observing limits for the initial and terminal conditions of zero order in (3.1),

\[
W(\zeta^0_0) = 0 < G(\zeta^0_0).
\]

(iii) Also, \(\pi_0^+ < \frac{1}{\varepsilon}\). Assume, for a contradiction, that \(\pi_0^+ = \frac{1}{\varepsilon}\). Then \(G(\zeta_0^+(\gamma_k)) \to \infty\), as \(k \to \infty\), and, since \(\zeta_0^- < \zeta_0^0\), the average trading costs corresponding to \(\gamma_k\) satisfy (by Lemma C.3)

\[
ATC(k) := \frac{\sigma^2 \left( \frac{2\mu}{\sigma^2} - 1 \right)}{2 - \frac{\mu}{\sigma^2 - 1}},
\]

as \(k \to \infty\). Denote by \(\varphi^k\) the trading strategy which only buys (resp. sells) at \(\pi_-(\gamma_k)\) (resp. \(\pi_+(\gamma_k)\)). By the results of Appendix C the value function satisfies for each \(k\)

\[
F_\infty(\varphi^k) = r + \int_{\pi_-(\gamma_k)}^{\pi_+(\gamma_k)} (\mu \pi - \frac{\gamma \sigma^2}{2} \pi^2) \rho(d\pi) - ATC(k) \leq r + \frac{\mu}{\varepsilon} - ATC(k) \to -\infty,
\]

as \(k \to \infty\). In particular, for sufficiently large \(k \geq k_0\), a buy-and-hold strategy \(\varphi\) satisfies

\[
F_\infty(\varphi) = r + \mu - \frac{\gamma k \sigma^2}{2} > F_\infty(\varphi^k),
\]

which contradicts the assumption concerning optimality of the trading strategy \([\pi_-(\gamma_k), \pi_+(\gamma_k)]\). Hence \(\pi_0^+ < 1/\varepsilon\).

Since the sequence \(\zeta_-(\gamma_k)\) converges, by (Keller-Ressel et al., 2010, Lemma 9) the solutions of the initial value problem associated with (3.1) and \(\gamma_k\), namely \(W(\zeta; \zeta_-(\gamma_k))\), converges to the solution of the initial value problem (3.1) (where \(\gamma = 0\)),

\[
W^0(\zeta) = -\frac{2}{\sigma^2 \zeta^2} \int_{\zeta_0^-}^{\zeta} (\mu - \frac{\zeta}{1 + \zeta} - \mu \frac{\zeta_0^-}{1 + \zeta_0^-}) \zeta' \zeta^{-2} d\zeta.
\]

The terminal conditions are met by \(W^0\), because \(G\) is continuous on \((-\infty, \frac{1}{1-\varepsilon})\). Also, for each \(k, k = 1, 2, \ldots\), by assumption the HJB equation (B.21) is satisfied. Non-negativity is preserved.
by taking limits, hence, \((\hat{W}(\gamma;0),\lambda(0))\) satisfies the HJB equation as well. Using the verification arguments of the proof of Proposition B.6 it follows that the trading strategies associated with the intervals \([\pi_-(\gamma),\pi_+(\gamma)]\) are optimal for risk-aversion levels \(\gamma \in [0,\bar{\gamma}]\), but also \([\pi_0^-,\pi_0^+]\) is optimal for a risk-neutral investor.

\(\zeta_-(\gamma)\) can have only one accumulation point for \(\gamma \downarrow 0\), because \(\lambda^0\) is the value function. Uniqueness of \(\zeta^0\) is therefore clear and it follows that \(\zeta^0 = \zeta_-(0)\). By assumption, the free boundary problem has a unique solution, hence it follows that \(\pi_{\pm}(0) = \pi_{\pm}^0\). In particular, the curves \((0,\bar{\gamma}) \to \mathbb{R} : \gamma \mapsto \pi_{\pm}(\gamma)\) each have a unique limit \(\pi_{\pm}^0\) as \(\gamma \downarrow 0\), which equals \(\pi_{\pm}(0)\), the solution of the free boundary problem.

References


