



## Leveraged funds: robust replication and performance evaluation

Paolo Guasoni & Eberhard Mayerhofer

To cite this article: Paolo Guasoni & Eberhard Mayerhofer (2023) Leveraged funds: robust replication and performance evaluation, Quantitative Finance, 23:7-8, 1155-1176, DOI: 10.1080/14697688.2023.2222753

To link to this article: <https://doi.org/10.1080/14697688.2023.2222753>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 27 Jun 2023.



Submit your article to this journal [↗](#)



Article views: 1235



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 3 View citing articles [↗](#)



This article has been awarded the Centre for Open Science 'Open Data' badge.



This article has been awarded the Centre for Open Science 'Open Materials' badge.

# Leveraged funds: robust replication and performance evaluation

PAOLO GUASONI <sup>†‡\*</sup> and EBERHARD MAYERHOFER <sup>§</sup>

<sup>†</sup>School of Mathematical Sciences, Dublin City University, Dublin D09 W6Y4, Ireland

<sup>‡</sup>Dipartimento di Statistica, Università di Bologna, Via Belle Arti 41, Bologna 40126, Italy

<sup>§</sup>Department of Mathematics and Statistics, University of Limerick, Limerick V94 T9PX, Ireland

(Received 11 January 2023; accepted 30 May 2023; published online 27 June 2023)

Leveraged and inverse exchange-traded funds seek daily returns equal to fixed multiples of indexes' returns, but the ensuing rebalancing costs create a tension between a high correlation with the index and a low average deviation from the leveraged index' performance. With proportional trading costs, we find that the optimal replication policy is robust to the index' dynamics and obtain a sufficient statistic for index replication performance, the *implied spread*, which is insensitive to risk-premia and enables comparisons of funds tracking different factors of an index. Overall, the impact of trading costs on replication performance is comparable to or higher than the effect of management fees.

**Keywords:** Transaction costs; Performance evaluation; Tracking error; Leverage; ETFs

**MSC(2010):** 91G10, 91G80

**JEL classifications:** G11, G23

## 1. Introduction

Leveraged and inverse exchange-traded funds (LIETFs) aim to deliver daily returns that scale an index' return by a constant factor. Introduced in the United States in 2006, the 12 funds available at the end of that year barely held \$2 billion. Since then, new funds have become available for equity indexes specific to countries and industry sectors, bonds, commodities, currencies, and real estate. At the end of 2022, over 240 leveraged or inverse funds held over \$65 billion in assets, almost all of them managed by two firms, Profunds and Direxion.<sup>†</sup>

Typical factors for these funds range from  $-3$  for inverse funds to  $+3$  for leveraged ones, which may lead to large losses in turbulent markets, thereby raising regulators' concerns—and doubts. In 2015, an SEC proposal sought to cap leverage to 150%, but was not implemented. In 2017, the SEC initially approved the listing of funds with factors of  $\pm 4$ , only to later reconsider its decision.<sup>‡</sup> Overall, LIETFs appear to have grown faster than their research.

Despite the ostensible simplicity of LIETFs, even the determinants of their returns are not fully understood. For example, since inception in 2006 to the end of 2022, the ProShares Ultra S&P500 (SSO), which doubles the daily return of the S&P 500 index, had a cumulative return of 396%. However, compounding the daily returns of its benchmark strategy would imply a gross (i.e. before management fees) theoretical return of 594%. Net of such fees, the return would have been 486%, which still leaves an unexplained gap of 1.01% per year.

Such deviations of LIETF returns from their benchmarks are economically significant, and of the same order of magnitude as management fees. They are also systematic, and call for sharper analytical tools to understand and evaluate the performance of these funds, which this paper aims to provide.

LIETFs represent shares of ownership of an asset pool, managed so that its exposure to the underlying index matches approximately the target factor of the fund. In practice, for leveraged funds on indexes with liquid constituents such as the S&P 500, a fraction of the exposure is achieved by investing in the index' components. The remainder of the exposure—for inverse funds, all of the exposure—is obtained through over-the-counter index swap contracts collateralized by the fund's assets, the bulk of which is in cash instruments. The funds' holdings and net asset value are published daily.

\*Corresponding author. Email: [paolo.guasoni@dcu.ie](mailto:paolo.guasoni@dcu.ie)

<sup>†</sup> These figures exclude other leveraged and inverse exchange-traded products such as exchange-traded notes, which have significantly different characteristics.

<sup>‡</sup> See <https://www.sec.gov/news/pressrelease/2015-276.html> and <https://www.wsj.com/articles/sec-reconsidering-staff-approval-of-first-quadruple-leveraged-etf-1494970106>.

The mechanism that keeps the shares of such funds aligned with their net asset value is the activity of authorized participants (typically market makers and broker dealers), who exchange blocks of funds' shares (creation units) for the underlying basket of securities at their net asset value (minus a transaction fee of approximately 0.10%). As a result, authorized participants quickly exploit arbitrage opportunities created by minimal departures of the market price of the fund's shares from its net asset value, thereby ensuring that the former closely tracks the latter.

Keeping a constant leverage ratio requires rebalancing the portfolio daily, as to keep the amount of index exposure near the target multiple of the fund's value: both leveraged and inverse funds buy the index when it increases and sell it when it decreases (Cheng and Madhavan 2009). Rebalancing requires a significant amount of trading: for example, a typical 1% daily return (one standard deviation, for an annual volatility of 16%) implies a daily turnover of 2% for a fund with factor +2 or −1, 6% with factors +3 or −2, and a remarkable 12% with factor −3.<sup>†</sup> Such trading volume generates substantial trading costs, especially for funds with high leverage ratios and for indexes with high volatility and low liquidity. While funds managers seek to mitigate trading frictions using swaps and derivatives, the question is to what extent the trading costs generated by frequent rebalancing are responsible for the observed underperformance of LIETFs.

The tradeoff between correlation and underperformance underlies the two main metrics used to evaluate passive investment funds, which focus on the difference between the index' multiple and the fund's daily returns. The tracking error (TrE), the annualized standard deviation of such difference, reflects the precision with which the fund mimics its target at a daily frequency. The tracking difference (TrD), the annualized average of such difference, measures the systematic gap between the benchmark index and the fund. Tracking error and tracking difference are the main indicators used by academics (Charteris and McCullough 2020), practitioners (Frino and Gallagher 2001, Johnson *et al.* 2013), and regulators (ESMA 2014) to evaluate the performance of ETFs.<sup>‡</sup>

This tension raises two central problems for managers and investors. Fund managers need to trade as efficiently as possible to achieve the desired tracking error while minimizing the deviation from the leveraged index. Investors compare competing funds that differ in leverage ratios, tracking error, and tracking difference, and need performance measures that control for such differences. The contribution of this paper is to answer these questions in a model with arbitrary volatility dynamics, continuous trading, and trading costs proportional to volume, leading to two main results for managers and investors.

A manager's optimal replication policy is independent of volatility, while it depends only on the trading cost  $\varepsilon$ , the target factor  $\Lambda$ , and the desired tradeoff between tracking error and tracking difference, summarized by a positive parameter  $\gamma$ , which captures investors' aversion to tracking error. The

manager should keep the fund's index exposure, as a multiple of the fund's assets, within the approximate thresholds

$$\Lambda \pm \left( \frac{3}{4\gamma} \Lambda^2 (\Lambda - 1)^2 \right)^{1/3} \varepsilon^{1/3}$$

by minimally increasing and decreasing the fund's exposure as it reaches the lower and upper levels, respectively. In this formula, higher values of the parameter  $\gamma$  lead to a narrower no-trade region, which results in lower tracking error, higher trading costs, and consequently a more negative tracking difference. The formula is similar to the optimal trading boundaries arising in problems of optimal portfolio choice with transaction costs, with one crucial difference: here the approximate policy is insensitive to the index price dynamics, both volatility and expected return.<sup>§</sup> Even as these boundaries remain constant, in more volatile times rebalancing is higher as the index position is more volatile, and boundaries reached more frequently.

The robust replication of leveraged funds has some analogies with the robust hedging of variance swaps, developed by Dupire (1993), Neuberger (1994), and Carr and Madan (2001): if the underlying index follows a continuous process, the optimal replication policy does not depend on the particular volatility dynamics, and in the absence of frictions the replication is perfect. In the present context, such robustness remains valid in the presence of trading frictions, though the optimal strategy's performance does depend on realized volatility.

The optimal replication policy under frictions explains the underexposure puzzle observed by Tang and Xu (2013), who report that average exposures of both leveraged and inverse funds in US markets are significantly smaller than their target factor. We confirm this effect in recent data on US funds and show that it is consistent with the model's predictions: though the optimal rebalancing boundaries are symmetric (at the first order) around the target factor, the fund's volatility is lower when its exposure is closer to zero, which increases the time spent at such smaller exposures and generates a bias towards zero for the average exposure. The effect is stronger for larger factors and more volatile indexes.

The main implication for investors is that a summary of a fund's overall tracking performance is the *implied spread*, defined as

$$\tilde{\varepsilon} := \frac{12}{\sqrt{3}} \frac{(-\text{TrD}) \cdot \text{TrE}}{\sigma^3 \Lambda^2 (1 - \Lambda)^2}. \quad (1)$$

This quantity estimates the hypothetical bid-ask spread that would make the investor indifferent between using the fund or replicating the index by trading with such a spread.

Formula (1) offers a tool for comparing the performance of funds with different factors and tracking errors: the implied spread combines the tracking difference (TrD) and tracking error (TrE) in a single number, measuring the fund's performance after controlling for the factor  $\Lambda$  and the index'

<sup>†</sup> A fund with factor  $\Lambda$  generates a daily turnover of  $\Lambda(\Lambda - 1)$  times the index return (Cheng and Madhavan 2009).

<sup>‡</sup> The website <https://www.trackingdifferences.com/> publishes the tracking differences of ETFs listed in several exchanges.

<sup>§</sup> See Magill and Constantinides (1976), Constantinides (1986), Taksar *et al.* (1988), Davis and Norman (1990), Dumas and Luciano (1991), and the asymptotic results in Shreve and Soner (1994), Rogers (2004), Gerhold *et al.* (2014), Kallsen and Muhle-Karbe (2017), and Melnyk and Seifried (2018).

average volatility  $\bar{\sigma}$ , which otherwise magnify the effect of trading frictions. The formula implies that, holding the factor constant, an optimally managed fund that seeks to halve its (negative) tracking difference should be prepared to approximately double its tracking error. In addition, the formula does not depend on the index' risk premium, which is notoriously hard to estimate with precision.

Most implied spreads in US funds range between 2 and 20 basis points, with smaller spreads for funds with larger factors, implying that more volatile funds offer a better tradeoff between tracking difference and tracking error. An explanation for this phenomenon lies in the management fees charged by leveraged funds, which are nearly the same for funds with different factors, making funds with larger factors comparatively cheaper in their unit cost of exposure.

These findings shed new light on the growing literature on leveraged funds. Tang and Xu (2013) report 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 leveraged returns, and a rebalancing component, due to trading frictions<sup>†</sup>. Wagalath (2014) derives an asymptotic expression for the *slippage* that results from rebalancing at fixed intervals. Anderson *et al.* (2012) note the sensitivity to trading frictions of risk-parity strategies, a popular class of leveraged strategies, and Anderson *et al.* (2014) propose a leveraged performance attribution formula that emphasizes, in addition to transaction costs, the comovement between portfolio exposure and index return. The recent work of Dai *et al.* (2022) studies the problem of continuous intraday replication of leveraged ETFs with quadratic transaction costs.

The paper is organized as follows: section 2 introduces the model and the optimization problem. Section 3 contains the main results on optimal leverage replication and performance evaluation (theorem 3.1). Section 4 describes how the replication of LIETFs takes place in practice, providing a quantitative analysis of the trading costs associated to rebalancing. With such context, section 5 investigates the paper's implications empirically, focusing on families of leveraged and inverse ETFs traded in US markets. Section 6 discusses the robustness of the paper's results to risk-premia proportional to return variance, finite horizons, and discrete trading. Section 7 concludes. All proofs are in the appendix.

## 2. Model

The market has a safe asset earning an interest rate  $r_t$  and a risky asset (the index) with ask (buying) price  $S_t$ , where

$$\frac{dS_t}{S_t} = (r_t + \mu_t) dt + \sigma_t dB_t, \quad S_0 \geq 0,$$

and with the index' bid (selling) price  $(1 - \varepsilon)S_t$ , which implies a constant relative bid-ask spread of  $\varepsilon > 0$ , or, equivalently, constant proportional transaction costs. Here  $(B_t)_{t \geq 0}$  is a standard Brownian motion on a filtered probability space

$(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathcal{F}, \mathbb{P})$ , while  $(r_t)_{t \geq 0}$  is an adapted process such that  $\int_0^T |r_t| dt < \infty$  a.s. Volatility is integrable and stationary:

**ASSUMPTION 2.1** *The process  $(\sigma_t)_{t \geq 0}$  satisfies  $\mathbb{E}[\int_0^T \sigma_t^2 dt] < \infty$  for all  $T > 0$  and is weakly ergodic with stationary variance  $\bar{\sigma}^2 > 0$ , i.e.  $\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \sigma_t^2 dt = \bar{\sigma}^2 \mathbb{P}$  - a.s.*

Thus the dynamics of the safe rate and volatility are essentially arbitrary, though volatility should neither explode nor vanish: as in Melnyk and Seifried (2018), this ergodicity condition ensures that average volatility is finite, hence the problem is well posed.

The baseline model in this section focuses on the case of a zero risk premium, i.e.  $\mu_t = 0$ , while section 6 extends the analysis to include a nonzero risk premium proportional to return variance and finds that it has only a second-order effect on the main results. In addition, the assumption of a zero risk premium is substantively appropriate for the optimal tracking problem at hand, as it removes a manager's incentive to generate positive tracking difference by systematically deviating from the target's exposure to earn a risk premium. Such hypothetical deviation would be counterfactual, as it would imply a positive bias for both leveraged and inverse funds. On the contrary, the 'underexposure puzzle' of Tang and Xu (2013) documents negative bias for leveraged funds and positive bias for inverse funds.

A manager's trading strategy is described by the number of shares  $\varphi_t$  of the index held at time  $t$ . The corresponding fund's value  $F_t = X_t + Y_t$  at time  $t$  is the sum of the index position  $Y_t = \varphi_t S_t$  and the safe position  $X_t$ , which follows the dynamics<sup>‡</sup>:

$$dX_t = r_t X_t dt - S_t d\varphi_t^\uparrow + (1 - \varepsilon) S_t d\varphi_t^\downarrow, \quad (2)$$

where  $\varphi_t^\uparrow$  and  $\varphi_t^\downarrow$  are the minimal increasing functions that satisfy  $\varphi_t = \varphi_t^\uparrow - \varphi_t^\downarrow$  and represent the cumulative number of shares purchased and sold, respectively. Furthermore, a strategy is required to be solvent, in that its corresponding wealth  $F_t$  is strictly positive at all times. (Admissible strategies are formally described in Definition A.1 below.)

Thus the fund value  $F_t$  satisfies the dynamics

$$\frac{dF_t}{F_t} - r_t dt = \pi_t \left( \frac{dS_t}{S_t} - r_t dt \right) - \varepsilon \pi_t \frac{d\varphi_t^\downarrow}{\varphi_t}, \quad (3)$$

where  $\pi_t = Y_t / (X_t + Y_t) = \varphi_t S_t / F_t$  represents the ratio of the index' position and the fund value at time  $t$ .

Absent any motive to systematically deviate from the target, the manager's objective is to trade off the fund's tracking error against its tracking difference, defined as follows. The (cumulative) difference  $D_t$  between the fund's return above the safe rate and the multiple of the index' return above the safe rate is defined as

$$dD_t = \frac{dF_t}{F_t} - r_t dt - \Lambda \left( \frac{dS_t}{S_t} - r_t dt \right), \quad D_0 = 0. \quad (4)$$

<sup>†</sup> Cf. Cheng and Madhavan (2009), Jarrow (2010), Avellaneda and Zhang (2010), and Lu *et al.* (2012).

<sup>‡</sup> The convention of evaluating the risky position at the ask price is inconsequential. Using the bid price instead leads to the same results up to a change in notation.



Accordingly, the definitions of annualized tracking difference and tracking error between the fund's and the index' multiple are

$$\begin{aligned} \text{TrD}_T &:= \frac{D_T}{T} = \lim_{\Delta t \rightarrow 0} \frac{1}{T} \sum_{k=1}^{\lfloor T/\Delta t \rfloor} (D_{k\Delta t} - D_{(k-1)\Delta t}) \\ &= \frac{1}{T} \int_0^T (\pi_t - \Lambda) \sigma_t dB_t - \frac{\varepsilon}{T} \int_0^T \pi_t \frac{d\varphi_t^\downarrow}{\varphi_t}, \end{aligned} \quad (5)$$

$$\begin{aligned} \text{TrE}_T &:= \left( \frac{\langle D \rangle_T}{T} \right)^{1/2} \\ &= \left( \lim_{\Delta t \rightarrow 0} \frac{1}{T} \sum_{k=1}^{\lfloor T/\Delta t \rfloor} (D_{k\Delta t} - D_{(k-1)\Delta t})^2 \right)^{1/2} \\ &= \left( \frac{1}{T} \int_0^T (\pi_t - \Lambda)^2 \sigma_t^2 dt \right)^{1/2}, \end{aligned} \quad (6)$$

where  $\langle D \rangle_T$  denotes the quadratic variation of the process  $D$ .<sup>†</sup>

The manager's objective is to find a trading policy that minimizes tracking error without deteriorating performance with excessive trading costs, i.e. for a fixed level of tracking difference. Such cost-adjusted tracking error is summarized by the quantity<sup>‡</sup>

$$\begin{aligned} \mathbb{E} \left[ -\text{TrD}_T + \frac{\gamma}{2} \text{TrE}_T^2 \right] \\ = \frac{1}{T} \mathbb{E} \left[ \frac{\gamma}{2} \int_0^T (\pi_t - \Lambda)^2 \sigma_t^2 dt + \varepsilon \int_0^T \pi_t \frac{d\varphi_t^\downarrow}{\varphi_t} \right]. \end{aligned} \quad (7)$$

The parameter  $\gamma > 0$  is interpreted as aversion to tracking error, and the empirical results in section 5 suggest that typical values of  $\gamma$  range between 5 and 10, as LIETFs aim at closely tracking the *daily* returns of their indexes, at the price of a moderate but significant average underperformance.

The final term in (7) represents trading costs, which hinder continuous portfolio rebalancing and make constant-proportion strategies infeasible. (Otherwise, in the absence of trading frictions the optimal choice is to set  $\pi_t$  constantly equal to  $\Lambda$ , which achieves perfect replication, with realized tracking difference and tracking error both zero.)

The appeal of the above objective is twofold: first, it makes quantitatively explicit the tradeoff between tracking error and trading costs, which is implicitly acknowledged in LIETF documents. For example, the prospectus of the fund Proshares Ultra S&P 500 states that 'The Fund seeks daily investment results, before fees and expenses, that correspond to two times (2x) the daily performance of the Index' before warning that 'A number of other factors may also adversely affect the Fund's correlation with the Index, including fees, expenses, transaction costs, financing costs [...]'.<sup>†</sup>

Second, the long-term average of (7) admits the interpretation of *equivalent expense ratio*, i.e. the hypothetical expense ratio that a frequent user of these funds would be willing to pay on the fund's assets to avoid both trading costs and

tracking error in the fund's payoff:<sup>§</sup>

$$\begin{aligned} \text{EER}(\varphi) &:= \\ \liminf_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \frac{\gamma}{2} \int_0^T (\pi_t - \Lambda)^2 \sigma_t^2 dt + \varepsilon \int_0^T \pi_t \frac{d\varphi_t^\downarrow}{\varphi_t} \right]. \end{aligned} \quad (8)$$

As leveraged and inverse funds are typically open-ended, without a specific maturity target, this paper focuses on the long-term objective (8).

Note that such ergodic average may be interpreted either as the long-term average that appears in the definition, or as the unconditional expectation of the daily cost-adjusted tracking error, given a stationary initial condition for the fund's composition at the beginning of the trading period.

Indeed, suppose that the investor could choose between the fund  $F_t$ , which tracks imperfectly the leveraged index, and another contract, which can be exercised anytime, and delivers the payoff  $\tilde{F}_t$ , defined as follows:

$$\frac{d\tilde{F}_t}{\tilde{F}_t} - r_t dt = \Lambda \left( \frac{dS_t}{S_t} - r_t dt \right) - \phi dt,$$

which does not entail any tracking error or trading costs, but rather a management fee  $\phi$ . For such a contract, the difference process is simply  $\tilde{D}_t = -\phi t$ , as trading costs and tracking error are zero, which means that the right-hand side of (8) is precisely  $\phi$ , justifying its interpretation as the equivalent expense ratio that would make an investor indifferent between the fund  $F_t$  with its tracking error and trading costs, and a hypothetical substitute that has neither, but rather pays such management fee.

The above comparison is more than a thought experiment:  $F_t$  reflects the characteristics of leveraged and inverse funds typical of US markets, while  $\tilde{F}_t$  mirrors the structure of factor certificates prevalent in Europe, derivatives contracts issued by a financial institution, which pay the holder the value of a leveraged index minus a management fee, rather than the value of a replicating portfolio.

### 3. Main results

The first main result characterizes the optimal policy and its performance in terms of the solution to a one-dimensional free-boundary problem that is independent of the volatility process.

<sup>§</sup> This quantity is well defined by virtue of assumption 2.1, which guarantees that tracking error does not diverge. The *lim inf* guarantees a good definition a priori. A posteriori, optimal strategies exist in which the limit inferior is a limit, hence the similar problem defined in terms of *lim inf* has the same solution. Note also that, if volatility  $\sigma$  is a function of some stationary Markov process, the first term in (8) is equivalent to  $\mathbb{E}[\frac{\gamma}{2}(\pi(Y) - \Lambda)^2 \sigma(Y)^2]$ , where the random variable  $Y$  has the stationary distribution under the expectation  $\mathbb{E}$ . In particular, the formulation in (7) in terms of a time-average is equivalent to a direct ergodic formulation of the problem. We keep the current formulation because it does not involve the Markov property and because an ergodic formulation of the trading cost in (8) would require the use of local times as well as further restrictions on trading strategies.

<sup>†</sup> Both limits hold in probability. The first convergence follows by the continuity of  $D$ , and the second one from theorem IV.1.3 in Revuz and Yor (1999) by localization.

<sup>‡</sup> The equality follows from lemma A.2, see also Remark A.3.

**THEOREM 3.1** Let  $\gamma > 0$  and  $\Lambda \neq 0, 1$ . For  $\varepsilon > 0$  small enough:

(i) *The free boundary problem*

$$\frac{1}{2}\zeta^2 W''(\zeta) + \zeta W'(\zeta) - \frac{\gamma}{(1+\zeta)^2} \left( \Lambda - \frac{\zeta}{1+\zeta} \right) = 0, \quad (9)$$

$$W(\zeta_-) = 0, \quad W'(\zeta_-) = 0, \quad (10)$$

$$W(\zeta_+) = \frac{\varepsilon}{(1+\zeta_+)(1+(1-\varepsilon)\zeta_+)},$$

$$W'(\zeta_+) = \frac{\varepsilon(\varepsilon - 2(1-\varepsilon)\zeta_+ - 2)}{(1+\zeta_+)^2(1+(1-\varepsilon)\zeta_+)^2} \quad (11)$$

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

(ii) *The trading strategy  $\hat{\varphi}$  that buys at  $\pi_- := \zeta_-/(1+\zeta_-)$  and sells at  $\pi_+ := \zeta_+/(1+\zeta_+)$  as little as necessary to keep the risky weight  $\pi_t$  within the interval  $[\pi_-, \pi_+]$ , maximizes (8).*

(iii) *The minimum Equivalent Expense Ratio is*

$$\min_{\varphi \in \Phi} \text{EER}(\varphi) = \frac{\gamma \bar{\sigma}^2}{2} (\pi_- - \Lambda)^2,$$

where  $\Phi$  is the set of admissible strategies (see appendix 1).

(iv) *The optimal Tracking Difference and Tracking Error are*

$$\text{TrD} := \lim_{T \rightarrow \infty} \frac{D_T}{T} = -\frac{\bar{\sigma}^2}{2} \frac{\pi_- \pi_+ (\pi_+ - 1)^2}{(\pi_+ - \pi_-)(1/\varepsilon - \pi_+)}, \quad (12)$$

$$\text{TrE} := \lim_{T \rightarrow \infty} \sqrt{\frac{\langle D \rangle_T}{T}} = \bar{\sigma} \sqrt{\pi_- \pi_+ + \Lambda(\Lambda - 2\bar{\beta})}, \quad (13)$$

where  $\bar{\beta}$  is defined in (14) below.

(v) *The average volatility  $\bar{s}$ , exposure  $\bar{\beta}$  and squared correlation  $R^2$  are:*

$$\bar{s} := \bar{\sigma} \sqrt{\pi_- \pi_+},$$

$$\bar{\beta} := \lim_{T \rightarrow \infty} \frac{\langle \int_0^T \frac{dF}{F}, \int_0^T \frac{dS}{S} \rangle_T}{\langle \int_0^T \frac{dS}{S} \rangle_T}$$

$$= \lim_{T \rightarrow \infty} \frac{\int_0^T \sigma_t^2 \pi_t dt}{\int_0^T \sigma_t^2 dt} = \pi_+ \pi_- \frac{\log(\pi_+/\pi_-)}{\pi_+ - \pi_-}, \quad (14)$$

$$R^2 := \lim_{T \rightarrow \infty} \frac{\langle \int_0^T \frac{dF}{F}, \int_0^T \frac{dS}{S} \rangle_T^2}{\langle \int_0^T \frac{dS}{S} \rangle_T^2 \langle \int_0^T \frac{dS}{S} \rangle_T^2}$$

$$= \lim_{T \rightarrow \infty} \frac{\left( \frac{1}{T} \int_0^T \sigma_t^2 \pi_t dt \right)^2}{\bar{\sigma}^2 \frac{1}{T} \int_0^T (\sigma_t \pi_t)^2 dt}$$

$$= \pi_- \pi_+ \left( \frac{\log(\pi_+/\pi_-)}{\pi_+ - \pi_-} \right)^2. \quad (15)$$

(vi) *The following asymptotic expansions hold:*

$$\pi_{\pm} = \Lambda \pm \left( \frac{3}{4\gamma} \Lambda^2 (\Lambda - 1)^2 \right)^{1/3} \varepsilon^{1/3}$$

$$- \frac{\Lambda}{\gamma} \left( \frac{\gamma \Lambda (\Lambda - 1)}{6} \right)^{1/3} \varepsilon^{2/3} + O(\varepsilon), \quad (16)$$

$$\text{EER} = \frac{\gamma \bar{\sigma}^2}{2} \left( \frac{3}{4\gamma} \Lambda^2 (\Lambda - 1)^2 \right)^{2/3} \varepsilon^{2/3} + O(\varepsilon),$$

$$\text{TrD} = -\frac{3\bar{\sigma}^2}{\gamma} \left( \frac{\gamma \Lambda (\Lambda - 1)}{6} \right)^{4/3} \varepsilon^{2/3} + O(\varepsilon), \quad (17)$$

$$\text{TrE} = \bar{\sigma} \sqrt{3} \left( \frac{\Lambda (\Lambda - 1)}{6\sqrt{\gamma}} \right)^{2/3} \varepsilon^{1/3} + O(\varepsilon), \quad (18)$$

$$\bar{s} = \bar{\sigma} \Lambda - \frac{\bar{\sigma}(7\Lambda - 3)}{4\gamma} \left( \frac{\gamma \Lambda (\Lambda - 1)}{6} \right)^{1/3} \varepsilon^{2/3} + O(\varepsilon),$$

$$\bar{\beta} = \Lambda - \frac{2\Lambda - 1}{\gamma} \left( \frac{\gamma \Lambda (\Lambda - 1)}{6} \right)^{1/3} \varepsilon^{2/3} + O(\varepsilon), \quad (19)$$

$$R^2 = 1 - \frac{1}{2\gamma} \left( 1 - \frac{1}{\Lambda} \right) \left( \frac{\gamma \Lambda (\Lambda - 1)}{6} \right)^{1/3} \varepsilon^{2/3} + O(\varepsilon). \quad (20)$$

In particular, the following relation holds:

$$\text{TrD} \cdot \text{TrE} = -\frac{\sqrt{3}}{12} \bar{\sigma}^3 \Lambda^2 (\Lambda - 1)^2 \varepsilon + O(\varepsilon^{4/3}). \quad (21)$$

*Proof* See appendix 4. ■

The main message of this theorem is that the solution to the optimal tracking problem is completely determined by the two buy and sell boundaries  $\pi_{\pm}$ , in terms of which all performance statistics follow in closed form. As volatility dynamics does not enter the free boundary problem,  $\pi_{\pm}$  are independent of the volatility process, and depend only on the factor  $\Lambda$ , the tracking-error aversion  $\gamma$ , and the spread  $\varepsilon$ . In consequence, also the average exposure  $\bar{\beta}$  and the squared correlation  $R^2$  are independent of volatility dynamics. By contrast, the tracking difference  $\text{TrD}$ , tracking error  $\text{TrE}$ , and the fund's volatility  $\bar{s}$  also depend on the average volatility  $\bar{\sigma}$ .

Part (vi) turns these qualitative insights into quantitative implications by deriving explicit series expansions of  $\pi_{\pm}$  and of the performance statistics. Each of these expansions depends on the unobservable parameter  $\gamma$ : eliminating this parameter from (17) and (18) yields the relation (21) among tracking error, tracking difference, and spread, which must hold regardless of the value of  $\gamma$ . In summary, the above theorem summarizes the optimal trading policy, its performance, and their statistical attributes. The next sections describe their main normative implications for replication and performance evaluation, while the following sections examine such implications empirically.

### 3.1. Trading boundaries

The optimal trading boundaries  $\pi_{\pm}$  identified by theorem 3.1 define a range of leverage ratios around the target factor  $\Lambda$ , on which it is optimal for the manager to refrain from buying and selling. For small trading costs, the first-order term of

the asymptotic expansion in (16) coincides with the optimal policy in Gerhold *et al.* (2014) for an investor with constant relative risk aversion  $\gamma$ , constant investment opportunities, and a target portfolio equal to  $\Lambda$ : superficially, tracking optimally is like investing optimally after replacing the investor's desired portfolio with the target factor.

The crucial difference between optimal tracking and portfolio choice is that the optimal tracking boundaries are insensitive to the volatility process. This means that in more volatile times the manager does rebalance more often, but only because the leverage ratio reaches the boundaries more often, not because the rebalancing policy changes.

An intuitive explanation of such robustness is that, because volatility is stationary, the objective function is invariant to a time change that stretches periods of high volatility and compresses periods of low volatility, as to normalize it to a constant: while in the original setting tracking error is high when volatility is high, in the time-changed problem volatility remains fixed, but time ticks faster, thereby accumulating more tracking error on average. As the optimal policy of the time-changed problem is independent of volatility, so is the optimal policy of the original one.

Figure 1 displays the optimal trading boundaries against tracking error: low tracking error (left, corresponding to high  $\gamma$ ) implies a narrow no-trade region, and thus frequent rebalancing. This regime is the most relevant for leveraged ETFs, as the users of these funds are primarily short-term traders,

for whom the precise replication of daily returns is more important than average performance (Cheng and Madhavan 2009).

As the aversion to tracking error decreases (right), the priority shifts to reducing long-term average rebalancing costs, at the expense of higher tracking error. For leveraged funds, widening the no-trade region serves this purpose—up to a point. As the emphasis on low trading costs dominates, it becomes virtually impossible for a leveraged fund to track its target: the next best alternative is to closely replicate the index, which can be done essentially for free.

For inverse funds the situation is different: the closest factor that is replicable at no cost is zero, therefore the no-trade region continues to widen until its sell boundary becomes exactly zero, while remaining symmetric around its target multiple to preserve pre-existing short positions. At that point, negative exposure is seldom decreased (at the buy boundary, for the sake of solvency) and never increased (as the sell boundary of zero is never reached).

Overall, the small tracking errors that are typically sought by managers of leveraged and inverse funds lead to similar replication strategies, based on nearly symmetric no-trade regions. Excessive emphasis on small costs would break this symmetry: because the only costless multiples are zero and one, leveraged funds would veer toward one, while inverse funds would veer toward preservation of negative exposure while retaining solvency.

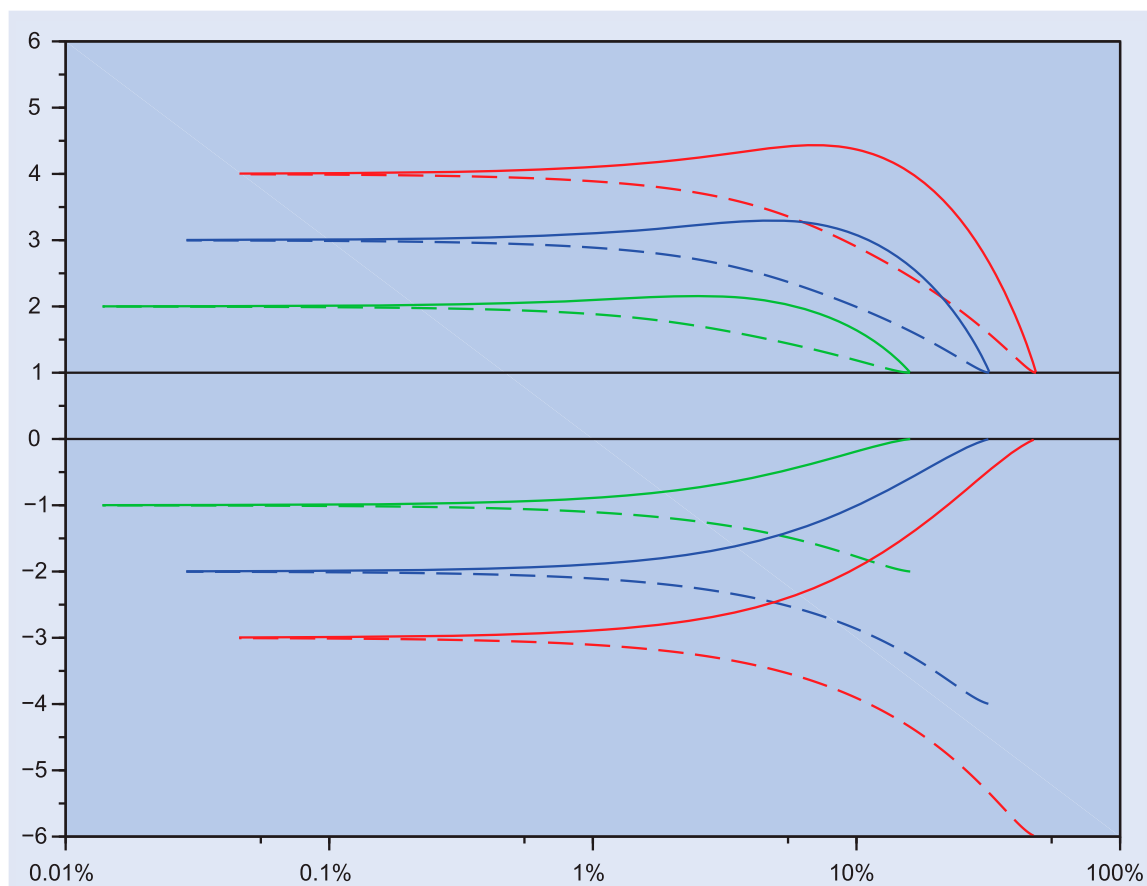


Figure 1. Buy (dashed) and sell (solid) boundaries (vertical, as risky weights  $\pi$ ) versus average tracking error TrE (horizontal) for leveraged (top) and inverse (bottom) funds, with multipliers 4 (top), 3, 2,  $-1$ ,  $-2$ ,  $-3$  (bottom). As aversion to tracking error  $\gamma$  decreases from left ( $\approx 10^6$ ) to right ( $\approx 10^{-4}$ ), for inverse funds (bottom) the trading boundaries widen around the target, whereas for leveraged funds (top) they first widen and then collapse to one.  $\varepsilon = 1\%$  and zero risk premium.

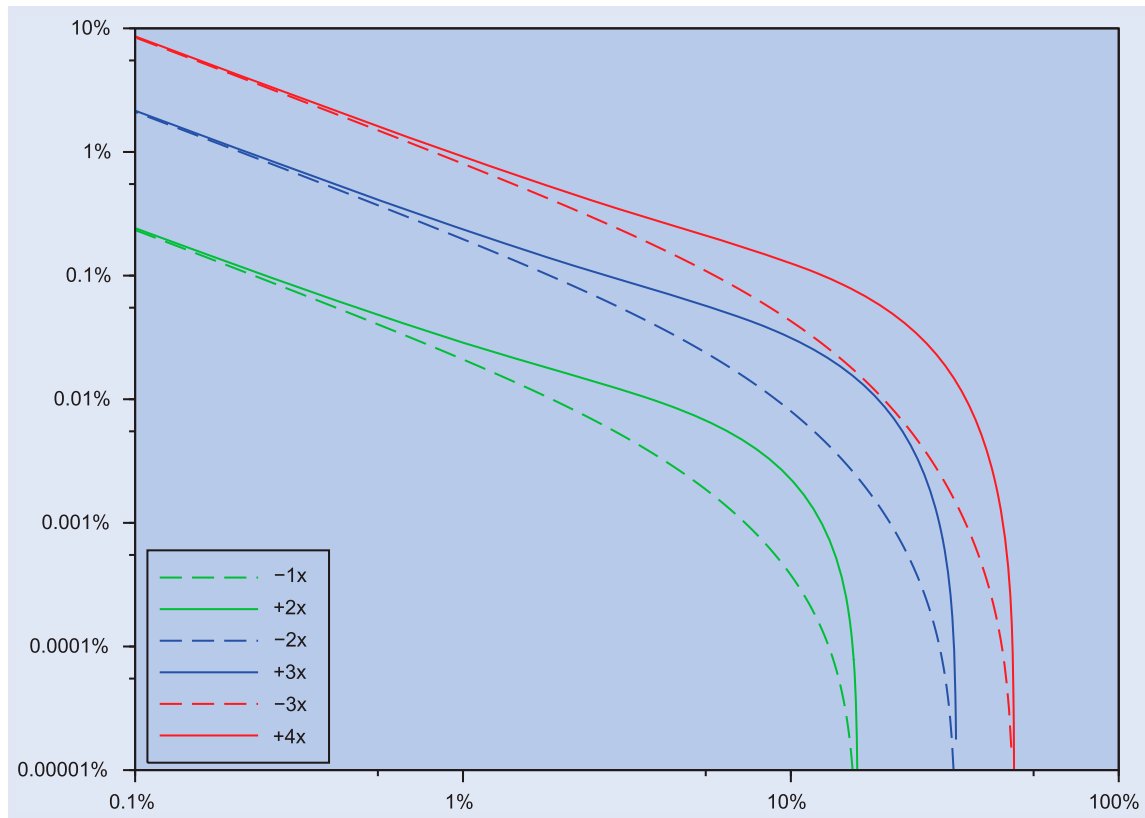


Figure 2. Negative tracking difference ( $-\text{TrD}$ , vertical) against tracking error (horizontal) for leveraged (solid) and inverse (dashed) funds, in logarithmic scale, from  $-3$ ,  $+4$  (top), to  $-2$ ,  $+3$  (middle), and  $-1$ ,  $+2$  (bottom). As aversion to tracking error  $\gamma$  decreases from left ( $\approx 10^6$ ) to right ( $\approx 10^{-4}$ ), a  $k + 1$ -leveraged fund is akin to a  $-k$  inverse one, as the respective curves (same color) approach low and high TrE aversion.  $\varepsilon = 1\%$ ,  $\sigma = 16\%$ , and zero risk premium.

### 3.2. Tracking difference versus tracking error

Figure 2 displays in log scale the tradeoff between tracking difference and tracking error for funds with different factors. For small trading costs, (21) implies that this tradeoff is approximately

$$\begin{aligned} \log(-\text{TrD}) + \log \text{TrE} \\ = \log \left( \frac{\sqrt{3}}{12} \bar{\sigma}^3 \Lambda^2 (\Lambda - 1)^2 \varepsilon \right) + O(\varepsilon^{4/3}), \end{aligned}$$

which corresponds to the negative, approximately linear dependence in the left of the graphs. As the right-hand side remains constant if  $\Lambda$  is replaced by  $-\Lambda + 1$ , for small tracking error a  $-(\Lambda - 1)$  inverse fund faces a tradeoff similar to the one of a leveraged fund with factor  $\Lambda$ . Put differently, a  $+3$  leveraged fund is as difficult to manage as a  $-2$  inverse fund. This is clear from the comparison of solid and dashed lines with same color in the figure.

As tracking error increases, the inverse fund's tradeoff improves, in that its optimal policy yields a less negative tracking difference than the symmetric leveraged fund because the boundaries of inverse funds keep widening even as those of leveraged funds start shrinking. As the tracking error increases further, their performances become again trivial, as both leveraged and inverse exposures are allowed to depart from their targets to avoid trading costs.

### 3.3. Implied spread

Equation (21) links the tracking difference, tracking error, leverage factor, index volatility, and trading spread  $\varepsilon$ . As the tracking error represents the standard deviation of the estimate of the tracking difference  $\text{TrD}$ , a low tracking error means that in this equation all quantities are estimated accurately, with the exception of the spread  $\varepsilon$ . In principle, one could estimate  $\varepsilon$  as the average bid-ask spread observed in the index, but in practice such an estimate would overstate the rebalancing costs of leveraged and inverse funds, as they achieve their index exposure primarily through total return swaps and futures, which offer much lower trading costs than the index components. In practice (cf. section 4), direct holdings of the index' components are typically less than the fund's assets for leveraged funds and are actually zero for inverse funds (i.e. there are no short positions in index' components but only in swaps and futures).

This observation suggests to use (21) to define the *implied spread* as

$$\tilde{\varepsilon} = -\frac{12}{\sqrt{3}} \frac{\text{TrD} \cdot \text{TrE}}{\bar{\sigma}^3 \Lambda^2 (1 - \Lambda)^2},$$

in a similar way as the familiar Black-Scholes formula is used to define implied volatility rather than to price options. The implied spread offers a more attractive measure of the fund's performance than the tracking difference or the tracking error,



as it controls for the effects of both volatility and the leverage factor.

This definition measures the trading cost, above which the investor is better off replicating the fund's payoff by trading, and below which the fund offers a cheaper alternative. The theoretical (population) spread is always positive, as the corresponding tracking difference is necessarily negative in view of trading costs. However, sample estimates can and do lead to occasionally negative implied spreads, which materialize if the fund deviates from a close replication, and happens to outperform its target by mere chance. In fact, such imprecise estimates of the implied spread result from the imprecise estimates of the tracking difference when the tracking error is large.

In summary, the implied spread offers a synthetic figure that represents the cost for investors of leveraged and inverse funds, adjusting for the effects of the factor, volatility, and the tradeoff between tracking error and tracking difference. However, such a measure is accurate only when the tracking difference is accurate, which is when the tracking error is sufficiently low.

### 3.4. *R-Squared*

The squared correlation of a fund's return with its benchmark is an intuitive, scale-free statistic, which reflects the fraction of the fund's variance that is explained by the benchmark's own variance. For funds with relatively low factors, such as the ones traded in US markets, the squared correlation is

similarly close to the frictionless level of 100%, leaving few insights to be gleaned from this quantity.

Equations (15) and (20) provide respectively the exact and asymptotic formulas of squared correlation. Importantly, the  $R^2$  does not depend on the volatility process, but only on the trading cost  $\varepsilon$ , the leverage factor  $\Lambda$ , and the aversion parameter  $\gamma$  that controls the tradeoff between tracking difference and tracking error. Figure 3 plots the squared correlation against the leverage factor, highlighting several effects. First, squared correlation declines for large positive and negative factors, as the optimal no-trade region widens and the fund's exposure is increasingly variable.

Second, squared correlation is asymmetric in the factor and systematically lowers for inverse funds. For positive factors, correlation is poor near zero, becomes perfect at the benchmark value of one, and deteriorates as leverage increases further. As negative factors do not include the benchmark, their correlation is generally lower, and achieves its maximum near minus one.

The low correlation near the zero factor, though empirically irrelevant, deserves a comment, in view of its differences with the factor of 1. The explanation of such low correlation is that a fund with near-zero exposure combines a very low volatility with a relatively wide no-trade region, resulting in a fund's exposure that is rather variable, and poorly explained by the benchmark's return.

There are two extreme cases,  $\Lambda \in \{0, 1\}$ , which are explicitly excluded in theorem 3.1. In these cases, an optimal trading policy is not to engage in trading at all; thus both transaction costs and tracking error vanish. Therefore, the tracking error

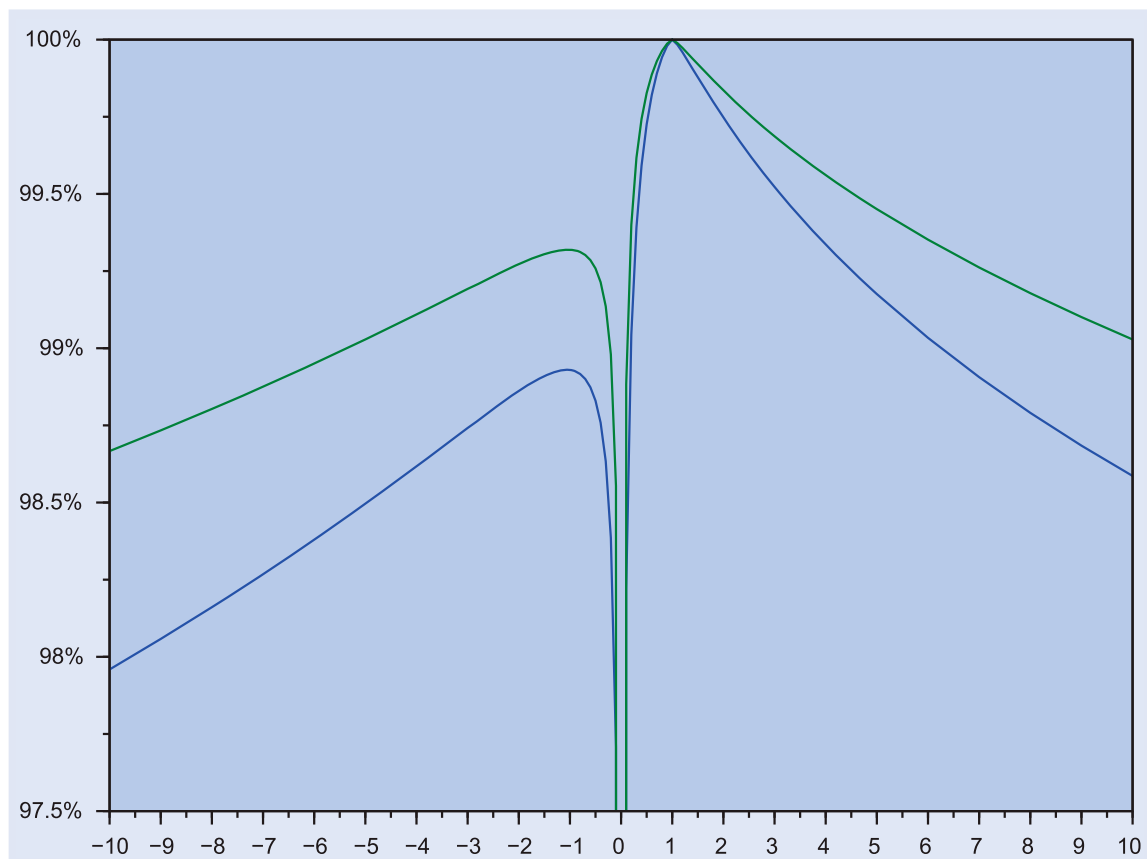


Figure 3. Exact  $R^2$  (vertical) for inverse and leveraged factors  $\Lambda$  (horizontal), with aversion parameter  $\gamma = 0.5$  (blue) and  $\gamma = 1$  (green).

cannot be freely chosen, as the value function vanishes and the optimal policy is independent of the value of  $\gamma$ .

Indeed, equations (13), (14) and (15) imply the relation between  $R^2$  and tracking error

$$\left(\frac{\text{TrE}}{\bar{\sigma}\Lambda}\right)^2 = 1 - R^2 + \left(\sqrt{\frac{\pi_-}{\Lambda} \frac{\pi_+}{\Lambda}} - R\right)^2,$$

which yields the asymptotic formula

$$R^2 = 1 - \left(\frac{\text{TrE}}{\bar{\sigma}\Lambda}\right)^2 + O(\varepsilon^{2/3}).$$

This formula validates the usual interpretation of  $1 - R^2$  as fraction of variance in the fund unexplained by the variation in the benchmark and shows that for  $\Lambda$  near zero even a small but positive tracking error can lead to a low  $R^2$ , as the factor  $\Lambda$  appears in the denominator.

#### 4. Replication of LIETFs

The model in this paper relies on the assumption that trading costs incurred by fund managers are a small constant proportion of amounts traded. Thus it is opportune to investigate the relevance of such an assumption for the replication of LIETFs.

Engle *et al.* (2012) estimate average transaction costs of 8.8 and 13.8 basis points (bps) for NYSE and NASDAQ respectively, based on orders executed by Morgan Stanley in 2004. Their estimates for trading cost rise to 27 bps for relatively larger orders—higher than 1% of the stock's daily trading volume. Similarly, Frazzini *et al.* (2012) estimate median transaction cost of 4.9 bps, with a value-weighted average of 9.5 bps, which reflect the higher cost of larger trades. In short, empirical studies suggest that trading costs are near 0.1%, except for orders that are higher than 1% of daily trading volume. Thus, to evaluate whether the assumption of constant trading costs is appropriate, it is necessary to understand (i) the details of the rebalancing strategies of leveraged and inverse ETFs and (ii) the relative size of their trading volume compared to that in the underlying index.

To investigate rebalancing strategies, we obtained from ProShares the daily holdings at the security level of the main leveraged and inverse ETFs on the S&P 500 index from January 26 to April 6, 2018. This period included a number of days with large market movements, which offer the opportunity to observe rebalancing behavior in volatile times. Because the institutional objective of these funds is to offer a multiple of the *daily* return on the underlying index, the daily

frequency is the natural one to study rebalancing. It is also the highest frequency for which public data are available.

Note also that LIETFs trade both to rebalance in response to the index' return, and to adjust exposure in response to the creation and redemption of shares by authorized participants. Thus the index return alone is not a reliable indicator of the trading volume. Instead, the direct inspection of daily portfolio composition reflects the overall trading volume resulting from both managers' rebalancing and investors' flows.

The left panel in table 1 shows the source of exposure to the index for each of the S&P 500 funds: most of the exposure is achieved through index swaps contracts with various counterparties (such as Goldman Sachs, Bank of America, JP Morgan, Credit Suisse, and other financial institutions). Such contracts are typically multiples of 10 or 100 million USD and are rebalanced in such multiples. A small fraction of the exposure is achieved through E-mini futures contracts on the S&P 500. For leveraged funds, a substantial amount of exposure, approximately 80% percent of the fund's assets, is achieved through direct ownership of each of the 505 index' components. By contrast, inverse funds do not take short positions in the index' components, thereby avoiding short sale costs associated with borrowing shares of individual stocks, while achieving exposure entirely through swaps and futures.

Comparing the changes in each security's daily holdings over time reveals the magnitude of the trading amounts generated by ProShares funds on the S&P 500 index. The right panel in table 1 displays the average daily turnover for each factor and asset class as a percentage of the fund's assets. As observed in the introduction, the overall turnover is significantly higher for larger negative factors, with the factor  $-3$  generating approximately twice as much turnover as  $-2$  or  $3$ . In addition, the table reveals that in practice most of the turnover is concentrated in index swaps, with a significant minority of trading in futures and, for positive factors, in stocks. Such a pecking order reflects the lower trading costs of index swaps relative to futures and of futures relative to stocks.

To understand the size of trading volume in LIETFs compared to trading volume in the index, we examined daily assets under management for all the Proshares funds on the S&P 500 index since June 23, 2009, the first day on which all funds multiples are available. This information allows to compute the total volume generated by all funds, both from rebalancing and from the subscription or redemption of shares. Comparing such daily volume to the total volume generated by trading in the stocks included in the S&P 500 index, yields for each day the ratio between ETFs volume and stock market volume. The summary statistics are in table 2: trading volume generated by LIETFs activity averages approximately

Table 1. Average exposure (left) and average daily turnover (right), both as percent of funds' values, in different asset classes for S&P 500 LIETFs: SPXU ( $-3$ ), SDS ( $-2$ ), SH ( $-1$ ), SSO ( $+2$ ), UPRO ( $+3$ ), from 2018-01-26 to 2018-04-06.

Factor	Futures	Stocks	Index Swaps	Factor	Futures	Stocks	Index Swaps
$-3.00$	$-12.45$	$0.00$	$-288.67$	$-3.00$	$2.96$	$0.00$	$10.65$
$-2.00$	$-9.85$	$0.00$	$-190.81$	$-2.00$	$1.60$	$0.00$	$5.24$
$-1.00$	$-5.18$	$0.00$	$-95.15$	$-1.00$	$0.82$	$0.00$	$2.37$
$2.00$	$5.90$	$83.20$	$110.89$	$2.00$	$0.84$	$0.46$	$1.77$
$3.00$	$6.90$	$83.93$	$209.16$	$3.00$	$1.17$	$0.78$	$5.89$

Table 2. Summary statistics for the ratio (as percent) between (i) daily volume generated by LIETFs on the S&P 500 and (ii) daily volume in the S&P 500 components. Data from 2009-06-23 to 2018-03-27.

1% Quantile	Median	99% Quantile	Mean
0.01	0.15	0.86	0.20

0.20% of trading volume in the S&P 500 index components, and on 99 out of 100 days it is below 0.86%.

These figures imply that, even in the implausible worst case scenario that all trading generated by LIETFs took place in underlying assets, such volume would be on average five times below the 1% threshold above which price impact starts to become relevant. In practice, as shown by table 1, only a minority of trading takes place in the underlying assets, with the bulk concentrated in index swaps and futures, which entail much lower trading costs, thereby indicating that the assumption of constant trading costs is germane to the present application.

## 5. Empirical results

This section examines the model's implications for leveraged and inverse exchange-traded funds listed in the USA.

Table 3 displays the performance statistics for some of the largest families of leveraged and inverse ETFs. To facilitate comparisons among different factors of the same benchmark, observations for each family are trimmed to the longest period that includes all factors (the youngest funds are usually the ones with triple exposure). Each row displays the relative performance of a leveraged or inverse fund compared to a hypothetical portfolio that trades without costs in the nonlevered fund on the same benchmark (e.g. SPY before management fees for the S&P 500), using as safe rate the 1-month treasury rate. The beta and tracking differences are calculated from the linear regression of the fund's daily returns on the returns of the corresponding hypothetical portfolio.

Average exposures (betas) are closely aligned with their targets, but consistently biased towards zero, and such differences are statistically significant for higher factors, as documented by their  $t$ -statistics. This is the underexposure puzzle, observed by Tang and Xu (2013), and explained in our model by equation (22) as the result of optimal rebalancing policies. The higher tracking error of funds with larger factors follow from the wider no-trade regions that are optimal for larger factors. Likewise, more negative tracking differences for larger factors may reflect higher rebalancing costs rather than management issues.

The implied spreads in table 3 tend to be lower for larger factors ( $-3$ ,  $-2$ ,  $+3$ ) than they are for smaller ones ( $-1$ ,  $+2$ ). In other words, the tracking differences of a larger factor is less negative than it should be if the spread was the same as for smaller factors. Such differences are large, ranging from around 5 basis points for larger factors to about 20 for small factors.

An explanation of this difference is in the relatively flat management fees charged by funds with different factors:

for example, the funds that replicate multiples of the S&P 500 index have expense ratios of 0.90% ( $-3$ ), 0.90% ( $-2$ ), 0.89% ( $-1$ ), 0.89% ( $+2$ ), and 0.91% ( $+3$ ). These expenses detract directly from the tracking difference of the fund, and make it much more onerous for an investor (in terms of implied rebalancing cost) to pay 0.90% for a fund that generates twice the index return, than to pay 0.91% for three times. For inverse funds it is even more obvious, as the once and twice inverse funds have virtually the same expenses.

The relevance of this explanation is further confirmed by the *gross* implied spreads, that is, the implied spreads obtained by the tracking differences before management fees (obtained in the data by adding to the market return the daily management fee of the fund). In accordance with the above explanation, such gross implied spreads are significantly closer to each other across different factors. Some of the gross spreads for the most liquid funds take small negative values due to the effect of tracking error, as the realized return in the sample period is minimally positive before fees.<sup>†</sup> Consistently with the model, gross implied spreads are higher for less liquid asset classes, such as smaller stocks (S&P Midcap 400 and Russel 2000).

In summary, while the gross implied spread is useful to understand the impact of management fees and managers' tracking efficiency, the relevant performance measure for investors is the net implied spread, as it captures the only return that is available to them, i.e. net of fees.

### 5.1. Underexposure

In their empirical work on leveraged and inverse ETFs listed in US exchanges, Tang and Xu (2013) report that 'LETFs show an underexposure to the index that they seek to track'. That is, empirical average exposures are significantly closer to zero than their target multiple. Table 3 confirms the underexposure phenomenon for several families of leveraged and inverse ETFs traded in US markets: the comparison of realized average exposure with target factors yields significant  $t$ -statistics for the riskiest funds, with deviations consistently biased toward zero.

This empirical regularity is explained by equations (14) and (19) as a consequence of managers' optimal rebalancing policies. Indeed, the transaction cost correction in

$$\bar{\beta} = \Lambda - \frac{2\Lambda - 1}{\gamma} \left( \frac{\gamma \Lambda (\Lambda - 1)}{6} \right)^{1/3} \varepsilon^{2/3} + O(\varepsilon) \quad (22)$$

is negative for  $\Lambda > 1$  and positive for  $\Lambda < 0$ . This effect may seem puzzling at first, as the buy and sell boundaries are approximately symmetric around  $\Lambda$ .

The crucial point is that when the exposure  $\pi_t$  is larger (farther from zero), it is also more variable. As the trading policy forces exposure to remain in a fixed range, the implication is that, within such range, the exposure on average spends more time on less variable values—where it is closer to zero. Thus the underexposure effect arises from the combination of

<sup>†</sup> The notable exception is the EET, which has a large negative tracking difference and hence a large implied spread, even after controlling for management fees.

Table 3. Performance of LIETFs on selected indexes through December 31st, 2022. For each index, observations begin on the earliest date for which all inverse ( $-3$ ,  $-2$ ,  $-1$ ) and leveraged ( $2$ ,  $3$ ) multiples are available. (Earliest date is November 19th, 2008.) Historical prices from Yahoo Finance, Interest rates from US Treasury.

Index	Ticker	X	Track. Error (bp)	Tracking Diff. (%)	Tracking Diff. (Gross) (%)	Implied Spread (bp)	Implied Spread (Gross) (bp)	Beta	T-stat (Beta) (%)	Vol. (%)	Years Data
S&P 500 (SPY)	SPXU	-3	5.41	-1.18	-0.26	0.57	0.13	-2.99	4.00	52.59	13.51
	SDS	-2	2.94	-0.74	0.15	0.77	-0.16	-2.00	-2.66	35.26	13.51
	SH	-1	1.96	-0.68	0.21	4.24	-1.31	-1.01	-6.52	17.73	13.51
	SSO	2	2.93	-1.59	-0.70	14.92	6.59	2.01	3.42	35.28	13.51
	UPRO	3	5.31	-2.29	-1.34	4.31	2.52	3.00	0.88	52.85	13.51
S&P Midcap 400 (MDY)	SMDD	-3	16.45	-2.45	-1.50	2.19	1.34	-2.96	5.10	61.90	12.88
MSCI Emerging Markets (EEM)	MZZ	-2	11.16	-2.31	-1.36	5.61	3.30	-1.94	11.75	40.54	12.88
	MYV	-1	4.69	-1.11	-0.16	10.18	1.46	-1.00	1.89	20.76	12.88
	MVV	2	4.22	-1.42	-0.47	11.78	3.92	1.99	-3.78	41.26	12.88
	UMDD	3	13.73	-1.58	-0.63	4.72	1.88	2.92	-12.06	60.88	12.88
	EDZ	-3	6.75	-1.39	-0.44	0.39	0.12	-2.96	13.31	66.89	13.57
Nasdaq 100 (QQQ)	EEV	-2	4.51	-1.63	-0.68	1.23	0.51	-1.99	3.22	45.05	13.57
	EUM	-1	2.83	-1.14	-0.19	4.87	0.81	-1.00	2.98	22.53	13.57
	EET	2	12.26	-2.74	-1.79	50.78	33.20	1.95	-8.81	44.61	13.57
	EDC	3	7.74	-5.12	-4.18	6.65	5.42	2.93	-19.07	66.35	13.57
	SQQQ	-3	5.81	-2.28	-1.33	0.71	0.41	-2.96	14.15	61.63	12.88
Russell 2000 (IWM)	QID	-2	3.05	-1.21	-0.26	0.79	0.17	-1.99	5.39	41.45	12.88
	PSQ	-1	2.13	-1.00	-0.05	4.10	0.20	-1.00	3.90	20.74	12.88
	QLD	2	3.06	-1.66	-0.71	9.75	4.16	1.99	-4.38	41.48	12.88
	TQQQ	3	5.43	-2.24	-1.30	2.61	1.51	2.95	-17.84	61.47	12.88
	TZA	-3	6.55	-4.95	-4.00	1.09	0.88	-2.99	5.10	72.59	14.11
Barclays 20+ Treasury (TLT)	TWM	-2	4.79	-3.12	-2.18	2.01	1.40	-1.99	5.47	48.37	14.11
	RWM	-1	2.66	-1.61	-0.67	5.20	2.14	-1.00	-1.80	24.37	14.11
	UWM	2	4.47	-1.67	-0.67	9.03	3.62	1.99	-3.77	48.45	14.11
	TNA	3	6.03	-3.33	-2.38	2.70	1.93	2.96	-15.29	71.99	14.11
	TTT	-3	11.75	-2.13	-1.18	4.01	2.22	-2.87	15.84	41.88	10.75
Dow Jones Industrial Average (DIA)	TBT	-2	8.90	-1.66	-0.74	9.43	4.20	-1.90	16.46	27.79	10.75
	TBF	-1	5.08	-1.24	-0.31	36.24	9.03	-0.96	12.38	14.05	10.75
	UBT	2	11.07	-1.07	-0.12	68.49	7.95	1.93	-9.49	28.40	10.75
	TMF	3	5.87	-1.71	-0.76	6.44	2.87	2.94	-13.53	42.61	10.75
	SDOW	-3	4.61	-1.39	-0.45	0.61	0.19	-2.97	12.03	51.04	12.88
Industrial Average (DIA)	DXD	-2	2.66	-0.92	0.03	0.93	-0.03	-1.99	6.84	34.20	12.88
	DOG	-1	1.77	-0.85	0.10	5.14	-0.59	-1.00	4.18	17.13	12.88
	DDM	2	3.01	-1.58	-0.63	16.28	6.50	1.98	-10.35	34.08	12.88
	UDOW	3	5.48	-1.91	-0.96	3.96	1.99	2.95	-14.49	50.83	12.88

symmetric trading boundaries with the asymmetric volatility of the fund.

Substituting in equation (22) a fund's average realized exposure as  $\beta$ , it is possible to solve for the ratio  $\varepsilon/\gamma$ , which, combined with a value of the trading cost  $\varepsilon$ , yields an estimate for  $\gamma$ . Larger deviations of the average exposure  $\beta$  from the target factor  $\Lambda$  imply a lower  $\gamma$  (the manager favors lower trading costs over lower tracking error) while smaller deviations imply a larger value of  $\gamma$ .

For example, the SPXU has a realized  $\beta = -2.99$  and a factor  $\Lambda = -3$ , which imply  $\gamma \approx 10$  assuming that  $\varepsilon = 0.1\%$ , a figure consistent with the discussion in the previous section. Table 4 displays the average exposure implied for funds with typical factors for different values of  $\varepsilon$  and  $\gamma$ . Small factors ( $-1$  and  $2$ ) lead to average exposures that are almost indistinguishable from their targets, especially when statistical error is accounted for. On the other hand, larger factors ( $-3$ ,  $-2$ , and  $3$ ) lead to discernible differences across TrE aversion  $\gamma$  and trading costs  $\varepsilon$ . In particular, the configuration  $\gamma = 10$ ,  $\varepsilon = 0.1\%$  closely approximates the panel of S&P 500 funds and  $\gamma = 5$ ,  $\varepsilon = 0.1\%$  the panel of Dow Jones funds. In general, values of  $\gamma$  between 5 and 10 with  $\varepsilon = 0.1\%$

closely reproduce the empirical estimates on funds tracking large-capitalization stocks, while higher trading costs  $\varepsilon$  are warranted for small-capitalization stocks and treasuries.

**Granularity.** It is worthwhile to consider whether the observed underexposure may have different explanations. One ostensible source of underexposure could be granularity. Indeed, the minimum trade size for the E-mini S&P 500 Futures contract (which is held by the ETFs tracking this index, as explained in the previous section) equals \$50 times the index value, which means that when the index is 2800, the contract size is \$140,000. Thus for a fund with a \$1 billion exposure (the SSO has more than 2 billion in assets, hence an exposure of over 4 billions), the granularity of futures contracts could explain a discrepancy in average exposure of the order of  $1.4 \times 10^5 / 10^9 \approx 10^{-4}$ .

In fact, the inspection of daily portfolio holdings described above reveals that, because daily rebalancing of leveraged (but not inverse) funds is partly achieved by trading incremental quantities of individual equities, the minimum trade unit for a fund is the price of one share, typically between \$10 and \$1000. Thus, for a fund with \$1 billion exposure,



Table 4. Average exposure ( $\tilde{\beta}$ ) for typical factors ( $\Lambda$ ) across trading costs ( $\varepsilon$ ) and TrE aversion ( $\gamma$ ), obtained from equation (22).

$\varepsilon = 0.1\%$			
$\Lambda$	$\gamma$		
	1	5	10
−3	−2.91	−2.97	−2.98
−2	−1.95	−1.98	−1.99
−1	−0.98	−0.99	−1.00
2	1.98	1.99	2.00
3	2.95	2.98	2.99
$\varepsilon = 0.5\%$			
$\Lambda$	$\gamma$		
	1	5	10
−3	−2.74	−2.91	−2.94
−2	−1.85	−1.95	−1.97
−1	−0.94	−0.98	−0.99
2	1.94	1.98	1.99
3	2.85	2.95	2.97

the granularity of equity prices could justify a discrepancy in average exposure of the order of  $1000/10^9 = 10^{-6}$ . Note also that granularity—whether from futures contract sizes or equity prices—would generate either underexposure or overexposure with comparable probabilities.

Empirically, the underexposure observed even for the largest ETFs, with assets under management above \$1 billion, is of the order of  $0.01 = 10^{-2}$ , while overexposure never appears in any of the 40 funds in our original sample. Taken together, these observations indicate that granularity does not reproduce quantitatively the observed underexposure of major leveraged and inverse exchange-traded funds.

**Noisy Returns.** Another potential source of underexposure is measurement error in the returns of LIETFs and their underlying index. For example, such an error can arise if such returns are calculated from net asset values (NAVs) rather than market prices, which differ slightly from NAVs, due to time lags in their calculation and to small deviations, within the cost of subscription and redemption, which cannot be arbitrated away.

Thus, denote the observed return of a fund as  $\tilde{R}_t^F := R_t^F + \varepsilon_t^F$ , where  $R_t$  is the true return and  $\varepsilon_t$  is a measurement error and, likewise, the observed return of the underlying index as  $\tilde{R}_t^I := R_t^I + \varepsilon_t^I$ . Then, assuming that returns are stationary and that errors are uncorrelated with true returns and with each other, the large sample average exposure converges in probability to

$$\tilde{\beta} = \frac{\text{Cov}(\tilde{R}_t^F, \tilde{R}_t^I)}{\text{Var}(\tilde{R}_t^I)} = \frac{\text{Cov}(R_t^F, R_t^I)}{\text{Var}(R_t^I) + \text{Var}(\varepsilon_t^I)} = \frac{\beta}{1 + \frac{\text{Var}(\varepsilon_t^I)}{\text{Var}(R_t^I)}} \approx \beta \left( 1 - \frac{\text{Var}(\varepsilon_t^I)}{\text{Var}(R_t^I)} \right),$$

where  $\beta = \frac{\text{Cov}(R_t^F, R_t^I)}{\text{Var}(R_t^I)}$  is the true average exposure. In other words, measurement error in the fund's return does not bias

the estimate of average exposure, but measurement error in the index's return generates a spurious proportional underexposure of  $\text{Var}(\varepsilon_t^I)/\text{Var}(R_t^I)$ .

Note that such underexposure is independent of the true or observed fund returns, and in particular it would be independent of the leverage factor, while the underexposure observed in table 3 is much bigger for larger factors. This qualitative observation already casts doubt on such mechanism as a source of the observed underexposure. (In addition, the results in table 3 do not rely on NAVs but on market prices only.)

A detailed analysis of returns of NAV and market prices sheds additional light on this issue. Indeed, for the SPY, which replicates the S&P 500 index, the standard deviation of the difference between daily NAV returns and market returns is 0.16%, while the standard deviations of both the market and the NAV daily returns are 1.15%, which implies that

$$\text{Var}(\varepsilon_t^I)/\text{Var}(R_t^I) = \frac{(0.16\%)^2}{(1.15\%)^2} \approx 1.9\%.$$

(The observation period is 2003-12-02 to 2018-12-31.) Put differently, a measurement error equivalent to the discrepancy between NAV and market price would generate a reduction in average exposure of nearly 2%, irrespective of the factor considered. By contrast, the underexposures in table 4 are much smaller for small factors, as NAVs are never used in the analysis.

The above observation helps understand why the underexposures reported here are smaller than the ones observed by Tang and Xu (2013). Indeed, they calculate average exposures of LIETFs with respect to the underlying indexes rather than the market return of the nonlevered fund on the same index (this paper's approach). The benchmark index (which is not itself traded) may exhibit deviations from the market price of the same order as the NAV, and in fact Tang and Xu (2013, table 1 B) report for the SPY an exposure of 98.48%, i.e. an underexposure of 1.52%, which is broadly consistent with the effect of noisy returns.

## 6. Robustness

This section discusses, theoretically and numerically, the robustness of previous results to risk premia with stochastic volatility, finite horizons, and discrete (as opposed to continuous) trading.

### 6.1. Risk premia: trading strategies and performance

Assume that the underlying  $S$  has non-zero risk premia,

$$\frac{dS_t}{S_t} = (r_t + \mu_t) dt + \sigma_t dB_t, \quad S_0 \geq 0,$$

where  $\mu_t = \kappa \sigma_t^2$ , with  $\kappa \geq 0$ . Thus this specification prescribes that the Sharpe ratio increases as return variance increases, as in the asset pricing models of Campbell and Cochrane (1999), Bansal and Yaron (2004), and others, and consistent with the empirical countercyclicity of risk premia.



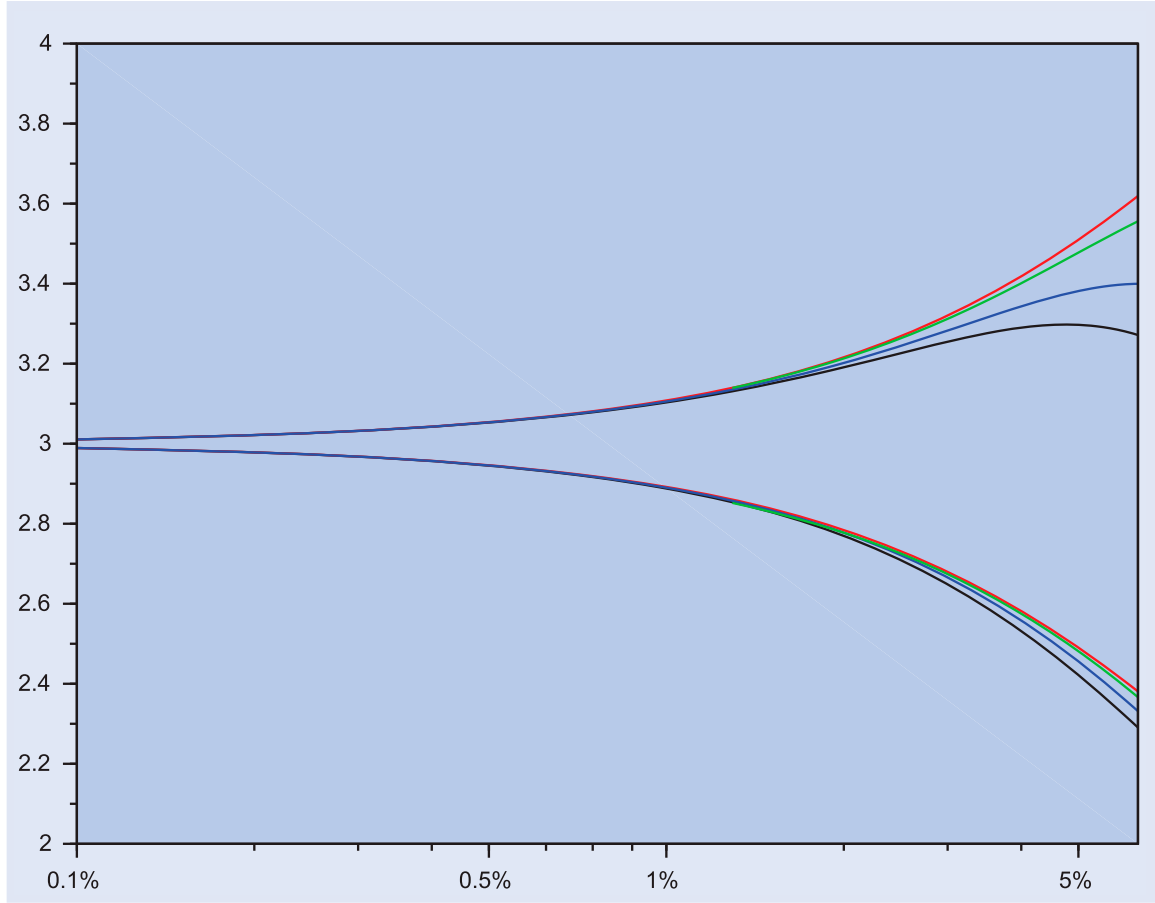


Figure 4. Trading boundaries (vertical) for factor +3 against tracking error (horizontal), from first-order asymptotics (25) (red) and exact solutions for  $\kappa = \mu_t/\sigma_t^2 = 0$  (black),  $\kappa = 1.5625$  (blue), and  $\kappa = 3.125$  (green). For realistic levels of tracking error, the impact of  $\kappa$  is of second order.

In view of the excess returns  $\mu_t$ , the cumulative returns process  $D_t$  defined by (4) includes a component  $\mu_t(\pi_t - \Lambda) dt$  that can be ascribed to the deviation of the fund's exposure from its target. As the exposure  $\pi_t$  is observable to investors,<sup>†</sup> they can accordingly control for overexposure while evaluating performance, by modifying the definition of the deviation process as

$$dD_t := \frac{dF_t}{F_t} - r_t dt - \Lambda \left( \frac{dS_t}{S_t} - r_t dt \right) - \mu_t(\pi_t - \Lambda) dt, \quad D_0 = 0. \quad (23)$$

Replacing (4) with (23), the previous identities (5)–(6) for Tracking Difference (TrD) and Tracking Error (TrE) remain valid, and thus the objective of a leveraged ETF manager is to find a trading policy  $\pi_t$  that minimizes the average equivalent expense ratio

$$EER(\varphi) := \liminf_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \frac{\gamma}{2} \int_0^T (\pi_t - \Lambda)^2 \sigma_t^2 dt + \varepsilon \int_0^T \pi_t \frac{d\varphi_t^\downarrow}{\varphi_t} \right]$$

over all  $\varphi \in \Phi$ , where  $\Phi$  is the set of admissible strategies (see appendix 1). Note that in this problem, although

<sup>†</sup> Note that  $d\langle F, S \rangle_t = F_t S_t \pi_t \sigma_t^2 dt$  and  $d\langle S \rangle_t = S_t^2 \sigma_t^2 dt$ , whence  $\pi_t = S_t/F_t \cdot d\langle F, S \rangle_t/d\langle S \rangle_t$ .

the objective function controls for extra exposure, the risk-premium  $\kappa$  has not disappeared completely, as it affects the dynamics of the portfolio weight  $\pi_t$ . The question is whether this effect is significant enough to alter the previous results.

The next theorem offers a negative answer, in which the effect of the risk premium is found to be of the second order:

**THEOREM 6.1 (Risk Premia)** *Let  $\gamma > 0$  and  $\Lambda \neq 0, 1$ . For  $\varepsilon > 0$  small enough:*

(i) *The free boundary problem*

$$\left( \frac{1}{2} \zeta^2 W'(\zeta) + \kappa \zeta W(\zeta) - \gamma \Lambda \left( \frac{\zeta}{1 + \zeta} \right) - \frac{\gamma}{2} \left( \frac{\zeta}{1 + \zeta} \right)^2 \right)' = 0, \quad (24)$$

*subject to (10)–(11) has a unique solution  $(W, \zeta_-, \zeta_+)$ .*

- (ii) *The trading strategy  $\hat{\varphi}$  that buys at  $\pi_- := \zeta_-/(1 + \zeta_-)$  and sells at  $\pi_+ := \zeta_+/(1 + \zeta_+)$  as little as to keep the risky weight  $\pi_t$  within the interval  $[\pi_-, \pi_+]$ , is optimal.*
- (iii) *The trading boundaries  $\pi_-$  and  $\pi_+$  depend only on the fraction  $\kappa = \mu_t/\sigma_t^2$ . They have the asymptotic*

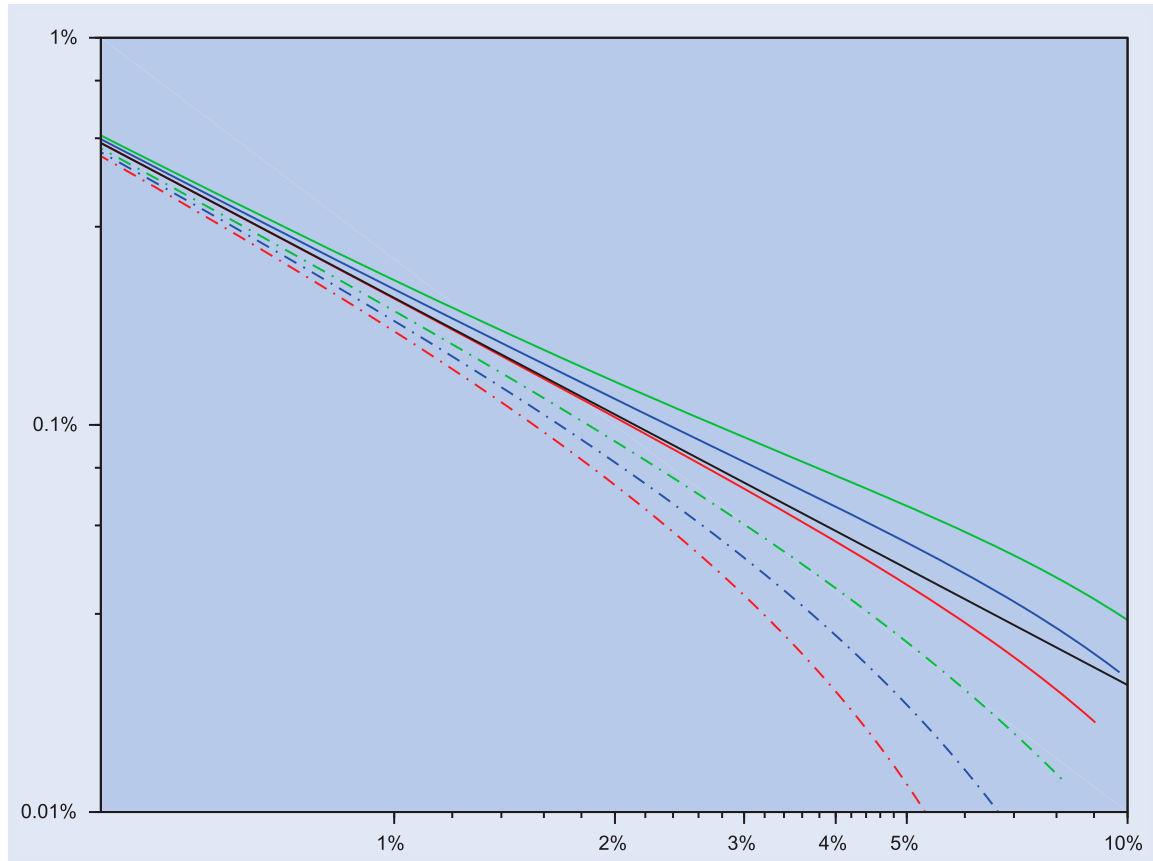


Figure 5. Negative tracking difference ( $-\text{TrD}$ ) for +3 leveraged fund (vertical, colored solid lines) and  $-2$  (dashed colored lines), against tracking error ( $\text{TrE}$ ) (horizontal), for  $\kappa = \mu_t/\sigma_t^2 = 0\%$  (black),  $\kappa = 1.5625$  (blue) and  $\kappa = 3.125$  (green). The solid red line represents the asymptotic formula in (1), which is the same for  $L$  and  $1-L$ .

expansions

$$\begin{aligned} \pi_{\pm} = & \Lambda \pm \left( \frac{3}{4\gamma} \Lambda^2 (\Lambda - 1)^2 \right)^{1/3} \varepsilon^{1/3} \\ & - \frac{(\Lambda - \kappa)}{\gamma} \left( \frac{\gamma \Lambda (\Lambda - 1)}{6} \right)^{1/3} \varepsilon^{2/3} + O(\varepsilon). \end{aligned} \quad (25)$$

(iv) The minimum Equivalent Expense Ratio is

$$\min_{\varphi \in \Phi} \text{EER}(\varphi) = \frac{\gamma \bar{\sigma}^2}{2} (\pi_- - \Lambda)^2, \quad (26)$$

where  $\Phi$  is the set of admissible strategies (see appendix 1).

Thus the risk premium  $\kappa$  affects the optimal trading boundaries only in the  $\varepsilon^{2/3}$  term, which is negligible for reasonable parameter values, as illustrated by figure 4. Similarly, the tracking difference and the average exposure of the fund are also affected at the second order. In particular, the performance formula (21) remains valid, as it involves only first-order terms.

On the issue of performance, figure 5 reproduces the relevant part of figure 2 with realistic risk premia. Again, the main message is that the tradeoff between tracking error and tracking difference remains virtually unchanged at typical levels of tracking error and target factors.

## 6.2. Finite horizons, trading frequency, and stochastic volatility

The use of a stationary objective raises the question of how robust the results are to fixed horizons, and how well stationary optimal strategies perform when trading takes place at discrete frequency, rather than continuously. Figure 6 plots the asymptotic tradeoff between tracking difference and tracking error to the one obtained from Monte Carlo simulations from geometric Brownian motion (Black–Scholes model) and from the Heston stochastic volatility model, in the presence of a risk premium, with daily rebalancing frequency, and for a finite horizon of 5 years.

For the Heston model, set  $v_t := \sigma_t^2$ , where  $v_t$  satisfies the dynamics

$$dv_t = \theta(k - v_t) dt + \sigma \sqrt{v_t} dW_t,$$

$$\frac{dS_t}{S_t} = (r_t + \kappa v_t) dt + \sqrt{v_t} dB_t,$$

with an instantaneous correlation  $d\langle B_t, W_t \rangle = \rho dt$ , and the parametrization  $v_0 = \bar{\sigma}^2$ ,  $\zeta_0 = \zeta_* = \frac{\Lambda}{1-\Lambda}$ , unconditional mean  $k = \bar{\sigma}^2$ , mean-reversion speed  $\theta = 5$  and diffusion coefficient  $\sigma$ . A minimal trade is executed whenever the proportion of wealth  $\pi$  leaves the no-trade region  $[\pi_-, \pi_+]$  as to return  $\pi$  to its closest boundary. The exact trading boundaries  $\pi_{\pm}$  are obtained by solving numerically (24) subject to the free boundary conditions (10)–(11).

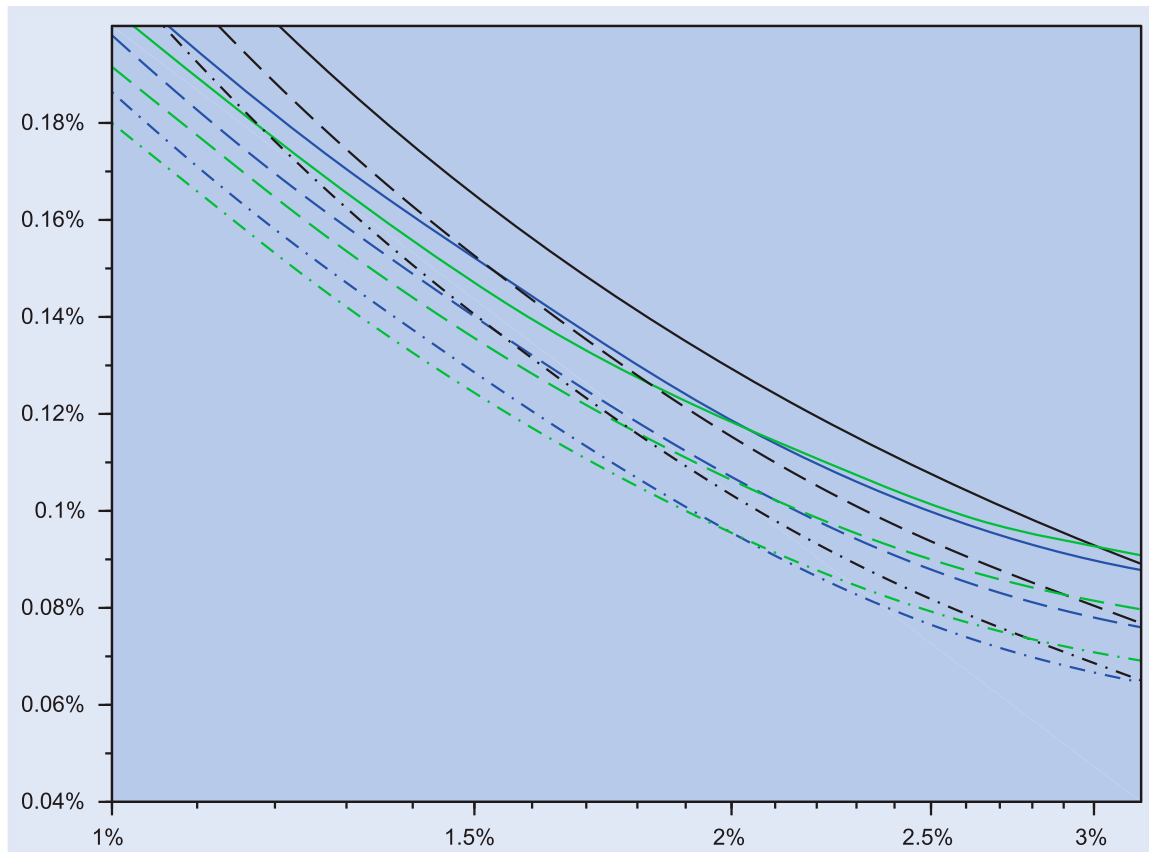


Figure 6. Negative tracking difference ( $- \text{TrD}$ ) for a  $+3$  leveraged fund (vertical), against tracking error (horizontal), for  $\mu_t/\sigma_t = 0$  (solid lines), 1.5625 (dashed) and 3.125 (dashed-dotted), when  $\varepsilon = 0.1\%$ . The black lines depicts the exact performance from (26) with  $\bar{\sigma} = 16\%$ . The blue and green lines correspond to Monte Carlo Estimates (15000 paths, daily grid) for the Black-Scholes (blue) and Heston (green) models, with time horizon of  $T = 5$  years, daily sampling and trading frequency, and correlation  $\rho = -0.9$ .

Again, the message of figure 6 is that, for realistic parameter values, risk premia, trading frequencies, and finite horizons have only a minor effect on the basic tradeoff identified by the asymptotic formulas in the paper. While these features make the optimization problem more complex, they remain largely inconsequential for performance evaluation and replication purposes.

## 7. Conclusion

Leveraged and inverse funds seek to replicate a multiple of the daily return on an index by frequently rebalancing their portfolio to keep a constant leverage ratio. In theory, in a frictionless market continuous rebalancing yields a perfect replication of a leveraged benchmark, i.e. zero alpha and tracking error. In practice, trading costs create a trade-off between the frequent rebalancing that generates low tracking error and the low trading costs that prevent alpha from becoming too negative. This trade-off becomes more relevant for products that seek to replicate larger multiples or less liquid benchmarks.

Portfolio performance measures based on the regression of a fund's return against its benchmark's return are ubiquitous, hence closely monitored by managers who may be evaluated on their basis. On average, the negative tracking difference of a leveraged fund results from management fees and trading costs, while the tracking error reflects the inevitable deviations from the target exposure that are necessary to

keep trading costs under control. Other things equal, a fund that seeks to replicate a larger multiple of an index' return has a higher tracking error and a more negative tracking difference.

It is misleading to compare funds with different factors, and to conclude that one is better managed than the other merely because its tracking error is lower, or because its alpha is less negative: two funds with the same factor may be optimally managed, as one may seek lower tracking error at the expense of more negative tracking difference. High tracking error is not necessarily evidence of poor manager performance if the index is less liquid. On the contrary, a savvy management strategy must accept higher tracking error to achieve less negative tracking differences, and low tracking error inevitably depresses average performance.

## Open Scholarship

This article has earned the Center for Open Science badges for Open Data and Open Materials through Open Practices Disclosure. The data and materials are openly accessible at <https://github.com/guasoni/leveraged-funds/>.

## Acknowledgments

For helpful comments, we thank Michael Brennan, Doug Breeden, Lisa Goldberg, Tim Leung, Johannes Muhle-Karbe,

Mathieu Rosenbaum, Ronnie Sircar, Matt Spiegel, René Stulz, Peter Tankov and seminar participants at University of Vienna, London School of Economics, Johns Hopkins University, Séminar Bachelier Paris, University of Warwick, the 9th Financial Risks International Forum (Paris), and the Byrne Workshop at the University of Michigan. A special thanks to the anonymous referee, who helped improve the interpretation of figure 1.

## Data Availability Statement

The code and data supporting this paper's findings are at <https://github.com/guasoni/leveraged-funds>.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work is supported by the European Research Council (ERC-279582), the National Science Foundation (DMS-1412529), and Science Foundation Ireland (16/IA/4443 and 16/SPP/3347).

## ORCID

Paolo Guasoni  <http://orcid.org/0000-0002-8562-3658>  
Eberhard Mayerhofer  <http://orcid.org/0000-0003-4051-1852>

## References

- Anderson, R.M., Bianchi, S.W. and Goldberg, L.R., Will my risk parity strategy outperform? *Financial Anal. J.*, 2012, **68**(6), 75–93.
- Anderson, R.M., Bianchi, S.W. and Goldberg, L.R., Determinants of levered portfolio performance. *Financial Anal. J.*, 2014, **70**(5), 53–72.
- Avellaneda, M. and Zhang, S., Path-dependence of leveraged ETF returns. *SIAM J. Financial Math.*, 2010, **1**(1), 586–603.
- Bansal, R. and Yaron, A., Risks for the long run: A potential resolution of asset pricing puzzles. *J. Finance*, 2004, **59**(4), 1481–1509.
- Campbell, J.Y. and Cochrane, J.H., By force of habit: A consumption-based explanation of aggregate stock market behavior. *J. Econom. Theory*, 1999, **107**(2), 205–251.
- Carr, P. and Madan, D., Towards a theory of volatility trading. In *Option Pricing, Interest Rates and Risk Management*, Handb. Math. Finance, pp. 458–476, 2001 (Cambridge University Press: Cambridge).
- Charteris, A. and McCullough, K., Tracking error vs tracking difference: does it matter? *Investment Anal. J.*, 2020, **49**(3), 269–287.
- Cheng, M. and Madhavan, A., The dynamics of leveraged and inverse exchange-traded funds. *J. Investment Manag.*, 2009, **16**(4), 43.
- Constantinides, G., Capital market equilibrium with transaction costs. *J. Polit. Economy*, 1986, **94**(4), 842–862.
- Dai, M., Kou, S., Soner, H.M. and Yang, C., Leveraged exchange-traded funds with market closure and frictions. *Manage. Sci.*, 2022, **69**(4), 2517–2535.
- Davis, M.H.A. and Norman, A.R., Portfolio selection with transaction costs. *Math. Oper. Res.*, 1990, **15**(4), 676–713.
- Dumas, B. and Luciano, E., An exact solution to a dynamic portfolio choice problem under transactions costs. *J. Finance*, 1991, **46**(2), 577–595.
- Dupire, B., Model art. *Risk*, 1993, **6**(9), 118–124.
- Engle, R., Ferstenberg, R. and Russell, J., Measuring and modeling execution cost and risk. *J. Portfolio Manag.*, 2012, **38**(2), 14–28.
- ESMA, Guidelines for competent authorities and units management companies: Guidelines on ETFs and other UCITS issues, 2014. European Securities and Markets Authority. [https://www.esma.europa.eu/sites/default/files/library/2015/11/esma-2014-0011-01-00\\_en\\_0.pdf](https://www.esma.europa.eu/sites/default/files/library/2015/11/esma-2014-0011-01-00_en_0.pdf).
- Frazzini, A., Israel, R. and Moskowitz, T.J., Trading costs of asset pricing anomalies. Fama-Miller Working Paper, 14–05, 2012.
- Frino, A. and Gallagher, D.R., Tracking S&P 500 index funds. *J. Portfolio Manag.*, 2001, **28**(1), 44–55.
- Gerhold, S., Guasoni, P., Muhle-Karbe, J. and Schachermayer, W., Transaction costs, trading volume, and the liquidity premium. *Finance Stoch.*, 2014, **18**(1), 1–37.
- Guasoni, P. and Mayerhofer, E., The limits of leverage. *Math. Finance*, 2019, **29**(1), 249–284.
- Gunning, R.C. and Rossi, H., *Analytic Functions of Several Complex Variables*, Vol. 2009 (American Mathematical Society).
- Jarrow, R.A., Understanding the risk of leveraged ETFs. *Finance Res. Lett.*, 2010, **7**(3), 135–139.
- Johnson, B., Bioy, H., Kellett, A. and Davidson, L., On the right track: measuring tracking efficiency in ETFs. *J. Index Investing*, 2013, **4**(3), 35–41.
- Kallsen, J. and Muhle-Karbe, J., The general structure of optimal investment and consumption with small transaction costs. *Math. Finance*, 2017, **27**(3), 659–703.
- Lu, L., Wang, J. and Zhang, G., Long term performance of leveraged ETFs. *Financial Services Rev.*, 2012, **21**(1). [https://go.gale.com/ps/retrieve.do?tabID=T002&resultListType=RESULT\\_LIST&searchResultsType=SingleTab&retrievalId=8f6fb502-7809-4a00-90cf-f06b2d717375&hitCount=6&searchType=AdvancedSearchForm&currentPosition=6&docId=GALE%7CA309587934&docType=Report&sort=Relevance&contentSegment=ZONE-Exclude-FT&prodId=AONE&pageNum=1&contentSet=GALE%7CA309587934&searchId=R2&userGroupName=anon%7E61bbbab&inPS=true](https://go.gale.com/ps/retrieve.do?tabID=T002&resultListType=RESULT_LIST&searchResultsType=SingleTab&retrievalId=8f6fb502-7809-4a00-90cf-f06b2d717375&hitCount=6&searchType=AdvancedSearchForm&currentPosition=6&docId=GALE%7CA309587934&docType=Report&sort=Relevance&contentSegment=ZONE-Exclude-FT&prodId=AONE&pageNum=1&contentSet=GALE%7CA309587934&searchId=R2&userGroupName=anon%7E61bbbab&inPS=true).
- Magill, M.J.P. and Constantinides, G.M., Portfolio selection with transactions costs. *J. Econom. Theory*, 1976, **13**(2), 245–263.
- Melnik, Y. and Seifried, F.T., Small-cost asymptotics for long-term growth rates in incomplete markets. *Math. Finance*, 2018, **28**(2), 668–711.
- Neuberger, A., The log contract. *J. Portfolio Manag.*, 1994, **20**(2), 74–80.
- Revuz, D. and Yor, M., *Continuous Martingales and Brownian Motion*, 3rd ed., 1999 (Springer: Berlin).
- Rogers, L.C.G., Why is the effect of proportional transaction costs  $O(\delta^{2/3})$ ? In *Mathematics of Finance*, Vol. 351 of *Contemp. Math.*, pp. 303–308, 2004 (American Mathematical Society: Providence, RI).
- Shreve, S.E. and Soner, H.M., Optimal investment and consumption with transaction costs. *Ann. Appl. Probab.*, 1994, **4**(3), 609–692.
- Taksar, M., Klass, M.J. and Assaf, D., A diffusion model for optimal portfolio selection in the presence of brokerage fees. *Math. Oper. Res.*, 1988, **13**(2), 277–294.
- Tang, H. and Xu, X.E., Solving the return deviation conundrum of leveraged exchange-traded funds. *J. Financial Quant. Anal.*, 2013, **48**(1), 309–342.
- Wagalath, L., Modelling the rebalancing slippage of leveraged exchange-traded funds. *Quant. Finance*, 2014, **14**(9), 1503–1511.

## Appendices

### Appendix 1. Admissible Trading Strategies

**DEFINITION A.1** A non-anticipative, self-financing strategy (compare (2)–(3)) is admissible if it is solvent and satisfies sufficient integrability:

- (i) its liquidation value is strictly positive at all times: There exists  $\varepsilon' > \varepsilon$  such that

$$F_t - \varepsilon' \varphi_t^+ S_t \geq 0 \quad \text{a.s. for all } t \geq 0. \quad (\text{A1})$$

- (ii) Let  $\|\varphi\|_t$  denote the total variation of  $\varphi$  on  $[0, t]$ , then

$$\begin{aligned} \mathbb{E} \left[ \int_0^t |\pi_u|^2 du \right] &< \infty, \\ \mathbb{E} \left[ \int_0^t \pi_u \frac{d\|\varphi\|_u}{\varphi_u} \right] &< \infty \quad \text{for all } t \geq 0. \end{aligned}$$

The family of admissible trading strategies is denoted by  $\Phi$ .

Let  $X_t$  be the safe position of an admissible, self-financing trading strategy  $\varphi$ , and  $Y_t = \varphi_t S_t$  the risky position. The following lemma describes the dynamics of the fund value  $F_t = X_t + Y_t$ , the risky weight  $\pi_t = Y_t/F_t$ , and the risky/safe ratio  $\zeta_t = Y_t/X_t$ .

**LEMMA A.2** Let  $\varphi$  be an admissible trading strategy, then

$$\begin{aligned} \frac{dF_t}{F_t} &= r_t dt + \pi_t \left( \sigma_t dB_t - \varepsilon \frac{d\varphi_t^\downarrow}{\varphi_t} \right), \\ \frac{d\pi_t}{\pi_t} &= (1 - \pi_t) \sigma_t dB_t - \pi_t (1 - \pi_t) \sigma_t^2 dt + \frac{d\varphi_t^\uparrow}{\varphi_t} \\ &\quad - (1 - \varepsilon \pi_t) \frac{d\varphi_t^\downarrow}{\varphi_t}, \\ \frac{d\zeta_t}{\zeta_t} &= \sigma_t dB_t + (1 + \zeta_t) \frac{d\varphi_t^\uparrow}{\varphi_t} - (1 + (1 - \varepsilon)\zeta_t) \frac{d\varphi_t^\downarrow}{\varphi_t}. \end{aligned} \quad (\text{A2})$$

Moreover, the functional

$$\Theta_T(\varphi) := \frac{1}{T} \mathbb{E} \left[ \int_0^T \left( \frac{dF_t}{F_t} - r_t dt \right) - \frac{\gamma}{2} \int_0^T \left( \frac{dF_t}{F_t} - \Lambda \frac{dS_t}{S_t} \right)_T \right]$$

can be written as

$$\Theta_T(\varphi) = \frac{1}{T} \mathbb{E} \left[ \int_0^T \frac{\gamma \sigma_t^2}{2} (\pi_t - \Lambda)^2 dt - \varepsilon \int_0^T \pi_t \frac{d\varphi_t^\downarrow}{\varphi_t} \right]. \quad (\text{A3})$$

**Proof** The proof follows from similar arguments as in Guasoni and Mayerhofer (2019, lemma A.2). ■

**REMARK A.3** Note that zero risk premia ( $\mu = 0$ ) imply  $\text{EER}(\varphi) = -\liminf_{T \rightarrow \infty} \Theta_T(\varphi)$  (as in this case  $D_T$  equals  $\int_0^T (\frac{dF_t}{F_t} - r_t dt)$  in expectation, compare (4)). Accordingly, the appendix (except part A) focuses on the equivalent problem of maximizing the functional

$$\begin{aligned} \Theta_\infty(\varphi) &:= \\ \liminf_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \int_0^T \left( \gamma \sigma_t^2 \Lambda \pi_t - \frac{\gamma \sigma_t^2}{2} \pi_t^2 \right) dt - \varepsilon \int_0^T \pi_t \frac{d\varphi_t^\downarrow}{\varphi_t} \right] \\ &= -\text{EER}(\varphi) - \frac{\gamma \Lambda^2 \sigma^2}{2}. \end{aligned} \quad (\text{A4})$$

### Appendix 2. Replication of Leveraged Benchmarks

This section contains a series of propositions, which culminate in a proof of theorem 3.1 in section 4. In particular, the optimal trading strategies are derived for replicating leveraged benchmarks under

transaction costs. Setting

$$G(\zeta) := \frac{\varepsilon}{(1 + \zeta)(1 + (1 - \varepsilon)\zeta)} \quad \text{and}$$

$$h(\zeta) := \gamma \Lambda \left( \frac{\zeta}{1 + \zeta} \right) - \frac{\gamma}{2} \left( \frac{\zeta}{1 + \zeta} \right)^2,$$

the free boundary problem (9)–(11) can be recast as

$$\frac{1}{2} \zeta^2 W''(\zeta) + \zeta W'(\zeta) - h'(\zeta) = 0, \quad (\text{A5})$$

$$W(\zeta_-) = 0, \quad W'(\zeta_-) = 0, \quad (\text{A6})$$

$$W(\zeta_+) = G(\zeta_+), \quad W'(\zeta_+) = G'(\zeta_+). \quad (\text{A7})$$

For the next statement, the third root of a negative number  $a$  is understood as the unique real root of the equation  $\zeta^3 = a$ .

**PROPOSITION A.4** Let  $\gamma > 0$ . For sufficiently small  $\varepsilon$ , the free boundary problem (A5)–(A7) has a unique solution  $(W, \zeta_-, \zeta_+)$ , with  $\zeta_- < \zeta_+$ . The free boundaries have the asymptotic expansion

$$\begin{aligned} \zeta_\pm &= \frac{\Lambda}{1 - \Lambda} \pm \left( \frac{3}{4\gamma} \right)^{1/3} \left( \frac{\Lambda}{(\Lambda - 1)^2} \right)^{2/3} \varepsilon^{1/3} \\ &\quad - \frac{5\Lambda}{2\gamma(\Lambda - 1)^2} \left( \frac{\gamma \Lambda (\Lambda - 1)}{6} \right)^{1/3} \varepsilon^{2/3} + O(\varepsilon). \end{aligned} \quad (\text{A8})$$

**Proof of proposition A.4** The differential equation (A5) is equivalent to  $(\zeta^2 W'(\zeta) - h(\zeta))' = 0$ . As  $\zeta_- \notin \{-1, 0\}$ , any solution of the initial value problem (A5)–(A6) is thus of the form

$$\begin{aligned} W(\zeta_-, \zeta) &= 2 \int_{\zeta_-}^{\zeta} (h(y) - h(\zeta_-)) \frac{dy}{y^2} \\ &= 2\gamma \Lambda \left( \log \left| \frac{\zeta}{1 + \zeta} \right| - \log \left| \frac{\zeta_-}{1 + \zeta_-} \right| \right) \\ &\quad + \gamma \left( \frac{1}{1 + \zeta} - \frac{1}{1 + \zeta_-} \right) + 2h(\zeta_-) \left( \frac{1}{\zeta} - \frac{1}{\zeta_-} \right). \end{aligned}$$

By the terminal conditions (A7) at  $\zeta_+$ , and setting  $\delta = \varepsilon^{1/3}$ ,  $(\zeta_-, \zeta_+)$  satisfy the system

$$\Psi_1(\zeta_-, \zeta_+) := W(\zeta_-, \zeta_+) - \frac{\delta^3}{(1 + \zeta_+)(1 + (1 - \delta^3)\zeta_+)} = 0, \quad (\text{A9})$$

$$\begin{aligned} \Psi_2(\zeta_-, \zeta_+) &:= \frac{2(h(\zeta_+) - h(\zeta_-))}{\zeta_+^2} - \frac{(1 - \delta^3)^2}{(1 + (1 - \delta^3)\zeta_+)^2} \\ &\quad + \frac{1}{(1 + \zeta_+)^2} = 0. \end{aligned} \quad (\text{A10})$$

The unique solution of the free boundary problem can be thus obtained by solving equations (A9)–(A10). Moreover, expanding  $\Psi_{1,2}$  around  $\zeta_\pm = \frac{\Lambda}{1 - \Lambda} + B_{2,1}\delta + O(\delta^2)$  with unknown coefficients  $B_{2,1}$  yields

$$\begin{aligned} \Psi_1(\zeta_\pm(\delta)) &= -\frac{\gamma(1 - \Lambda)^6}{3\Lambda^2} \left( 2B_1^3 - 3B_1^2 B_2 + B_2^3 + \frac{3\Lambda^2}{\gamma(1 - \Lambda)^4} \right) \delta^3 \\ &\quad + O(\delta^4), \\ \Psi_2(\zeta_\pm(\delta)) &= \frac{(B_1 - B_2)(B_1 + B_2)\gamma(\Lambda - 1)^6}{\Lambda^2} \delta^2 + O(\delta^3). \end{aligned}$$

Setting the relevant coefficients of the expansion to zero, it follows that

$$2B_1^3 - 3B_1^2 B_2 + B_2^3 + \frac{3\Lambda^2}{\gamma(1 - \Lambda)^4} = 0, \quad B_1 = -B_2, \quad (\text{A11})$$



which yields

$$B_1 = -\left(\frac{3}{4\gamma}\right)^{1/3} \left(\frac{\Lambda}{(1-\Lambda)^2}\right)^{2/3}. \quad (\text{A12})$$

Finally, it is shown that equations (A9)–(A10) have a unique solution around

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

This is equivalent to

$$\begin{aligned} \Phi_1 &:= \frac{\Psi_1(\zeta_-(\eta_-), \zeta_+(\eta_+))}{\delta^3} = 0, \\ \Phi_2 &:= \frac{\Psi_2(\zeta_-(\eta_-), \zeta_+(\eta_+))}{\delta^2} = 0 \end{aligned}$$

having a unique solution  $\eta_{\pm} := \frac{\zeta_{\pm} - \frac{\Lambda}{1-\Lambda}}{\delta}$  around  $(B_1, B_2)$  for sufficiently small  $\delta$ . As

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

the Implicit Function Theorem for analytic functions (Gunning and Rossi 2009, theorem I.B.4) implies the existence of a unique solution. The higher order coefficients in (A8) are obtained similarly. ■

**PROPOSITION A.5** *Let  $(W, \zeta_-, \zeta_+)$  be the solution of the free boundary problem (A5)–(A7) (provided by proposition A.4). For sufficiently small  $\varepsilon$ , the pair  $(V, \lambda)$  defined by*

$$V(\zeta) := \begin{cases} 0 & \text{for } \zeta < \zeta_-, \\ \int_{\zeta_-}^{\zeta} W(u) du & \text{for } \zeta \in [\zeta_-, \zeta_+], \\ G(\zeta_+) + \int_{\zeta_+}^{\zeta} G(u) du & \text{for } \zeta \geq \zeta_+ \end{cases} \quad \text{and } \lambda := h(\zeta_-),$$

is a solution of the HJB equation

$$\min \left( \frac{x^2}{2} V''(x) - h(x) + \lambda, G(x) - V'(x), V'(x) \right) = 0, \quad x \in \mathbb{R} \setminus [-(1-\varepsilon)^{-1}, 0]. \quad (\text{A13})$$

*Proof of proposition A.5* First, recall that either  $\zeta < -1/(1-\varepsilon)$  (implying  $\zeta < -1$ ) or  $\zeta > -1$  (implying  $\zeta > -1/(1-\varepsilon)$ ); thus for both ranges,

$$G(\zeta) = \frac{\varepsilon}{(1+\zeta)(1+(1-\varepsilon)\zeta)} > 0.$$

We show now the validity of (A13), by dividing the real line into the intervals parts  $\zeta < \zeta_-$ ,  $\zeta \in [\zeta_-, \zeta_+]$  and  $\zeta > \zeta_+$ . For  $\zeta \in [\zeta_-, \zeta_+]$ , equation (A5) implies

$$\begin{aligned} &(\zeta^2/2V''(\zeta) - h(\zeta) + h(\zeta_-))' \\ &= \frac{1}{2}\zeta^2W''(\zeta) + \zeta W'(\zeta) - h'(\zeta) = 0, \end{aligned}$$

and, the first initial condition in (A6) gives

$$(\zeta^2/2V''(\zeta) - h(\zeta) + h(\zeta_-))|_{\zeta=\zeta_-} = \zeta^2/2V''(\zeta)|_{\zeta=\zeta_-} = 0,$$

whence  $\zeta^2/2V''(\zeta) - h(\zeta) + h(\zeta_-) \equiv 0$ , for all  $\zeta \in [\zeta_-, \zeta_+]$ . Second, show  $V'(\zeta) \geq 0$ . As

$$W(\zeta) = V'(\zeta) = \frac{1}{2} \int_{\zeta_-}^{\zeta} (h(\eta) - h(\zeta_-)) \eta^{-2} d\eta, \quad (\text{A14})$$

$W'(\zeta_0) = 0$  if and only if  $h(\zeta_-) = h(\zeta_0)$ . Hence, either  $\zeta_0 = \zeta_-$ , or  $\zeta_0$  satisfies  $\frac{\zeta_0}{1+\zeta_0} = 2\Lambda - \pi_-$ . The first-order asymptotics of (A8)

imply  $\zeta_0 \notin [\zeta_-, \zeta_+]$  for sufficiently small  $\varepsilon$ , and therefore  $W' > 0$  on  $(\zeta_-, \zeta_+]$ , and by (A14) it follows that  $V' \geq 0$  on all of  $[\zeta_-, \zeta_+]$ .

Third, note that  $V' \leq G$  on  $[\zeta_-, \zeta_+]$ . Indeed, the function  $\Psi_1(\zeta) = W(\zeta) - G(\zeta)$  (see (A9)) satisfies

$$\Psi_1(\zeta_-) = -G(\zeta_-) = -(1-\Lambda)^2\varepsilon + O(\varepsilon^{4/3}), \quad \Psi_1(\zeta_+) = 0,$$

hence for sufficiently small  $\varepsilon$ ,  $\Psi_1(\zeta) < 0$  on some interval  $[\zeta_-, \bar{\zeta}]$ , where  $\bar{\zeta} \leq \zeta_+$ . Therefore, it suffices to show that  $\bar{\zeta} = \zeta_+$  to verify the claim.

By contradiction, suppose there exists a sequence  $\delta_k \downarrow 0$  such that for each  $k \in \mathbb{N}$   $\Psi_1(\bar{\zeta}(\delta_k)) = 0$ , and that  $\zeta_-(\delta_k) < \bar{\zeta}(\delta_k) < \zeta_+(\delta_k)$ . Changing to the variable  $u = \frac{\zeta - \zeta_-}{\delta}$ , and introducing the notation  $u_{\pm} = \frac{\zeta_{\pm} - \zeta_-}{\delta}$ ,  $\bar{u} = \frac{\bar{\zeta} - \zeta_-}{\delta}$  shall prove convenient. By selecting, if necessary, a subsequence, one may without loss of generality assume that  $\bar{u}(\delta_k)$  converges, hence it must satisfy  $\lim_{k \rightarrow \infty} \bar{u}(\delta_k) =: B_0 \in [B_1, B_2]$ , where  $B_1$  is defined in (A12), and  $B_2 = -B_1$ . The calculations leading to (A12) therefore entail that  $B_0$  must satisfy (A11) in place of  $B_2$ , i.e.

$$2B_1^3 - 3B_1^2B_0 + B_0^3 + \frac{3\Lambda^2}{\gamma(1-\Lambda)^4} = 0. \quad (\text{A15})$$

With  $B_1$  from (A12), the change of variable  $\xi = -B_0/B_1$  implies  $2 - 3\xi + \xi^3 = 0$ , which has the only solutions 1 and  $-2$ . Therefore, (A15) has the only relevant solution  $B_0 = -B_1 = B_2$ . Set

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

Then  $(u_-(\delta), u^*(\delta))$  satisfies  $\Phi(u_-, u_+) = 0$  near  $(B_1, B_2)$ , for sufficiently small  $\delta$ . By the uniqueness in the proof of proposition A.4,  $u^*(\delta) = u_+(\delta)$ , which contradicts the assumption  $\bar{\zeta} \neq \zeta_+$ .

Consider now  $\zeta \leq \zeta_-$ . As  $V' \equiv 0$ , it suffices to show

$$\varphi(\zeta) := \zeta^2/2V'' - h(\zeta) + h(\zeta_-) = h(\zeta_-) - h(\zeta) \geq 0.$$

Using

$$(h(\zeta) - h(\zeta_-))' = h'(\zeta) = \frac{\gamma}{(1+\zeta)^2} \left( \Lambda - \frac{1}{1+\zeta} \right),$$

the asymptotic expansion (A8) implies  $\varphi'(\zeta_-) < 0$  for sufficiently small  $\varepsilon$ . Furthermore,  $\varphi(\zeta_-) = 0$ , hence  $\varphi(\zeta) > 0$  for all  $\zeta < \zeta_-$  and for any sufficiently small  $\varepsilon$ .

Finally, consider  $\zeta \geq \zeta_+$ . As  $V' = W = G$ , only

$$\begin{aligned} L(\zeta) &:= \zeta^2/2V''(\zeta) - h(\zeta) + h(\zeta_-) \\ &= \frac{\zeta^2}{2}G'(\zeta) - h(\zeta) + h(\zeta_-) \geq 0 \end{aligned}$$

needs to be checked. Note that  $L$  is a rational function and satisfies  $L(\zeta_+) = 0$ , as  $(W, \zeta_-, \zeta_+)$  is a solution of the free boundary problem. Therefore it suffices to show  $L$  has no zeros on  $[\zeta_+, \infty) \setminus [-1/(1-\varepsilon), -1]$ , besides  $\zeta_+$ . The transformation  $z = \frac{\zeta}{1+\zeta}$ ,  $\varepsilon = \delta^3$  leads to a rational function  $F(z, \delta) := L(\zeta(z), \varepsilon(\delta))$ , for which strict positivity has to be shown on  $(\pi_+, 1/\varepsilon)$ . As  $F(\pi_+) = 0$ , polynomial division by  $(z - \pi_+)$  yields

$$F(z, \delta) = \frac{(z - \pi_+)}{(1 - \delta^3 z)^2} g(z),$$

where  $g$  is a cubic polynomial. The further transformation  $y = (z - \pi_+)/(\delta^3 - \pi_+)$  yields  $g(z(y)) = l(y)/\delta^3 + O(\delta)$ , where  $l = \frac{\gamma}{2}[(y-1)^2(\gamma-1)+1]$ . Therefore  $l(y) > 0$  for all  $y \in (0, 1)$  and for any  $\gamma > 0$ , which translates into positivity of  $F$ ,  $g$  and thus  $L$ . ■

**LEMMA A.6** *Let  $\eta_- < \eta_+$  be such that either  $\eta_+ < -1/(1-\varepsilon)$ , or  $-1 < \eta_- < \eta_+ < 0$  or  $\eta_- > 0$ . Then there exists an admissible trading strategy  $\hat{\varphi}$  such that the risky/safe ratio  $\eta_t$  satisfies*

SDE (A2). Moreover,  $(\eta_t, \hat{\varphi}_t^\uparrow, \hat{\varphi}_t^\downarrow)$  is a reflected diffusion on the interval  $[\eta_-, \eta_+]$ .

*Proof* See the first part of Guasoni and Mayerhofer (2019, lemma B.5). ■

The following is the verification of optimality for the trading strategy of lemma A.6 with the trading boundaries in proposition A.4:

**PROPOSITION A.7** Denote by  $\hat{\varphi}$  the trading strategy of lemma A.6 for the free boundaries  $\zeta_\pm$  in proposition A.4, and let

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

Then for any  $t > 0$ , the proportion of wealth  $\pi_t$  invested in the index lies in the interval  $[\pi_-, \pi_+]$ , almost surely, entails no trading whenever  $\pi \in (\pi_-, \pi_+)$ , sells the index at  $\pi_+$  and buys it at  $\pi_-$ . The strategy  $\hat{\varphi}$  is optimal for sufficiently small  $\varepsilon$ , and the value function is

$$\begin{aligned} \Theta_\infty(\hat{\varphi}) &= \max_{\varphi \in \Phi} \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \int_0^T \left( \gamma \sigma_t^2 \Lambda \pi_t - \frac{\gamma}{2} \sigma_t^2 \pi_t^2 \right) dt - \varepsilon \int_0^T \pi_t \frac{d\varphi_t^\downarrow}{\varphi_t} \right] \\ &= \gamma \bar{\sigma}^2 \Lambda \pi_- - \frac{\gamma \bar{\sigma}^2}{2} \pi_-^2. \end{aligned}$$

*Proof of proposition A.7* Recall from proposition A.5 that  $\lambda = h(\zeta_-)$  and  $(V, \lambda)$ , defined from the unique solution of the free boundary problem, is a solution of the HJB equation (A13). For the verification, the proportion  $\pi_t$  of wealth in the index is used, instead of the risky/safe ratio  $\zeta_t$ . With the change of variables  $\zeta = -1 + \frac{1}{1-\pi}$  we introduce the operator

$$(\mathcal{L}f)(\pi) := \frac{1}{2} f''(\pi) \pi^2 (1 - \pi)^2 - f'(\pi) \pi^2 (1 - \pi).$$

By proposition A.5 and the chain rule,  $\hat{V}(\pi) := V(\zeta(\pi))$  satisfies the HJB equation

$$\min \left( \mathcal{L}\hat{V}(\pi) - \hat{h}(\pi) + \lambda, \hat{V}'(\pi), \frac{\varepsilon}{1 - \varepsilon\pi} - \hat{V}'(\pi) \right) = 0, \quad (\text{A16})$$

for  $-\infty < \pi < 1/\varepsilon$ , where  $\hat{h}(\pi) = h(\zeta(\pi)) = \gamma \Lambda \pi - \frac{\gamma}{2} \pi^2$ . By Itô's formula,

$$\hat{V}(\pi_T) - \hat{V}(\pi_0) = \int_0^T \hat{V}'(\pi_t) d\pi_t + \frac{1}{2} \hat{V}''(\pi_t) d\langle \pi \rangle_t \quad (\text{A17})$$

$$\begin{aligned} &= \int_0^T \left( \mathcal{L}\hat{V}(\pi) - \hat{h}(\pi_t) + \lambda \right) \sigma_t^2 dt \\ &\quad + \int_0^T (\hat{h}(\pi_t) - \lambda) \sigma_t^2 dt \end{aligned} \quad (\text{A18})$$

$$+ \int_0^T \hat{V}'(\pi_t) \pi_t (1 - \pi_t) \sigma_t dB_t \quad (\text{A19})$$

$$- \int_0^T \hat{V}'(\pi_t) (1 - \varepsilon \pi_t) \pi_t \frac{d\varphi_t^\downarrow}{\varphi_t} \quad (\text{A20})$$

$$+ \int_0^T \hat{V}'(\pi_t) \pi_t \frac{d\varphi_t^\uparrow}{\varphi_t}. \quad (\text{A21})$$

The first term in line (A18) is non-negative, due to (A16). Furthermore, by (A1) there exists  $\varepsilon' > \varepsilon$  such that  $\pi_t \leq 1/\varepsilon' < 1/\varepsilon$ , for all  $t$ , a.s.. Using (A16) one thus obtains

$$0 \leq \hat{V}'(\pi_t) \leq \frac{\varepsilon \varepsilon'}{\varepsilon' - \varepsilon}, \quad \text{a.s. for all } t \geq 0. \quad (\text{A22})$$

Hence, the integral (A19) is a martingale – in view of the integrability condition in assumption 2.1 – and has zero expectation. Furthermore, (A16) yields

$$0 \leq \hat{V}'(\pi_t) \pi_t (1 - \varepsilon \pi_t) \leq \varepsilon \pi_t,$$

which implies that for (A20) one has

$$- \int_0^T \hat{V}'(\pi_t) (1 - \varepsilon \pi_t) \pi_t \frac{d\varphi_t^\downarrow}{\varphi_t} \geq -\varepsilon \int_0^T \pi_t \frac{d\varphi_t^\downarrow}{\varphi_t}.$$

Finally, (A21) is non-negative, because  $\hat{V}' \geq 0$  due to (A16). Thus, taking the expectation of (A17) gives the estimate

$$\begin{aligned} \frac{1}{T} \mathbb{E}[\hat{V}(\pi_T) - \hat{V}(\pi_0)] &\geq -\frac{\lambda}{T} \int_0^T \sigma_t^2 dt + \frac{1}{T} \mathbb{E} \left[ \int_0^T \sigma_t^2 \hat{h}(\pi_t) dt \right] \\ &\quad - \varepsilon \frac{1}{T} \int_0^T \pi_t \frac{d\varphi_t^\downarrow}{\varphi_t}. \end{aligned} \quad (\text{A23})$$

By (A22),

$$|\hat{V}(\pi_t) - \hat{V}(\pi_0)| \leq |\pi_T - \pi_0| \sup_{0 < u \leq 1/\varepsilon'} |\hat{V}'(u)| \leq \frac{\varepsilon}{\varepsilon' - \varepsilon},$$

therefore  $\lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}[\hat{V}(\pi_T) - \hat{V}(\pi_0)] = 0$ . Hence letting  $T \rightarrow \infty$  in (A23) reveals that  $\Theta_\infty(\varphi) \leq \lambda \bar{\sigma}^2$  for any admissible strategy  $\varphi$ . It is finally shown that this bound is attained by the admissible trading strategy  $\hat{\varphi}$  defined by lemma (A.6). Indeed, let  $\zeta_t$  be the corresponding risky/safe ratio. Itô's formula yields

$$dV(\zeta_t) = V'(\zeta_t) \zeta_t \sigma_t dB_t + 0 - \varepsilon \pi_t \frac{d\varphi_t^\downarrow}{\varphi_t} + (h(\zeta_t) - \lambda) \sigma_t^2 dt.$$

Integration with respect to  $t$  and division by  $T$  yields, in view of (A3),

$$\begin{aligned} \frac{1}{T} \mathbb{E} \left[ \int_0^T \left( \gamma \sigma_t^2 \Lambda \pi_t - \frac{\gamma}{2} \sigma_t^2 \pi_t^2 \right) dt - \varepsilon \int_0^T \pi_t \frac{d\varphi_t^\downarrow}{\varphi_t} \right] \\ = \frac{\lambda}{T} \int_0^T \sigma_t^2 dt + \frac{1}{T} \mathbb{E}[\hat{V}(\pi_T) - \hat{V}(\pi_0)]. \end{aligned}$$

Letting  $T \rightarrow \infty$ , one obtains  $\Theta_\infty(\hat{\varphi}) = \lambda \bar{\sigma}^2$ . ■

### Appendix 3. Performance Evaluation

While the trading strategies derived in the previous section are universal, in the sense that the trading boundaries are independent of (stochastic) volatility of the benchmark's returns, the various performance measures depend on the stationary volatility.

This section obtains explicit expressions for several moments related to the optimal trading strategy which allow in section 4 below to compute the performance statistics of theorem 3.1 are calculated in explicit form, in terms of the trading boundaries. Throughout the section,  $\zeta_t$  is the reflected diffusion†

$$\frac{d\zeta_t}{\zeta_t} = \sigma_t dB_t + dL_t - dU_t \quad (\text{A24})$$

on the interval  $[\zeta_-, \zeta_+]$ , where  $\zeta_- < \zeta_+$ , and where either  $-1 < \zeta_- < \zeta_+ < 0$ ,  $0 < \zeta_- < \zeta_+ < \infty$ , or  $\zeta_+ < -1$ . Accordingly, set  $\pi_\pm = \frac{\zeta_\pm}{1 + \zeta_\pm}$ . Note that we do not require that  $\pi_-, \pi_+$  are optimal in the sense of theorem 3.1; it is only assumed that  $\pi_+ \leq 1/\varepsilon$  to avoid bankruptcy.

The discussion begins with a lemma required in the proofs of several subsequent statements.

**LEMMA A.8** Let  $(f_t)_{t \geq 0}$  be an adapted process satisfying  $0 < \underline{f} \leq |f_t| \leq \bar{f} < \infty$ . Then,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f_t \sigma_t dB_t = 0 \quad \mathbb{P} - \text{a.s.}$$

†  $\zeta_- L_t, \zeta_+ U_t$  are increasing processes (therefore, if  $\zeta_+ < 0$ , then  $L, U$  are decreasing).

*Proof* Denote by  $M_t^f := \int_0^t f_u \sigma_u dB_u$  for each  $t \geq 0$ . The quadratic variation satisfies

$$\langle M^f \rangle_t = \int_0^t f_u^2 \sigma_u^2 du \geq f_-^2 \int_0^t \sigma_u^2 du,$$

which converges  $\mathbb{P}$ -a.s. to infinity as  $t \rightarrow \infty$  (cf. Assumption 2.1). By the Dambis–Dubins–Schwarz Theorem,  $W_t := M_{U_t}^f$ , where  $U_t = \inf\{s > 0 | \langle M^f \rangle_s = t\}$ , is a Brownian motion on  $(\Omega, \mathcal{F}, (\mathcal{F}_{U_t})_t, \mathbb{P})$  and, by construction,  $M_t^f = W_{U_t^{-1}}$ . Furthermore, as  $\langle M^f \rangle$  is strictly increasing,  $U_t^{-1} = \langle M^f \rangle_t$ , hence  $\lim_{T \rightarrow \infty} U_T^{-1}/T \leq \bar{\sigma}^2 \bar{f}^2$  holds  $\mathbb{P}$ -a.s. Thus, by the law of iterated logarithm,  $\lim_{T \rightarrow \infty} M_T^f/T = 0$ ,  $\mathbb{P}$ -a.s. ■

**PROPOSITION A.9** *Let*

$$\lim_{T \rightarrow \infty} \left| \frac{U_T}{T} \right| < \infty, \mathbb{P} - \text{a.s.} \quad (\text{A25})$$

*Then the following  $\mathbb{P}$ -a.s. limits hold:*

$$\lim_{T \rightarrow \infty} \frac{L_T}{T} - \lim_{T \rightarrow \infty} \frac{U_T}{T} = \frac{\bar{\sigma}^2}{2}, \quad (\text{A26})$$

$$\begin{aligned} & \frac{1}{2} \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \sigma_t^2 \pi_t dt \\ &= \zeta_+ \log |\pi_+| \cdot \lim_{T \rightarrow \infty} \frac{U_T}{T} - \zeta_- \log |\pi_-| \cdot \lim_{T \rightarrow \infty} \frac{L_T}{T}, \end{aligned} \quad (\text{A27})$$

$$\frac{1}{2} \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T |\sigma_t \pi_t|^2 dt = \pi_- \cdot \lim_{T \rightarrow \infty} \frac{L_T}{T} - \pi_+ \cdot \lim_{T \rightarrow \infty} \frac{U_T}{T}. \quad (\text{A28})$$

*Proof* First, let  $Y_t = \log |\zeta_t|$ , and  $u := \zeta_+, l := \zeta_-$ . Then  $Y$  is the reflected diffusion

$$Y_t = Y_0 - \int_0^T \frac{\sigma_t^2}{2} dt + \int_0^T \sigma_t dB_t + L_T - U_T$$

on  $[l, u]$ . By lemma A.8,  $\lim_{T \rightarrow \infty} \frac{L_T - U_T}{T} = \frac{\bar{\sigma}^2}{2}$ , and thus, by assumption (A25), equation (A26) follows. Furthermore, applying Itô's Formula to the function  $F(\zeta) := \zeta(\log |\zeta| - 1) - (1 + \zeta)(\log |1 + \zeta| - 1)$  yields

$$\begin{aligned} dF(\zeta_t) &= \frac{\sigma_t^2 \pi_t}{2} dt + \log |\pi_t| \zeta_t \sigma_t dB_t + \log |\pi_-| \zeta_- dL_t \\ &\quad - \log |\pi_+| \zeta_+ dU_t, \end{aligned}$$

and in the limit, (A27) follows using lemma A.8 once again. Similar arguments yield (A28) using the function  $G(\zeta) := \log |1 + \zeta|$  and noting that

$$dG(\zeta_t) = \pi_t \sigma_t dB_t + \pi_- dL_t - \pi_+ dU_t - \frac{1}{2} \sigma_t^2 \pi_t^2 dt. \quad \blacksquare$$

**LEMMA A.10** *If  $\lim_{T \rightarrow \infty} |\frac{U_T}{T}| < \infty$   $\mathbb{P}$ -a.s., then*

$$\lim_{T \rightarrow \infty} \frac{L_T/T}{U_T/T} = \frac{\zeta_+}{\zeta_-} \quad \mathbb{P} - \text{a.s.} \quad (\text{A29})$$

*Proof* By (A2) it follows that

$$\frac{1}{T} (\zeta_T - \zeta_0) = \frac{1}{T} \int_0^T \zeta_t \sigma_t dB_t + \zeta_- L_t - \zeta_+ U_t$$

and thus letting  $T \rightarrow \infty$ , the first integral vanishes almost surely by lemma A.8. The remaining limits hold in view of assumption (A25). ■

A straightforward application of the preceding statements is the following:

**PROPOSITION A.11** *For any (not necessarily optimal) trading strategy associated to  $[\zeta_-, \zeta_+]$  (A25) is satisfied, and  $\mathbb{P}$ -a.s.,*

$$\lim_{T \rightarrow \infty} \frac{L_T}{T} = \frac{\bar{\sigma}^2}{2} \frac{\zeta_+}{\zeta_+ - \zeta_-}, \quad \lim_{T \rightarrow \infty} \frac{U_T}{T} = \frac{\bar{\sigma}^2}{2} \frac{\zeta_-}{\zeta_+ - \zeta_-}, \quad (\text{A30})$$

$$\frac{1}{2} \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \sigma_t^2 \pi_t dt = \frac{\bar{\sigma}^2}{2} \log(\pi_+/\pi_-) \frac{\pi_+ \pi_-}{\pi_+ - \pi_-}, \quad (\text{A31})$$

$$\frac{1}{2} \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T |\sigma_t \pi_t|^2 dt = \frac{\bar{\sigma}^2}{2} \pi_- \pi_+. \quad (\text{A32})$$

**REMARK A.12** This section does not assume that  $\zeta_-, \zeta_+$  are the optimal trading boundaries of theorem 3.1. Invoking optimality, an alternative proof of proposition A.11 only requires the use of two of the four identities (A26)–(A29).

## Appendix 4. Proof of theorem 3.1

*Proof* The part (i) is covered by proposition A.4, and part (ii) is stated in proposition A.7. Part (iii) follows from proposition A.7 by using Remark A.3.

Proof of (iv) (Tracking Difference and Tracking Error): Let  $\varphi_t$  the optimal trading strategy defined by part (ii), and let  $\pi_t$  be the proportion of wealth in the risky asset. Recall the defining equation (5),

$$\text{TrD} := \lim_{T \rightarrow \infty} \left( \frac{1}{T} \int_0^T (\pi_t - \Lambda) \sigma_t dB_t - \frac{\varepsilon}{T} \int_0^T \pi_t \frac{d\varphi_t^\downarrow}{\varphi_t} \right). \quad (\text{A33})$$

By lemma A.8, the first integral in (A33) vanishes almost surely, as  $T \rightarrow \infty$ . Consider now the second term. Substituting

$$dL_t = (1 + \zeta_t) \frac{d\varphi_t^\uparrow}{\varphi_t} \quad \text{and} \quad dU_t = (1 + (1 - \varepsilon)\zeta_t) \frac{d\varphi_t^\downarrow}{\varphi_t}$$

into (A2), the risky/safe ratio  $\zeta$  satisfies equation (A24). Thus

$$\begin{aligned} \text{TrD} &:= - \lim_{T \rightarrow \infty} \frac{\varepsilon}{T} \int_0^T \pi_t \frac{d\varphi_t^\downarrow}{\varphi_t} \\ &= -\varepsilon \frac{\zeta_+}{(1 + \zeta_+)(1 + (1 - \varepsilon)\zeta_+)} \lim_{T \rightarrow \infty} \frac{U_T}{T} \end{aligned}$$

and equation (A30) yields (12). As

$$\begin{aligned} \text{TrE} &= \sqrt{\lim_{T \rightarrow \infty} \left\langle \frac{1}{T} \int_0^T \frac{dF_t}{F_t} - \frac{\Lambda}{T} \int_0^T \frac{dS_t}{S_t} \right\rangle_T} \\ &= \sqrt{\left( \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \sigma_t^2 \pi_t^2 dt - 2\Lambda \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \sigma_t^2 \pi_t dt + \bar{\sigma}^2 \Lambda^2 \right)}, \end{aligned}$$

formula (13) follows from equations (A31)–(A32).

Proof of (v) ( $\bar{s}, \bar{\beta}, R^2$ ): The average volatility  $\bar{s}$  of the optimal portfolio follows from (A32). Moreover, for fixed time horizons  $T$  the regression of  $\frac{1}{T} \int_0^T (\frac{dF_t}{F_t} - r_t dt)$  against  $\frac{1}{T} \int_0^T (\frac{dS_t}{S_t} - r_t dt)$  leads to the estimated slope, or beta, of this regression with the continuous-time approximation

$$\bar{\beta}_T \approx \frac{\langle \int_0^T \frac{dw}{w}, \int_0^T \frac{dS}{S} \rangle_T}{\langle \int_0^T \frac{dS}{S} \rangle_T} = \frac{\int_0^T \pi_t \sigma_t^2 dt}{\int_0^T \sigma_t^2 dt}.$$

According to equations (A31)–(A32), the limit

$$\bar{\beta} := (\bar{\sigma}^2)^{-1} \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \sigma_t^2 \pi_t dt \quad (\text{A34})$$

is well defined. Similarly, the  $R^2$  of the regression, defined as the ratio between the variance of the predicted return and the variance of the realized return, is

$$R^2 = \lim_{T \rightarrow \infty} \frac{(\bar{\sigma}^2)^{-1} \left( \frac{1}{T} \int_0^T \sigma_t^2 \pi_t dt \right)^2}{\frac{1}{T} \int_0^T |\sigma_t \pi_t|^2 dt}. \quad (\text{A35})$$

Plugging equations (A31)–(A32) into equations (A34)–(A35) yields equations (14) and (15).

For part (vi), the asymptotic expansions of the trading boundaries  $\pi_{\pm}$ , (16), follow directly from those for the free boundaries  $\zeta_{\pm}$  in proposition A.4. Furthermore, (16) allows to calculate the respective expansions for EER, TrD, TrE, average volatility  $\bar{\sigma}$ , exposure  $\bar{\beta}$  and  $R^2$  by using their explicit expressions in terms of  $\pi_{\pm}$  provided in theorem 3.1. ■

## Appendix 5. Robustness

In the presence of excess returns ( $\mu_t = \kappa \sigma_t^2$ ,  $\kappa \neq 0$ ), the risky/safe ratio of any admissible trading strategy follows the dynamics

$$\frac{d\zeta_t}{\zeta_t} = \kappa \sigma_t^2 dt + \sigma_t dB_t + (1 + \zeta_t) \frac{d\varphi_t^{\uparrow}}{\varphi_t} - (1 + (1 - \varepsilon)\zeta_t) \frac{d\varphi_t^{\downarrow}}{\varphi_t}$$

and thus the ODE of the free boundary problem is slightly modified to (24), but formulated with the same boundary conditions (10)–(11). Similar proofs as those of theorem 3.1(i) carry over to all points of theorem 6.1. For the sake of completeness, we sketch the derivation of the free boundaries. The solution of (24) subject to the first boundary conditions (10) is of the form

$$W(\zeta_-, \zeta) = \frac{2}{\zeta^2} \int_{\zeta_-}^{\zeta} (h(y) - h(\zeta_-)) \left( \frac{y}{\zeta} \right)^{2\kappa-2} dy.$$

Insisting on the terminal boundary conditions of zero and first order (11) results in the following system of equations for the free boundaries  $\zeta_-, \zeta_+$ ,

$$\begin{aligned} 2\zeta_+^{-2\kappa} \int_{\zeta_-}^{\zeta_+} (h(y) - h(\zeta_-)) y^{2\kappa-2} dy - \frac{\varepsilon}{(1 + \zeta_+)(1 + (1 - \varepsilon)\zeta_+)} &= 0, \\ -4\kappa \zeta_+^{-2\kappa-1} \int_{\zeta_-}^{\zeta_+} (h(y) - h(\zeta_-)) y^{2\kappa-2} dy + \frac{2}{\zeta_+^2} (h(\zeta_+) - h(\zeta_-)) \\ &- \frac{(1 - \varepsilon)^2}{(1 + (1 - \varepsilon)\zeta_+)^2} + \frac{1}{(1 + \zeta_+)^2} = 0. \end{aligned}$$

For sufficiently small  $\varepsilon$ , we thus arrive at the asymptotic expansions

$$\begin{aligned} \zeta_{\pm} &= \frac{\Lambda}{1 - \Lambda} \pm \left( \frac{3}{4\gamma} \right)^{1/3} \left( \frac{\Lambda}{(\Lambda - 1)^2} \right)^{2/3} \varepsilon^{1/3} - \frac{(5\Lambda - 2\kappa)}{2\gamma(\Lambda - 1)^2} \\ &\times \left( \frac{\gamma\Lambda(\Lambda - 1)}{6} \right)^{1/3} \varepsilon^{2/3} + O(\varepsilon), \end{aligned}$$

from which the asymptotics (25) of the trading boundaries  $\pi_{\pm}$  follow.

For figure 6, the exact tracking differences and tracking errors for the optimal trading strategies are used (see theorem 6.1(ii)). To obtain the corresponding formulas, we start with the following modification<sup>†</sup> of proposition A.9:

LEMMA A.13 If

$$\lim_{T \rightarrow \infty} \left| \frac{U_T}{T} \right| < \infty, \quad \mathbb{P} - a.s., \quad (\text{A36})$$

then the following limits hold  $\mathbb{P}$ -a.s.:

$$\lim_{T \rightarrow \infty} \frac{L_T}{T} - \lim_{T \rightarrow \infty} \frac{U_T}{T} = (1 - 2\kappa) \frac{\bar{\sigma}^2}{2}, \quad (\text{A37})$$

$$\begin{aligned} \frac{1}{2} \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T |\sigma_t \pi_t|^2 dt &= \kappa \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \sigma_t^2 \pi_t dt + \pi_- \\ &\cdot \lim_{T \rightarrow \infty} \frac{L_T}{T} - \pi_+ \cdot \lim_{T \rightarrow \infty} \frac{U_T}{T}. \end{aligned} \quad (\text{A38})$$

The appropriate generalization of lemma A.10 is as follows:

LEMMA A.14 If  $\lim_{T \rightarrow \infty} |U_T/T| < \infty$ ,  $\mathbb{P} - a.s.$ , then  $\frac{\lim_{T \rightarrow \infty} L_T/T}{\lim_{T \rightarrow \infty} U_T/T} = |\frac{\zeta_-}{\zeta_+}|^{2\kappa-1}$ .

*Proof* Apply Itô's formula to the functional  $|\zeta_t|^{1-2\kappa}$ . All drift terms vanish. Division by  $T$  and letting  $T \rightarrow \infty$  yields the claim (the Brownian term vanishes due to lemma A.8). ■

PROPOSITION A.15 For any (not necessarily optimal) trading strategy associated with  $[\zeta_-, \zeta_+]$  (A36) is satisfied, and  $\mathbb{P}$ -a.s.,

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{U_T}{T} &= \frac{\frac{\bar{\sigma}^2}{2}(1 - 2\kappa)}{|\zeta_-/\zeta_+|^{2\kappa-1} - 1}, \\ \lim_{T \rightarrow \infty} \frac{L_T}{T} &= \lim_{T \rightarrow \infty} \frac{U_T}{T} + \frac{\bar{\sigma}^2}{2}(1 - 2\kappa). \end{aligned} \quad (\text{A39})$$

The tracking difference is

$$\text{TrD} = -\frac{\bar{\sigma}^2}{2} \zeta_+ G(\zeta_+) \left( \frac{2\kappa - 1}{1 - |\zeta_-/\zeta_+|^{2\kappa-1}} \right). \quad (\text{A40})$$

Furthermore, if the trading strategy is optimal, and if  $\Lambda \neq \kappa$ , then average excess returns (minus transaction costs), and variance of the optimal portfolio's returns are:

$$\begin{aligned} \hat{m} &:= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \kappa \sigma_t^2 \pi_t dt \\ &= \frac{\kappa}{\gamma(\Lambda - \kappa)} \left( \bar{\sigma}^2 \left( h(\zeta_-) - \frac{(1 - 2\kappa)\gamma\pi_-}{2} \right) \right. \\ &\quad \left. - \text{TrD} \left( 1 - \frac{\gamma(\pi_+ - \pi_-)}{\zeta_+ G(\zeta_+)} \right) \right), \\ \hat{\sigma}^2 &:= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T |\sigma_t \pi_t|^2 dt \\ &= \frac{1}{\gamma(\Lambda - \kappa)} \left( -2 \text{TrD} \left( \kappa - \gamma\Lambda \frac{\pi_+ - \pi_-}{\zeta_+ G(\zeta_+)} \right) \right. \\ &\quad \left. + 2\kappa\bar{\sigma}^2 h(\zeta_-) + \gamma\Lambda(1 - 2\kappa)\bar{\sigma}^2\pi_- \right). \end{aligned}$$

<sup>†</sup> Note that in the present setting, equation (A27) generalizes to

$$\begin{aligned} \frac{1}{2} \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \sigma_t^2 \pi_t dt \\ = -\kappa \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \log |\pi_t| \zeta_t \sigma_t^2 dt + \zeta_+ \log |\pi_+| \cdot \lim_{T \rightarrow \infty} \frac{U_T}{T} \end{aligned}$$

$$- \zeta_- \log |\pi_-| \cdot \lim_{T \rightarrow \infty} \frac{L_T}{T},$$

but in view of the logarithmic term ( $\kappa \neq 0$ ), it is not tractable and therefore abandoned. (We therefore invoke in this section optimality of the trading strategy associated with the boundaries  $\zeta_-, \zeta_+$ , cf. remark A.12).

*Proof* Equations (A39) and (A40) are a consequence of equation (A37) and lemma A.14. Also, due to lemma A.2,

$$\begin{aligned}\bar{\sigma}^2 h(\zeta_-) &= -\text{EER} - \frac{\gamma \bar{\sigma}^2 \Lambda^2}{2} \\ &= \gamma \Lambda \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \sigma_t^2 \pi_t \, dt - \frac{\gamma}{2} \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \sigma_t^2 \pi_t^2 \, dt \\ &\quad - \frac{\varepsilon \pi_+}{1 + (1 - \varepsilon) \zeta_+} \lim_{T \rightarrow \infty} \frac{U_T}{T}.\end{aligned}$$

By definition of  $\hat{m}, \hat{s}^2$  and the identity (A40), the last equation may be rewritten as

$$\frac{\gamma \Lambda}{\kappa} \hat{m} - \frac{\gamma}{2} \hat{\sigma}^2 = -\text{TrD} + \bar{\sigma}^2 h(\zeta_-). \quad (\text{A41})$$

Further, rewriting identity (A38) yields  $\hat{m} - \frac{\hat{\sigma}^2}{2} = \pi_+ \lim_{T \rightarrow \infty} \frac{U_T}{T} - \pi_- \lim_{T \rightarrow \infty} \frac{L_T}{T}$ . Thus, by equations (A39) and (A40),

$$\hat{m} - \frac{\hat{\sigma}^2}{2} = -\text{TrD} \cdot \frac{(\pi_+ - \pi_-)}{\zeta_+ G(\zeta_+)} - \frac{\bar{\sigma}^2}{2} \pi_- (1 - 2\kappa). \quad (\text{A42})$$

The two equations (A41), (A42) can now be solved for  $\hat{m}, \hat{\sigma}^2$ . ■