

# Portfolio Optimisation under Transaction Costs with Recursive Preferences\*

Martin Herdegen<sup>†</sup>

David Hobson<sup>‡</sup>

Alex S.L. Tse<sup>§</sup>

January 26, 2026

## Abstract

The Merton investment-consumption problem is fundamental, both in the field of finance and in stochastic control. An important extension of the problem adds transaction costs, which is highly relevant from a financial perspective but also challenging from a control perspective because the solution now involves singular control. A further significant extension takes us from additive utility to stochastic differential utility (SDU), which allows time preferences and risk preferences to be disentangled.

In this paper, we study this extended version of the Merton problem with proportional transaction costs and Epstein-Zin SDU. We fully characterise all parameter combinations for which the problem is well posed (which may depend on the level of transaction costs) and provide a full verification argument that relies on no additional technical assumptions and uses primal methods only.

Even in the special case of (additive) power utility, our arguments are significantly simpler, more elegant and more far-reaching than the ones in the extant literature. This means that we can easily analyse aspects of the problem which previously have been very challenging, including comparative statics, boundary cases which heretofore have required separate treatment and the situation beyond the small transaction cost regime. A key and novel idea is to parametrise consumption and the value function in terms of the shadow fraction of wealth, which may be of much wider applicability.

**Mathematics Subject Classification (2010):** 93E20 , 60H20, 49L20, 91B16, 91G10

**Keywords:** Merton problem, lifetime investment and consumption, transaction costs, Epstein–Zin stochastic differential utility

## 1 Introduction

One of the fundamental problems of economic theory is to determine the optimal consumption and investment behaviour of agents who are endowed with an initial capital, who face a stochastic opportunity set and who are averse to fluctuations of consumption both with respect to time and with respect to stochastic innovations. In a classic formulation of the problem, known as the Merton problem [29, 30], an agent with constant relative risk aversion (CRRA) seeks to maximise the expected, integrated, discounted utility of consumption over the infinite horizon, where consumption is financed from initial wealth and from investments in a risky asset driven by Brownian motion. The important outputs are the optimal instantaneous rate of consumption and the optimal fraction of wealth the agent should invest in the risky asset.

---

\*We thank three anonymous referees for their helpful comments and suggestions.

<sup>†</sup>University of Stuttgart, Department of Mathematics, 70569 Stuttgart, Germany, and University of Warwick, Department of Statistics, Coventry, CV4 7AL, UK. Email: martin.herdegen@isa.uni-stuttgart.de

<sup>‡</sup>University of Warwick, Department of Statistics, Coventry, CV4 7AL, UK. Email: d.hobson@warwick.ac.uk.

<sup>§</sup>University College London, Department of Mathematics, London, WC1H 0AY, UK. Email: alex.tse@ucl.ac.uk.

As well as being a key problem in economics, the Merton problem is also a classical problem in stochastic control. In his seminal paper [29], Merton showed how to derive the Hamilton-Jacobi-Bellman equation for the problem and found the value function and optimal strategy. The solution is as follows: when the problem is well-posed the agent should consume at a rate which is a constant proportion of wealth and invest a constant fraction of wealth in the risky asset, where these constants can be expressed in terms of the parameters of the financial market and the agent's preferences.

Merton's solution of the investment-consumption problem is a beautiful application of stochastic control, and was one of the inspirations behind many subsequent developments of the theory, including the use of Hamilton-Jacobi-Bellman equations, verification arguments and the martingale optimality principle, and the dual approach, see, for example, Fleming and Soner [13], and Pham [31]. But, from a practical perspective, the final answer is unsatisfactory in two important ways. First, the suggestion that agents should invest a constant fraction of wealth in the risky asset is infeasible in markets with transaction costs: if the asset price follows a semi-martingale with non-trivial local martingale term, then to keep a constant fraction of wealth in the risky asset involves a trading strategy with infinite variation. Such a strategy would immediately exhaust the wealth of the agent if there are positive transaction costs. Second, the fraction of wealth that Merton proposes that agents should invest in the risky asset does not match the actual fraction that the representative agent invests in the financial market (as taken from US market data) – this is the equity premium puzzle of Mehra and Prescott [27]. The economics and mathematics literature has responded to these issues in two directions, first by adding transaction costs to the model, and second by assigning more sophisticated preferences to the agent in the form of recursive utility, and its continuous-time analogue stochastic differential utility.

Regarding the first issue with the Merton solution, Magill and Constantinides [26] and Constantinides [9] were the first to introduce transaction costs into the investment-consumption model and to formulate the intuition which underpins the solution, namely that in  $(X, Y\Phi) = (\text{Cash}, \text{Value of investments in the risky asset})$  space<sup>1</sup>, instead of trading to be on the Merton line  $\frac{Y\Phi}{X+Y\Phi} = p$  (where  $p$  is a constant) agents should trade to keep the fraction of wealth in the risky asset in an interval (or equivalently to keep  $(X, Y\Phi)$  in a no-transaction wedge). Davis and Norman [10] formalised the model, and explained that the resulting problem involved singular stochastic control, where the controlled process  $(X_t, Y_t\Phi_t)_{t \geq 0}$  is kept in the no-transaction region using local time pushes on the boundary. For some parameter combinations, [10] showed that the value function could be expressed via the solution of two coupled first order ODEs with free-boundaries, but in terms of understanding the solution, and even of numerically solving the problem, further progress is challenging with this formulation. Shreve and Soner [34] improved on the Davis-Norman solution (in terms of the range of parameter combinations that were covered) using methods from viscosity solutions, but issues with giving a simple, accessible formulation of the value function and optimal strategy remained.<sup>2</sup>

Subsequently, efforts mainly turned to using asymptotic methods to give an expansion for the value function (and thence to derive an expansion for the optimal no-transaction region and consumption rate) in the regime of small transaction costs; with key contributions given by [33, 34, 23, 3, 15, 35, 32, 22]. The small transaction cost literature will be discussed in more detail in Section 7.

Returning to the non-asymptotic case, significant progress was made when Choi et al. [8] and Hobson et al. [22] separately showed how the problem could be decoupled into first, solving a initial-value problem for a first order ODE (and then adjusting the initial value so that an integral equation is satisfied) and second, solving a further first order ODE. Indeed, very important quantities such as the locations of the boundaries of the no-transaction region can be determined from the solution of the first ODE only. Choi et al. [8] took a dual approach based on the shadow price for the asset with transaction costs whereas Hobson et al. [22] use the primal approach. Understanding the final form of the problem (including the set of possible behaviours of the first order ODE, and how they relate to the fundamental properties of

---

<sup>1</sup>Here  $Y$  is the price of the risky asset, and  $\Phi$  is the number of shares held.

<sup>2</sup>Other important earlier papers on the Merton problem with proportional transaction costs include [36, 11] (long-horizon terminal wealth) and [1, 2] (multidimensional case)

the financial problem, eg, when is it well-posed) is simpler in the setting used by Hobson et al. [22], but both approaches can be criticised for appearing to use ‘magic’ to transform the problem into a simpler version — especially in [22] these transformations appear to have no economic motivation.

Regarding the second issue with the Merton solution, the response in the literature is to extend the analysis to allow for recursive preferences. Additive preferences can be criticised for the fact that a single parameter is used to describe the agents’ attitude to fluctuations in consumption due to stochastic risk and their attitude to fluctuations in consumption over time. Recursive preferences postulate that the utility of current consumption depends on the value of future consumption. This makes it possible to separate risk aversion from the elasticity of inter-temporal substitution (EIS). (In fact, for notational convenience we express our problem in terms of the reciprocal of the EIS.) The analogue of recursive utility in continuous-time is stochastic differential utility, and in this setting the natural generalisation of the additive CRRA utility is Epstein-Zin stochastic differential utility (EZ-SDU) as proposed in Epstein and Zin [12].

The extension to EZ-SDU brings several challenges. In the finite-horizon setting, for a given consumption the associated value function is the solution of a backward stochastic differential equation (BSDE). But, over the infinite horizon it is not completely clear how to formulate an analogue of the terminal condition. Then, there are significant issues which arise concerning existence and uniqueness of solutions. Indeed, for many parameter combinations for EZ-SDU uniqueness fails for essentially every consumption stream and we must identify which value process we wish to associate to a given consumption stream (using an economic criterion, rather than an ad-hoc selection). Only when these questions have been resolved is it possible to optimise over admissible investment-consumption strategies (i.e. those which keep wealth non-negative). Fortunately these issues have been considered in depth in Herdegen et al. [18, 19, 20]. Since the real issue is to define a value process given a consumption process, we can ‘piggyback’ on their analysis – the main impact of the introduction of transaction costs is to change the set of attainable consumption streams, and the value associated to a given consumption stream is unchanged.

Although the literature on optimal investment and consumption with recursive utility is growing rapidly (see Chacko and Viceira [4], Kraft et al. [25] and Xing [39]) it focuses on the frictionless case. The exception is Melnyk et al. [28] who look for an asymptotic expansion of the value function in the case of small transaction costs and derive expressions for the asymptotic expansion for the no-transaction region and the optimal consumption, albeit under a restricted set of parameter combinations.

In this article we consider the lifetime investment and consumption problem in a financial market with transaction costs for an agent with recursive preferences. Our main contributions are as follows:

- (a) First, we give a complete solution of the consumption-investment problem in the case of EZ-SDU which is valid for all levels of transaction costs. This includes giving exact conditions for when the problem is well-posed. The methods are an extension of the analysis in the additive case in Hobson et al. [22] and the solution involves finding a family of solutions to a first crossing problem for a first order ODE (indexed by its starting point on a curve) and fixing the solution for a given level of transaction costs by choosing the solution which satisfies an integral equation. The start and end point of this solution determine the boundaries of the no-transaction region. Other important quantities, such as the shadow price of the risky asset and the optimal consumption are given by the solutions of a further first order ODE (this time with fixed start and end points).
- (b) Second, we make the shadow fraction of wealth the base variable in which all other quantities are expressed. This is a key and novel idea that simplifies the analysis considerably.<sup>3</sup> We believe that using the shadow fraction of wealth as base variable will lead to significant simplification in other problems and settings, too. Three immediate candidate examples are the Merton problem with proportional transaction costs and dividends, the Merton problem with proportional transactions cost and different interest rates for buying and selling or the Merton problem in higher dimensions, where only one stock pays proportional transaction costs.

---

<sup>3</sup>The shadow fraction of wealth (called the shadow risky portfolio weight in [16]) has appeared in [16] and [14] before but these papers do not make it the underlying variable in their analysis, which in their case is the centered log stock-cash ratio.

- (c) Third, we ‘break the magician’s code’ and give interpretations to many of the quantities which arose ‘by magic’ in the analysis in [22]. There, the transformations lead to a great simplification due to an act of serendipity; here, we explain where they come from and how they may be interpreted. Thus  $q$ ,  $n(q)$  and  $m(q)$  in [22] are here given interpretations as the shadow fraction of wealth, the optimal consumption rate per unit of wealth as a function of the shadow fraction of wealth, and the optimal consumption rate in a frictionless Merton model if the agent invests a constant fraction  $q$  of wealth in the risky asset.
- (d) Fourth, use of the shadow fraction of wealth as a key variable simplifies many of the arguments. At points in the analysis we have to consider a highly non-standard ODE which passes through a singular point (and has multiple solutions to the left of this point, and just one to the right). In our coordinate system this analysis becomes relatively straightforward (whereas in other systems understanding the phase-space of the solutions can be difficult). One implication is that we only need a one-dimensional version of Tanaka’s results to define the underlying processes via a local time, whereas to date much of the literature has been forced to rely on a multidimensional version. Further, we can treat several boundary cases as part of the main analysis, including the case when one of the boundaries of the no-transaction region corresponds to a portfolio with no cash component, and then the leading term in a small-transaction cost expansion for the locations of the boundary of the no-transaction region is of a different order to the standard case.
- (e) Fifth, we show how the risk aversion  $R$  mainly governs the optimal investment strategy and the no-transaction region ( $S$ , the reciprocal of the EIS, only appears indirectly through the frictionless consumption rate in any of the first three terms of the expansion in small transaction costs for the boundaries of the no-trade interval) and  $S$  mainly governs the rate of consumption. Further, (following on from Janeček and Shreve [23] in the additive case  $R = S$  and Melnyk et al. [28] for EZ-SDU) we show that consumption may increase or decrease with the introduction of small transaction costs, and we give an explanation which shows that the introduction of transaction costs will increase the instantaneous consumption rate if  $S$  is small and decrease it if  $S$  is large.
- (f) Sixth, we investigate analytically the comparative statics with respect to the risk aversion parameter  $R$  and the reciprocal of EIS  $S$ . One of the advances of this paper is that we do not (only) consider the additive case, and therefore it is possible to disentangle the impact of the risk aversion from the impact of the EIS. We find that in many circumstances (for example, if  $R < 1$  or  $R \geq 2$ ) the boundaries of the no-transaction region are monotonic in the risk aversion parameter in the sense that the agent with higher risk aversion buys and sells the risky asset when the fraction of wealth in the risky asset is at a lower threshold than an agent with lower risk aversion investing and consuming in the same market. However, this is not universal and for  $R \in (1, 2)$  this monotonic relationship can fail. The dependence of the no transaction wedge on  $S$  is less clear cut, and whether the no-transaction region moves towards or away from a cash portfolio depends on a crucial combination of parameters, see Section 8.

Along the way we give several other interesting, original and important results:

- (a) We give precise statements for when the problem under EZ-SDU is well posed for all levels of transaction costs, when it is ill-posed for all levels of transaction costs, and when it is well-posed for some levels of transaction costs and ill-posed for others. In the latter case we give an expression for the threshold level of transaction costs at which the problem becomes ill-posed.
- (b) Unlike most other authors we allow for different transaction costs on buying and selling; often the literature assumes that these costs are the same, or that the transaction cost on selling is zero. Whilst there is a straightforward transformation between the various parameterisations of transaction costs, covering the general case facilitates an easy comparison to the individual cases. It also highlights

that, for small transaction costs, the individual transaction costs only enter the expansion at a higher order level, and the first few terms only involve the round-trip transaction cost.

- (c) We show that if the Merton line lies in the first quadrant (i.e. in the case of zero transaction costs the optimal portfolio involves a long position in both the risky asset and the risk-less asset) then it is always inside the no-transaction region; however if the Merton line is not in the first quadrant then it can lie outside the no transaction wedge (and thus becomes a very poor approximation of a reasonable investment strategy).
- (d) We show that if (for large enough transaction costs) the positions corresponding to all wealth being held in the risky asset lie inside the no-transaction region then (for large transaction costs) the location of the sale-boundary on the no-transaction region is independent of the transaction cost for buying.
- (e) We give the asymptotic expansion for the value function and other quantities in the small transaction cost regime to a higher order than has been given elsewhere.

The remainder of the paper is structured as follows. In Section 2 we introduce stochastic differential utility and discuss the formulation of the investment consumption problem under Epstein-Zin SDU over the infinite horizon. In Section 3 we solve the frictionless investment-consumption problem for EZ-SDU. The arguments are extended to the case with frictions in Sections 4 and 5 with Section 4 focussing on the heuristics and Section 5 concentrating on the rigorous statements. Section 6 gives some illustrative numerical examples and Section 7 covers the asymptotics in the small transaction regime. In Section 8 we consider the comparative statics with respect to the two key parameters of EZ-SDU, namely the risk aversion  $R$  and the reciprocal of the EIS  $S$ . Some of the more technical material is relegated to the appendices.

## 2 Epstein-Zin Stochastic Differential Utility

Let  $R, S \in (0, \infty) \setminus \{1\}$  and  $\delta \in \mathbb{R}$  and set  $\mathbb{V} = (1 - R)[0, \infty]$ . The Epstein-Zin (EZ) aggregator (Epstein and Zin [12]) is the function  $g_{EZ} : \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{V} \rightarrow \mathbb{V}$ , given by

$$g_{EZ}(t, c, v) := e^{-\delta t} \frac{c^{1-S}}{1-S} ((1-R)v)^{\frac{S-R}{1-R}}. \quad (2.1)$$

Here,  $R \in (0, \infty) \setminus \{1\}$  denotes the coefficient of *relative risk aversion*,  $S \in (0, \infty) \setminus \{1\}$  denotes the reciprocal of the elasticity of inter-temporal substitution (EIS) and  $\delta \in \mathbb{R}$  denotes some discount parameter.

It is convenient to introduce the parameters  $\theta := \frac{1-R}{1-S}$  and  $\rho = \frac{S-R}{1-R} = 1 - \frac{1}{\theta}$ , so that (2.1) simplifies to

$$g_{EZ}(t, c, v) = e^{-\delta t} \frac{c^{1-S}}{1-S} ((1-R)v)^\rho.$$

Note that  $\theta = 1$  (or equivalently,  $\rho = 0$ ) corresponds to the case  $R = S$ , which is additive power utility.

Given a nonnegative progressively measurable stochastic process  $C = (C_t)_{t \geq 0}$ , called a *consumption stream*, we seek to find a utility process  $V^C = (V_t^C)_{t \geq 0}$  that solves the equation

$$V_t = \mathbb{E} \left[ \int_t^\infty e^{-\delta s} \frac{C_s^{1-S}}{1-S} ((1-R)V_s)^\rho \, ds \mid \mathcal{F}_t \right]. \quad (2.2)$$

The problem is to find  $\hat{C} \in \mathcal{C}$  such that

$$V_0^{\hat{C}} = \sup_{C \in \mathcal{C}} V_0^C,$$

where  $V^C = (V_t^C)_{t \geq 0}$  denotes an appropriate solution of (2.2), and  $\mathcal{C}$  is an appropriate class of consumption streams.

Throughout the paper, we make the following standing assumption

**Standing Assumption 1.** *We assume  $\theta \in (0, \infty)$ .*

The necessity of this standing assumption can be seen from multiplying (2.2) on both sides by  $(1 - R)$ . Then, with  $\tilde{V} = (1 - R)V$ , which implicitly must be positive for it to be possible to take arbitrary powers, we have  $\tilde{V}_t = \theta \mathbb{E}[\int_t^\infty e^{-\delta s} C_s^{1-S} \tilde{V}_s^\rho ds | \mathcal{F}_t]$  and we must have  $\theta > 0$  for the two sides of this equation to have the same sign. See Herdegen et al. [18] for further discussion of this issue.

For  $\theta \in (0, 1]$ , we have existence and uniqueness of a utility process  $V^C = (V_t^C)_{t \geq 0}$  solving (2.2) (in a generalised sense) for all consumption streams, see Herdegen et al. [18]. By saying that  $V^C$  solves (2.2) in a generalised sense we mean that we allow  $(1 - R)V^C$  to take the value  $+\infty$  provided that  $\mathbb{E}[\int_t^\infty e^{-\delta s} C_s^{1-S} ((1 - R)V_s^C)^\rho ds | \mathcal{F}_t]$  also takes the value  $+\infty$ . We denote by  $\mathcal{P}_+$  the collection of all nonnegative progressively measurable processes.

**Theorem 2.1** ([19, Theorem 6.9]). *Suppose  $\theta \leq 1$ . For each  $C \in \mathcal{P}_+$ , there exists a unique generalised utility process  $V^C = (V_t^C)_{t \geq 0}$  solving (2.2) (in a generalised sense).*

By contrast, if  $\theta > 1$ , the situation is more involved because (2.2) has multiple solutions. For example,  $V \equiv 0$  is always a solution to (2.2) when  $\rho > 0$ . To get uniqueness we restrict ourselves to so-called proper solutions; see Herdegen et al. [20] for a discussion of the issues.

**Definition 2.2.** Suppose  $C \in \mathcal{P}_+$  and suppose that  $V = (V_t)_{t \geq 0}$  is a solution to (2.2). Then  $V$  is called a *proper solution* if  $(1 - R)V_t > 0$  on  $\{\mathbb{E}[\int_t^\infty e^{-\delta s} C_s^{1-R} ds | \mathcal{F}_t] > 0\}$  up to null sets for all  $t \geq 0$ .

Denote by  $\mathbb{UP}^*$  the collection of right-continuous processes  $C \in \mathcal{P}_+$  for which there exists a unique proper utility process  $V^C = (V_t^C)_{t \geq 0}$  solving (2.2). The results of Herdegen et al. [20] (especially Proposition 4.8) show that is a wide class of processes. When  $\theta > 1$ , we will optimise over consumption streams in  $\mathbb{UP}^*$ .

### 3 Frictionless problem

We consider a Black–Scholes–Merton market consisting of a riskless asset with interest rate  $r \in \mathbb{R}$ , whose price process  $Y^0 = (Y_t^0)_{t \geq 0}$  is given by  $Y_t^0 = y_0^0 e^{rt}$  (for  $y_0^0 > 0$ ), and a risky asset given by geometric Brownian motion with drift  $\mu \in \mathbb{R}$ , volatility  $\sigma > 0$  and initial value  $y_0 > 0$ , whose price process  $Y = (Y_t)_{t \geq 0}$  is given by  $Y_t = y_0 \exp(\sigma B_t + (\mu - \frac{1}{2}\sigma^2)t)$ , for a Brownian motion  $B = (B_t)_{t \geq 0}$ .

At each time  $t \geq 0$ , the agent chooses to consume at a rate  $C_t \in \mathbb{R}_+$  per unit time, and to hold  $\Phi_t \in \mathbb{R}$  numbers of shares of the risky asset. If we denote total wealth at time  $t$  by  $Z_t$ , the number of shares invested into the riskless asset is given by  $\Phi_t^0 = \frac{Z_t - \Phi_t Y_t}{Y_t^0}$ . Then, by the self-financing condition, the (total) wealth process of the agent  $Z = (Z_t)_{t \geq 0}$  satisfies the SDE

$$dZ_t = \Phi_t^0 dY_t^0 + \Phi_t dY_t - C_t dt = \Phi_t Y_t \sigma dB_t + (Z_t r + \Phi_t Y_t (\mu - r) - C_t) dt, \quad (3.1)$$

subject to the initial condition  $Z_0 = z$ , where  $z > 0$  is the agent's initial wealth.

**Remark 3.1.** It is customary in the literature to parameterise trading strategies in terms of fractions of wealth (rather than numbers of shares). However, this parametrisation is less suitable for the problem with transaction costs. For this reason, and to make an easy comparison with the frictionless case, we also parameterise trading in the frictionless problem in numbers of shares.

**Definition 3.2.** Given  $z > 0$ , an *admissible investment-consumption strategy* (for  $z$ ) is a pair  $(\Phi, C) = (\Phi_t, C_t)_{t \geq 0}$  of progressively measurable processes, where  $\Phi$  is real-valued and  $C$  is nonnegative, such that the SDE (3.1) has a unique strong solution  $Z^{z, \Phi, C} = (Z_t^{z, \Phi, C})_{t \geq 0}$  that is  $\mathbb{P}$ -a.s. nonnegative. We denote the set of admissible investment-consumption strategies for  $z > 0$  by  $\mathcal{A}_0(z)$ .

Since the utility value associated to an admissible investment-consumption strategy only depends upon consumption, and not upon the amount invested into the risky asset, we also introduce the following definition:

**Definition 3.3.** Given  $z > 0$ , an *attainable consumption stream* (for  $z$ ) is a nonnegative progressively measurable stochastic process  $C = (C_t)_{t \geq 0}$  for which there exists a progressively measurable process  $\Phi = (\Phi_t)_{t \geq 0}$  such that  $(\Phi, C) \in \mathcal{A}_0(z)$ . Denote the set of attainable consumption streams for  $z > 0$  by  $\mathcal{C}_0(z)$ .

The frictionless problem is to find  $\hat{C} \in \mathcal{C}_0(z)$  such that

$$V_0^{\hat{C}} = \sup_{C \in \mathcal{C}_0^*(z)} V_0^C, \quad (3.2)$$

where, if  $\theta \in (0, 1]$ ,  $V^C = (V_t^C)_{t \geq 0}$  denotes the unique generalised utility process from Theorem 2.1 solving (2.2) (in a generalised sense) and  $\mathcal{C}_0^*(z) = \mathcal{C}_0(z)$ , and, if  $\theta \in (0, 1]$ ,  $V^C = (V_t^C)_{t \geq 0}$  denotes the unique proper solution of (2.2) and  $\mathcal{C}_0^*(z) = \mathcal{C}_0(z) \cap \mathbb{UP}^*$ .

### 3.1 Solution for constant proportional strategies

In order to motivate our solution ansatz in the case with frictions, it is insightful to study the optimisation problem (3.2) for constant proportional strategies only. To this end, consider the investment-consumption strategy  $(\Phi, C)$ , where  $\Phi = q \frac{Z}{Y}$  and  $C = mZ$  for  $q \in \mathbb{R}$ ,  $m \in (0, \infty)$ . Then, the wealth process  $Z^{z, \Phi, C} = Z = (Z_t)_{t \geq 0}$  from (3.1) is given by

$$Z_t = z \exp \left( q\sigma B_t + \left( r + q\sigma\lambda - m - \frac{q^2\sigma^2}{2} \right) t \right),$$

where  $\lambda := \frac{\mu - r}{\sigma}$  denotes the *market price of risk*. We make the ansatz

$$V_t = ae^{-bt} \frac{Z_t^{1-R}}{1-R} \quad (3.3)$$

for some constants  $a, b \in \mathbb{R}$  to be determined. Substituting this ansatz into (2.2), and recalling that  $\rho = \frac{S-R}{1-R}$  yields

$$V_t = \mathbb{E} \left[ \int_t^\infty e^{-\delta s} \frac{(mZ_s)^{1-S}}{1-S} \left( ae^{-bs} Z_s^{1-R} \right)^\rho ds \middle| \mathcal{F}_t \right] = a^\rho \frac{m^{1-S}}{1-S} \mathbb{E} \left[ \int_t^\infty e^{-(\delta+b\rho)s} Z_s^{1-R} ds \middle| \mathcal{F}_t \right]. \quad (3.4)$$

A straightforward calculation gives

$$\mathbb{E} \left[ e^{-(\delta+b\rho)s} Z_s^{1-R} \middle| \mathcal{F}_t \right] = e^{-(\delta+b\rho)t} Z_t^{1-R} e^{-\theta \tilde{H}_b(q, m)(s-t)}, \quad s > t,$$

where  $\tilde{H}_b : \mathbb{R} \times (0, \infty) \mapsto \mathbb{R}$  is given by  $\tilde{H}_b(q, m) = \frac{1}{\theta}(\delta + b\rho) + (S-1) \left( r + \lambda\sigma q - m - \frac{q^2\sigma^2}{2} R \right)$ . Provided that  $\tilde{H}_b(q, m) > 0$ , (3.4) simplifies to

$$V_t = \frac{a^\rho m^{1-S}}{\tilde{H}_b(q, m)} e^{-(\delta+b\rho)t} \frac{Z_t^{1-R}}{1-R}.$$

Our ansatz (3.3) now implies that  $\frac{a^\rho m^{1-S}}{\tilde{H}_b(q, m)} = a$  and  $\delta + b\rho = b$ . We deduce that  $b = \delta\theta$  and that  $a = \left( \frac{m^{1-S}}{\tilde{H}_b(q, m)} \right)^\theta$ . If we substitute this value of  $b$  into  $\tilde{H}_b$  and define  $H = \tilde{H}_{\delta\theta}$  then we have

$$H(q, m) = \delta + (S-1) \left( r + \lambda\sigma q - m - \frac{q^2\sigma^2}{2} R \right). \quad (3.5)$$

and provided  $H(q, m) > 0$ ,

$$V_t = \left( \frac{m^{1-S}}{H(q, m)} \right)^\theta e^{-\delta\theta t} \frac{Z_t^{1-R}}{1-R}. \quad (3.6)$$

If we now maximise (3.6) for fixed  $q$  over  $m$ , we obtain that the optimal consumption rate expressed per unit of total wealth is given by  $m = m(q)$  where

$$m(q) = \frac{\delta}{S} - \frac{1-S}{S} \left( r + \lambda\sigma q - \frac{q^2\sigma^2}{2} R \right) = \frac{\alpha}{S} - \frac{1-S}{S} \lambda\sigma q + R \frac{1-S}{S} \frac{\sigma^2}{2} q^2, \quad (3.7)$$

where  $\alpha := \delta - r(1-S)$  denotes the *impatience rate*, see [17, p.1237] or [18, p.140] for a justification and discussion of this terminology. Plugging (3.7) into (3.6) yields

$$V_t = e^{-\delta\theta t} \frac{Z_t^{1-R}}{1-R} m(q)^{-\theta S} \quad (3.8)$$

Differentiating (3.8) with respect to  $q$  and setting the derivative equal to zero, gives that the turning point of  $m$  is at  $q_M = \frac{\lambda}{R\sigma}$ . Clearly,  $q_M$  is the famous *Merton ratio*. Thus, in frictionless market the optimal fraction of wealth to invest in the risky asset is independent of the elasticity of intertemporal substitution.

The above analysis yields a candidate value function and a candidate optimal strategy. It is well known that:

**Theorem 3.4** ([28, 19, 20]). *Consider the frictionless market and recall that  $q_M = \frac{\lambda}{\sigma R}$ . Suppose that  $m_M := m(q_M) > 0$ . Then the frictionless optimal-investment consumption problem is well-posed, and the value function is given by*

$$V_0 = \sup_{C \in \mathcal{C}_0^*(z)} V_0^C = \frac{z^{1-R}}{1-R} m_M^{-\theta S}$$

*The optimal strategy is  $(C^* = m_M Z, \Phi = \frac{\lambda}{\sigma R} \frac{Z}{Y})$ , i.e.,  $q_M$  is the optimal fraction of wealth and  $m_M$  is the optimal consumption rate.*

*If  $m_M \leq 0$  then the problem is ill-posed, i.e.,  $V_0 = +\infty$  if  $R < 1$  or  $V_0 = -\infty$  if  $R > 1$ .*

**Remark 3.5.** The shape of  $m = m(q)$  will play a key role in later arguments. It is determined by the shape of  $H(q, m)$  as given in (3.5), and as a function parameterised by  $R$  and  $S$ , the shape of  $H$  is determined by the sign of  $(S - 1)$ . In particular, when  $S < 1$ ,  $H$  is convex in  $q$ , and this property is inherited by  $m(q)$  (see Section 8.3 for details of how this inheritance works). Then,  $m(q)$  has a *minimum* over  $q \in \mathbb{R}$  at  $q_M$ . Conversely, when  $S > 1$ ,  $H$  is concave in  $q$ , this property is inherited by  $m$ , and  $m$  has a *maximum* (over  $q \in \mathbb{R}$  such that  $H(q, m) > 0$ ) at  $q_M$ . Indeed, given the sign of  $V$ , this must be the case for  $V$  in (3.8) to have an interior *maximum* in  $q$ .

Economically, the explanation for the shape of  $m$  is as follows. Consider the impact of investing a suboptimal fraction of wealth in the risky asset upon the rate of instantaneous consumption (as opposed to the level of future consumption). Using the wrong  $q$  has two opposite effects, an ‘income effect’ and a ‘substitution effect’. Because of the sub-optimality of  $q$ , the agent is less wealthy in certainty equivalent terms and this leads to an ‘income effect’: a reduction in consumption, and hence a fall in instantaneous consumption rate (as a fraction of current wealth). At the same time, the reduction in certainty equivalent is primarily a reduction in the expected value of future consumption, rather than current consumption, and this leads to a ‘substitution effect’—an increase of the current consumption rate from bringing forward consumption from the future to the present time.

When  $S$ , the reciprocal of the EIS, is relatively small (less than one)—equivalently the EIS is relatively large (greater than one)—the desire to smooth consumption over time is small (see Herdegen et al. [18, Section 4.1]). This implies that the ‘substitution effect’ dominates and  $m$  is concave. By contrast, when  $S$  is relatively large (greater than one), the desire to seek smooth consumption over time is increased. This implies that the ‘income effect’ dominates and  $m$  is convex.

We return to this issue in Remark 7.6(c) and Section 8.3.

## 4 Problem with friction: notation and heuristics

Now we consider the problem with transaction costs. We assume that in order to buy one unit of the risky asset at time  $t$  one has to pay  $\gamma^\uparrow Y_t$ , and that the receipts from selling one unit of the risky asset at time  $t$  are given by  $\frac{1}{\gamma^\downarrow} Y_t$ . Here  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$  denote the transaction cost parameters for buying and selling. When  $\gamma^\uparrow \gamma^\downarrow > 1$ , we have to restrict ourselves to trading strategies  $\Phi = (\Phi_t)_{t \geq 0}$  (in numbers of shares) that are of finite variation. In this case, we denote by  $\Phi^\uparrow = (\Phi_t^\uparrow)_{t \geq 0}$  and  $\Phi^\downarrow = (\Phi_t^\downarrow)_{t \geq 0}$  the increasing and decreasing part of  $\Phi$ , respectively, so that  $\Phi = \Phi^\uparrow - \Phi^\downarrow$  where  $\Phi^\uparrow$  and  $\Phi^\downarrow$  are increasing processes. Moreover, in this case, it is useful not to describe the total wealth process but rather to consider the cash process and the holdings in the risky asset separately.

When choosing the investment-consumption strategy  $(\Phi, C)$ , the corresponding cash process  $X = (X_t)_{t \geq 0}$  satisfies the dynamics

$$dX_t = rX_t dt - C_t dt - \gamma^\uparrow Y_t d\Phi_t^\uparrow + \frac{Y_t}{\gamma^\downarrow} d\Phi_t^\downarrow. \quad (4.1)$$

This is paired with the initial condition

$$X_0 = x_0 - \gamma^\uparrow y(\Phi_0 - \phi_0)^+ + \frac{y}{\gamma^\downarrow}(\Phi_0 - \phi_0)^-, \quad (4.2)$$

where  $x_0 \in \mathbb{R}$  denotes the initial cash position and  $\phi_0$  denotes the initial stock position prior to trading at time 0. Note that we allow for an initial block trade just before time zero (i.e. at time  $0-$ ) that is not encoded in the strategy  $\Phi = (\Phi_t)_{t \geq 0}$ . This allows us to consider  $\Phi$  to be a right-continuous process.

For  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$ , we denote by

$$\mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow} := \left\{ (x, y, \phi) \in \mathbb{R} \times (0, \infty) \times \mathbb{R} : x + \frac{1}{\gamma^\downarrow} \phi^+ y - \gamma^\uparrow \phi^- y \geq 0 \right\}$$

the *solvency region*. We denote by  $\mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ$  its interior, which satisfies

$$\mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ = \left\{ (x, y, \phi) \in \mathbb{R} \times (0, \infty) \times \mathbb{R} : x + \frac{1}{\gamma^\downarrow} \phi^+ y - \gamma^\uparrow \phi^- y > 0 \right\}.$$

By analogy with Definitions 3.2 and 3.2 we can define what it means for an investment-consumption strategy  $(\Phi, C)$  to be admissible for  $(\gamma^\uparrow, \gamma^\downarrow)$ -transaction costs, and a consumption stream  $C$  to be attainable for  $(\gamma^\uparrow, \gamma^\downarrow)$ -transaction costs.

**Definition 4.1.** Given  $(x, y, \phi) \in \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ$ , a  $(\gamma^\uparrow, \gamma^\downarrow)$ -admissible investment-consumption strategy is a pair  $(\Phi, C) = (\Phi_t, C_t)_{t \geq 0}$  of progressively measurable processes, where  $\Phi$  is real-valued and of finite variation and  $C$  is nonnegative, such that the SDE (4.1) has a unique strong solution  $X^{x, y, \phi, \Phi, C} = (X_t^{x, y, \phi, \Phi, C})_{t \geq 0}$  such that  $(X_t^{x, y, \phi, \Phi, C}, \Phi_t, Y_t) \in \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}$   $\mathbb{P}$ -a.s. for all  $t \geq 0$ . We denote the set of  $(\gamma^\uparrow, \gamma^\downarrow)$ -admissible investment-consumption strategies for  $(x, y, \phi) \in \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ$  by  $\mathcal{A}_{\gamma^\uparrow, \gamma^\downarrow}(x, \phi, y)$ .

**Definition 4.2.** Given  $(x, y, \phi) \in \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ$ , a  $(\gamma^\uparrow, \gamma^\downarrow)$ -attainable consumption stream (for  $(x, y, \phi)$ ) is a nonnegative progressively measurable stochastic process  $C = (C_t)_{t \geq 0}$  for which there exists a progressively measurable process  $\Phi = (\Phi_t)_{t \geq 0}$  such that  $(\Phi, C) \in \mathcal{A}_{\gamma^\uparrow, \gamma^\downarrow}(x, \phi, y)$ . Denote the set of attainable consumption streams for  $(x, y, \phi) \in \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ$  by  $\mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}(x, y, \phi)$ .

**Remark 4.3.** It is straightforward to check that for all  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$ ,  $\mathcal{A}_{\gamma^\uparrow, \gamma^\downarrow}(x, \phi, y) \subseteq \mathcal{A}_0(x + \phi y)$  and  $\mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}(x, \phi, y) \subseteq \mathcal{C}_0(x + \phi y)$ . More generally, if  $\tilde{\gamma}^\uparrow, \tilde{\gamma}^\downarrow$  are such that  $\tilde{\gamma}^\uparrow \geq \gamma^\uparrow$  and  $\tilde{\gamma}^\downarrow \geq \gamma^\downarrow$  then  $\mathcal{C}_{\tilde{\gamma}^\uparrow, \tilde{\gamma}^\downarrow}(x, \phi, y) \subseteq \mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}(x, \phi, y)$ .

For  $(x, y, \phi) \in \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ$ , the problem with friction is to find  $\hat{C} \in \mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}(x, \phi, y)$  such that

$$V_0^{\hat{C}} = \sup_{C \in \mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}^*(x, \phi, y)} V_0^C, \quad (4.3)$$

where, if  $\theta \in (0, 1]$ ,  $V^C = (V_t^C)_{t \geq 0}$  denotes the unique generalised utility process from Theorem 2.1 solving (2.2) (in a generalised sense), and  $\mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}^*(x, \phi, y) = \mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}(x, \phi, y)$ , and if  $\theta > 1$ ,  $V^C = (V_t^C)_{t \geq 0}$  denotes the unique proper utility process solving (2.2) and  $\mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}^*(x, \phi, y) = \mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}(x, \phi, y) \cap \mathbb{UP}^*$ .

#### 4.1 Shadow fraction of wealth and relative shadow price

For a cash process  $X = (X_t)_{t \geq 0}$  corresponding to some  $(\gamma^\uparrow, \gamma^\downarrow)$ -admissible investment-consumption strategy  $(\Phi, C)$ , define the *fraction of wealth*  $P = (P_t)_{t \geq 0}$  invested into the risky asset by

$$P_t = \frac{\Phi_t Y_t}{X_t + \Phi_t Y_t}.$$

Let  $\tilde{Y} = (\tilde{Y}_t)_{t \geq 0}$  denote a *shadow stock price process*, which takes values in  $Y[\frac{1}{\gamma^\downarrow}, \gamma^\uparrow]$ .

Then the process  $Q = (Q_t)_{t \geq 0}$  defined by

$$Q_t = \frac{\Phi_t \tilde{Y}_t}{X_t + \Phi_t \tilde{Y}_t},$$

denotes the corresponding *shadow fraction of wealth*. For future reference, we note that the ratio of shadow wealth to real wealth can be expressed as

$$\frac{X + \Phi \tilde{Y}}{X + \Phi Y} = \frac{\frac{X}{X + \Phi Y}}{\frac{X}{X + \Phi \tilde{Y}}} = \frac{1 - P}{1 - Q}. \quad (4.4)$$

We make the ansatz that the *relative shadow price*  $\tilde{Y}/Y$  can be written in terms of the shadow fraction of wealth, and thus is given by

$$\frac{\tilde{Y}}{Y} = \kappa(Q)$$

for some function  $\kappa$  to be determined later. Note that by (4.4),

$$\kappa(Q) = \frac{\tilde{Y}}{Y} = \frac{\Phi \tilde{Y}}{\Phi Y} = \frac{Q(X + \Phi \tilde{Y})}{P(X + \Phi Y)} = \frac{Q/(1 - Q)}{P/(1 - P)}. \quad (4.5)$$

We also make the ansatz that

$$Q = q(P)$$

for some function  $q$  to be determined later. Then (4.5) gives

$$\kappa(q(p)) = \frac{q(p)/(1 - q(p))}{p/(1 - p)} \quad (4.6)$$

These equations may be inverted to express  $p$  in terms of  $q$  in which case we find  $P = p(Q)$  where

$$p(q) = \frac{q}{1 + q(1 - \kappa(q))}, \quad \kappa(q) = \frac{q}{1 - q} \frac{1 - p(q)}{p(q)}. \quad (4.7)$$

Differentiating (4.6) and rearranging yields the key identity

$$p(1 - p)q'(p) = \left( \frac{1}{q(p)(1 - q(p))} - \frac{\kappa'(q(p))}{\kappa(q(p))} \right)^{-1}. \quad (4.8)$$

## 4.2 Solution ansatz

Motivated by the structure (3.8) of the frictionless solution, we make the first ansatz that the value process  $V = (V_t)_{t \geq 0}$  is given by

$$V_t := e^{-\delta\theta t} \frac{(X_t + \Phi_t \tilde{Y}_t)^{1-R}}{1-R} n(Q_t)^{-\theta S} = e^{-\delta\theta t} \frac{(X_t + \Phi_t Y_t \kappa(Q_t))^{1-R}}{1-R} n(Q_t)^{-\theta S},$$

where the function  $n$  is to be determined. Here  $n$  represents the rate of consumption per unit of wealth, when wealth is measured using the shadow stock price.

Thus, setting  $p = \frac{\phi y}{x + \phi y \kappa}$ , we look for a value function of the form  $V = V(I, X, \Phi, Y)$  where

$$V(t, x, \phi, y) = e^{-\delta\theta t} \frac{(x + \phi y \kappa(q(p)))^{1-R}}{1-R} n(q(p))^{-\theta S}. \quad (4.9)$$

Inside the no trade region we make the second ansatz that the value function should be independent of the shadow fraction of wealth. Thus, we assume that

$$0 = \frac{\partial V}{\partial q} = (1-R) \frac{\phi y \kappa'(q)}{(x + \phi y \kappa(q))} V - \theta S \frac{n'(q)}{n(q)} V,$$

which, using  $\frac{\phi y}{x + \phi y \kappa} = \frac{q}{\kappa}$ , is equivalent to

$$\frac{\kappa'(q)}{\kappa(q)} = \frac{S}{1-S} \frac{1}{q} \frac{n'(q)}{n(q)}. \quad (4.10)$$

Using (4.9) together with the second ansatz we can easily compute the first-order partial derivatives of  $V$

$$V_t = -\delta\theta V, \quad (4.11)$$

$$\begin{aligned} xV_x &= (1-R)V \frac{x}{(x + \phi y \kappa(q))} = (1-R)V(1-q), \\ \phi V_\phi &= (1-R)V \frac{\phi y \kappa(q)}{(x + \phi y \kappa(q))} = (1-R)Vq, \\ yV_y &= (1-R)V \frac{\phi y \kappa(q)}{(x + \phi y \kappa(q))} = (1-R)Vq, \end{aligned} \quad (4.12)$$

and then using (4.8) and (4.10),

$$\begin{aligned} y^2 V_{yy} &= (1-R)^2 q^2 V - (1-R)Vq + (1-R)Vp(1-p)q'(p) \\ &= (1-R)V \left( (1-R)q^2 - q + \left( \frac{1}{q(1-q)} - \frac{\kappa'(q)}{\kappa(q)} \right)^{-1} \right) \\ &= (1-R)V \left( (1-R)q^2 - q + \left( \frac{1}{q(1-q)} - \frac{S}{1-S} \frac{1}{q} \frac{n'(q)}{n(q)} \right)^{-1} \right). \end{aligned} \quad (4.13)$$

## 4.3 Martingale optimality principle

Define the process  $M = (M_t)_{t \geq 0}$  by

$$M_t := \int_0^t e^{-\delta s} \frac{C_s^{1-S}}{1-S} ((1-R)V(s, X_s, Y_s, \Phi_s))^\rho ds + V(t, X_t, Y_t \Phi_t).$$

By the martingale optimality principle,  $M$  is a supermartingale for each pair  $(C, \Phi)$  and a martingale for the optimal pair  $(C^*, \Phi^*)$ . Itô's formula and (4.11) – (4.12) and (4.13) give

$$\begin{aligned}
dM_t &= \frac{C_t^{1-S}}{1-S}((1-R)V)^\rho e^{-\delta t} dt + V_t dt + V_x dX_t + V_y dY_t + V_\phi d\Phi_t + \frac{1}{2}V_{yy} d\langle Y \rangle_t \\
&= \left( \frac{C_t^{1-S}}{1-S}((1-R)V)^\rho e^{-\delta t} - \delta\theta V + r(1-R)V(1-Q_t) - V_x C_t + \mu(1-R)Q_t V \right) dt \\
&\quad + \sigma Y_t V_y dB_t + (V_\phi - \gamma^\uparrow V_x Y_t) d\Phi_t^\uparrow + \left( \frac{V_x}{\gamma^\downarrow} Y_t - V_\phi \right) d\Phi_t^\downarrow \\
&\quad + \frac{\sigma^2}{2}(1-R)V \left( (1-R)Q_t^2 - Q_t + \left( \frac{1}{Q_t(1-Q_t)} - \frac{S}{1-S} \frac{1}{Q_t} \frac{n'(Q_t)}{n(Q_t)} \right)^{-1} \right) dt \\
&= \left( \frac{C_t^{1-S}}{1-S}((1-R)V)^\rho e^{-\delta t} - V_x C_t \right) dt + \sigma Y_t V_y dB_t + (V_\phi - \gamma^\uparrow V_x Y_t) d\Phi_t^\uparrow + \left( \frac{V_x}{\gamma^\downarrow} Y_t - V_\phi \right) d\Phi_t^\downarrow \\
&\quad - \theta SV \left( \ell(Q_t) - \frac{\sigma^2}{2} \frac{1-S}{S} \left( \frac{1}{Q_t(1-Q_t)} - \frac{S}{1-S} \frac{1}{Q_t} \frac{n'(Q_t)}{n(Q_t)} \right)^{-1} \right) dt
\end{aligned} \tag{4.14}$$

where the function  $\ell$  is given by

$$\ell(q) = \frac{\alpha}{S} - \frac{1-S}{S} \left( \lambda\sigma - \frac{\sigma^2}{2} \right) q - (1-R) \frac{1-S}{S} \frac{\sigma^2}{2} q^2.$$

Maximising the first  $dt$ -term on the right hand side of (4.14) over  $C_t$  shows that the candidate optimal consumption stream  $C^* = (C_t^*)_{t \geq 0}$  satisfies

$$C_t^* := (X_t + \Phi_t \tilde{Y}_t) n(Q_t) = (X_t + \Phi_t Y_t \kappa(Q_t)) n(Q_t), \tag{4.15}$$

This implies in particular that  $n$  can be interpreted as the optimal consumption rate in terms of shadow wealth. Then we have

$$\left( \frac{(C_t^*)^{1-S}}{1-S} (e^{-\delta t} ((1-R)V)^\rho) - V_x C_t^* \right) = \frac{S}{1-S} (X_t + \Phi_t \tilde{Y}_t) n(Q_t) V_x = \theta SV n(Q_t)$$

Plugging this into (4.14) and using that the  $dt$ -term must vanish, after dividing by  $\theta SV$ , we obtain

$$0 = n(Q_t) - \ell(Q_t) + \frac{\sigma^2}{2} \frac{1-S}{S} \left( \frac{1}{Q_t(1-Q_t)} - \frac{S}{1-S} \frac{1}{Q_t} \frac{n'(Q_t)}{n(Q_t)} \right)^{-1}, \tag{4.16}$$

Rearranging (4.16) and noting that  $\ell(q) - m(q) = \frac{1-S}{S} \frac{\sigma^2}{2} q(1-q)$ , we obtain after a rearrangement that  $\kappa$  satisfies the ODE

$$\kappa'(q) = \frac{\kappa(q)}{q(1-q)} \frac{m(q) - n(q)}{\ell(q) - n(q)}. \tag{4.17}$$

Combining this with (4.10) we find that  $n$  satisfies the autonomous ODE

$$n'(q) = O(n, q) \quad \text{where} \quad O(n, q) = \frac{1-S}{S} \frac{n}{1-q} \frac{m(q) - n}{\ell(q) - n}. \tag{4.18}$$

Plugging (4.17) into (4.8) and noting the simple identify

$$\frac{m(q(p)) - \ell(q(p))}{m(p) - \ell(p)} = \frac{q(p)(1-q(p))}{p(1-p)}$$

we obtain after a rearrangement that  $q$  satisfies the ODE

$$q'(p) = \frac{\ell(q(p)) - n(q(p))}{\ell(p) - m(p)},$$

which may be re-expressed as

$$p'(q) = \frac{\ell(p(q)) - m(p(q))}{\ell(q) - n(q)}. \quad (4.19)$$

We find a candidate solution therefore by first solving the autonomous first equation (4.18) for  $n$ , and then solving (4.17) to determine  $\kappa$ . This then yields an expression for  $p$  in terms of  $q$  via (4.7) or (4.19). The next issue is to find the initial condition for the ODE (4.18) which will depend on the round-trip transaction cost  $\xi$ .

#### 4.4 No-trade region and boundary conditions

We assume that the no-trade region in fraction of shadow wealth is given by  $[q_*, q^*]$  for some  $q_* < q^* \in \mathbb{R}$ . This will correspond to a no-trade region in fractions of real wealth of the form  $[p_*, p^*]$  for some  $p_* < p^* \in \mathbb{R}$ .

Outside the no trade region, the shadow price coincides with the bid or ask price (depending on the side). Hence, we get the following boundary condition for the function  $\kappa$ .

$$\kappa(q) = \gamma^\uparrow, \quad q \leq q_*, \quad \text{and} \quad \kappa(q) = \frac{1}{\gamma^\downarrow}, \quad q \geq q^*. \quad (4.20)$$

Considering, (4.20) for  $q = q_*$  and  $q = q^*$ , integrating  $\kappa'(q)/\kappa(q)$  between  $q_*$  and  $q^*$ , and using the ODE (4.17) for  $\kappa$ , we obtain

$$\gamma^\uparrow \gamma^\downarrow = \frac{\kappa(q_*)}{\kappa(q^*)} = \exp \left( - \int_{q_*}^{q^*} \kappa'(q)/\kappa(q) \, dq \right) = \exp \left( \int_{q_*}^{q^*} \frac{1}{q(1-q)} \frac{n(q) - m(q)}{\ell(q) - n(q)} \, dq \right). \quad (4.21)$$

Moreover, assuming smooth fit of  $\kappa$  at  $q_*$  and  $q^*$ , (4.20) implies that  $\kappa'(q_*) = 0 = \kappa'(q^*)$ . This together with the ODE (4.17) for  $\kappa$ , this implies that

$$n(q_*) = m(q_*) \quad \text{and} \quad n(q^*) = m(q^*). \quad (4.22)$$

The idea is that the integral condition (4.21) together with the boundary condition (4.22) fixes  $q_*$  and  $q^*$  and then also the particular solution  $n = n_{\gamma^\uparrow \gamma^\downarrow}$  we want. The boundaries  $p_*$  and  $p^*$  can then be found by

$$p_* = \frac{q_*}{\gamma^\uparrow + (1 - \gamma^\uparrow)q_*} \quad \text{and} \quad p^* = \frac{\gamma^\downarrow q^*}{1 + (\gamma^\downarrow - 1)q^*}.$$

It remains to analyse solutions to  $n'(q) = O(n, q)$  started from points  $(z, n(z) = m(z))$ , and to look for the initial point  $z = q_*$  such that (4.21) holds.

**Remark 4.4.** It will follow from later analysis (see Proposition 5.3) that  $q_M \in [q_*, q^*]$  so that the Merton ratio lies inside the no-transaction region expressed in terms of fractions of shadow wealth. But this does imply that  $q_M \in [p_*, p^*]$ , and it is possible that the Merton line lies outside the no-transaction region. This can only happen when  $q_M \notin [0, 1]$ . For example, in the case  $q_M > 1$ , if  $\gamma^\downarrow > \frac{q_M}{q_M - 1}$  then the Merton line does not even lie in the solvency region  $\mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}$ . Further, if  $q_M > 1$  then there are cases when  $p^* = p^*(\gamma^\uparrow, \gamma^\downarrow)$  is a non-monotonic function of transaction costs. The fact that it is initially increasing will follow from the small transaction cost results and especially Corollary 7.3; the fact that it must subsequently decrease will follow from the fact that the no-transaction region must lie inside the solvency region. See also Figure 4(d) for a numerical example.

## 5 Problem with friction: rigorous statements

We assume henceforth that market price of risk  $\lambda := \frac{\mu-r}{\sigma}$  is non-zero. Otherwise the problem is trivial because the solution for the problem with frictions coincides with the one without friction. At  $t = 0$  the agent would re-balance their portfolio to leave a portfolio consisting of investments in the riskless asset only; thereafter they would continue to not invest in the risky asset.

**Standing Assumption 2.** *We assume  $\lambda \neq 0$ .*

Note that Standing Assumption 2 implies in particular that the Merton ratio  $q_M = \frac{\lambda}{\sigma R}$  is non-zero.

For better readability, all proofs and some results of a more technical nature can be found in Appendix A.

### 5.1 Well-posedness conditions

We begin our discussion by defining when the problem with friction (4.3) is well posed. It turns out that this does not depend on the transaction cost parameters  $\gamma^\uparrow$  and  $\gamma^\downarrow$  for buying and selling separately but only on their product  $\xi := \gamma^\uparrow \gamma^\downarrow$ .

**Definition 5.1.** Define the *threshold transaction cost* by<sup>4</sup>

$$\bar{\xi} := \begin{cases} 1 & \text{if } m \text{ has at most one zero,} \\ \exp\left(\int_{q_-^m}^{q_+^m} -\frac{1}{q(1-q)} \frac{m(q)}{\ell(q)} dq\right) & \text{if } m \text{ has two distinct zeros } q_-^m < q_+^m. \end{cases} \quad (5.1)$$

Then  $\xi \in (1, \infty)$  is said to be in the *well-posedness range of transaction costs* if one of the following two conditions is satisfied

- (a)  $R < 1$ ,  $m(0), m(1) > 0$  and  $\xi > \bar{\xi}$ .
- (b)  $R > 1$ ,  $m(q_M) > 0$  and  $\xi < \bar{\xi}$ .

To understand Definition 5.1, let us look at the cases  $R < 1$  and  $R > 1$  separately.

First, if  $R < 1$ , the problem with friction cannot be well posed if either investing nothing ( $q = 0$ ) or everything ( $q = 1$ ) in the risky asset leads to infinite wealth in the frictionless problem (which is the case if  $m(0) \leq 0$  or  $m(1) \leq 0$ , respectively) since these are the two cases where the frictionless strategy is also admissible with friction; the corresponding wealth is either the same if  $q = 0$  or reduced by the factor  $\frac{1}{\gamma^\downarrow}$  if  $q = 1$ . But even if  $m(0), m(1) > 0$ , the frictionless Merton problem will be ill-posed if  $m(q_M) \leq 0$ . Hence, the problem with friction is only well-posed if transaction costs are sufficiently large.

Next, if  $R > 1$ , the problem with friction cannot be well posed if the frictionless Merton problem is not well posed, i.e., if  $m(q_M) \leq 0$ . But even if  $m(q_M) > 0$ , the problem with friction is only well-posed if transaction costs are sufficiently small.

We shall show in Section 5.6 that if  $\xi := \gamma^\uparrow \gamma^\downarrow$  does not lie in the well-posedness range of transaction costs, the problem (4.3) is ill-posed.

**Remark 5.2.** If  $m$  in case (a) of Definition 5.1 has two distinct zeros  $q_-^m < q_+^m$ , then  $\bar{\xi} < \infty$  since either  $q_-^m < q_+^m < 0$  or  $0 < q_-^m < q_+^m < 1$  or  $1 < q_-^m < q_+^m$  by the assumption that  $m(0), m(1) > 0$ .

In case (b) of Definition 5.1, however, if  $m(q_M) > 0$ , then  $m$  always has two distinct zeros  $q_-^m < q_+^m$ . Moreover,  $\bar{\xi} = \infty$  if and only if  $m(0) \geq 0$  or  $m(1) \geq 0$  or there exists  $q \in (0, 1)$  with  $\ell(q) \geq 0$ . Using that  $m(0) = \ell(0)$  and  $m(1) = \ell(1)$  the latter three cases can be summarised by the single condition:  $\bar{\xi} = \infty$  if and only if  $\max_{q \in [0,1]} \ell(q) \geq 0$ .

---

<sup>4</sup>Here, we agree that  $\exp(\int_{q_-^m}^{q_+^m} -\frac{1}{q(1-q)} \frac{m(q)}{\ell(q)} dq) := \infty$  in case that  $\int_{q_-^m}^{q_+^m} |\frac{1}{q(1-q)} \frac{m(q)}{\ell(q)}| dq = \infty$ .

## 5.2 The free boundary problem

We first establish existence, uniqueness and further properties of the optimal consumption rate  $n$ . Mathematically, this is linked to a first-order free-boundary problem, which is additionally delicate due to the fact that the right-hand side of the ODE (5.2) is not defined for  $q = 1$ , which may lie inside the free-boundary region.

The region between the two free boundaries for  $n$  turns out to be the shadow no-trade region. Note that the latter does not depend on the transaction costs parameters  $\gamma^\uparrow, \gamma^\downarrow$  separately but only on their product  $\xi = \gamma^\uparrow \gamma^\downarrow$ .

**Proposition 5.3.** *Let  $\xi \in (1, \infty)$  be in the well-posedness range of transaction costs. Then there exists unique boundary points  $q_*(\xi), q^*(\xi) \in \mathbb{R}$  satisfying*

$$\begin{aligned} 0 < q_*(\xi) < q_M < q^*(\xi) < 1 & \text{ if } q_M \in (0, 1), \\ 0 < q_*(\xi) < q^*(\xi) = 1 & \text{ if } q_M = 1, \\ 0 < q_*(\xi) < q_M < q^*(\xi) < 2q_M - 1 & \text{ if } q_M > 1, \\ 2q_M < q_*(\xi) < q_M < q^*(\xi) < 0 & \text{ if } q_M < 0, \end{aligned}$$

as well as a unique continuously differentiable function  $n_\xi : [q_*(\xi), q^*(\xi)] \rightarrow (0, \infty)$  with the following properties:

- (a)  $n_\xi(q) \neq \ell(q)$  for  $q \neq 1$ .
- (b)  $n_\xi$  is decreasing if  $R < 1$  and increasing if  $R > 1$ .
- (c) On  $[q_*(\xi), q^*(\xi)] \setminus \{1\}$ ,  $n_\xi$  satisfies the ODE

$$n'_\xi(q) = \frac{1-S}{S} \frac{n_\xi}{1-q} \frac{m(q) - n_\xi(q)}{\ell(q) - n_\xi(q)}. \quad (5.2)$$

- (d)  $n_\xi$  satisfies the boundary conditions

$$\begin{aligned} n_\xi(q_*(\xi)) &= m(q_*(\xi)) \quad \text{and} \quad n_\xi(q^*(\xi)) = m(q^*(\xi)), \\ n'_\xi(q_*(\xi)) &= 0 \quad \text{and} \quad n'_\xi(q^*(\xi)) = 0. \end{aligned} \quad (5.3)$$

- (e) If  $1 \in [q_*(\xi), q^*(\xi)]$ ,  $n_\xi$  satisfies

$$n_\xi(1) = m(1) \quad \text{and} \quad n'_\xi(1) = m'(1).$$

- (f)  $n_\xi$  satisfies the integral constraint

$$\xi = \exp \left( \int_{q_*(\xi)}^{q^*(\xi)} -\frac{1}{q(1-q)} \frac{m(q) - n_\xi(q)}{\ell(q) - n_\xi(q)} dq \right). \quad (5.4)$$

Moreover,  $q_*(\xi)$  and  $q^*(\xi)$  satisfy the bounds

$$q_*(\xi) < \frac{\xi q^*(\xi)}{1 + (\xi - 1)q^*(\xi)} < \frac{\xi}{\xi - 1}, \quad (5.5)$$

$$q^*(\xi) > \frac{q_*(\xi)}{\xi - q_*(\xi)(\xi - 1)} > -\frac{1}{\xi - 1}. \quad (5.6)$$

We next turn to existence, uniqueness and further properties of the relative shadow price  $\kappa$ .

**Corollary 5.4.** Let  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$  with  $\xi := \gamma^\uparrow \gamma^\downarrow > 1$  be such that  $\xi$  lies in the well-posedness range of transaction costs. Then there exists a unique continuously differentiable function  $\kappa_{\gamma^\uparrow, \gamma^\downarrow} : [q_*(\xi), q^*(\xi)] \rightarrow (0, \infty)$  with the following properties:

- (a)  $\kappa_{\gamma^\uparrow, \gamma^\downarrow}$  is decreasing.
- (b) On  $[q_*(\xi), q^*(\xi)] \setminus \{1\}$ ,  $\kappa_{\gamma^\uparrow, \gamma^\downarrow}$  satisfies the ODE

$$\kappa'_{\gamma^\uparrow, \gamma^\downarrow}(q) = \frac{\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q)}{q(1-q)} \frac{m(q) - n_\xi(q)}{\ell(q) - n_\xi(q)}, \quad (5.7)$$

- (c)  $\kappa_{\gamma^\uparrow, \gamma^\downarrow}$  satisfies the boundary conditions

$$\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q_*(\xi)) = \gamma^\uparrow \quad \text{and} \quad \kappa_{\gamma^\uparrow, \gamma^\downarrow}(q^*(\xi)) = \frac{1}{\gamma^\downarrow}, \quad (5.8)$$

$$\kappa'_{\gamma^\uparrow, \gamma^\downarrow}(q_*(\xi)) = 0 \quad \text{and} \quad \kappa'_{\gamma^\uparrow, \gamma^\downarrow}(q^*(\xi)) = 0. \quad (5.9)$$

Moreover, if  $\tilde{\gamma}^\uparrow, \tilde{\gamma}^\downarrow \in [1, \infty)$  are such that  $\tilde{\gamma}^\uparrow \tilde{\gamma}^\downarrow = \xi$ , then

$$\kappa_{\tilde{\gamma}^\uparrow, \tilde{\gamma}^\downarrow} = \kappa_{\gamma^\uparrow, \gamma^\downarrow} \times \frac{\tilde{\gamma}^\uparrow}{\gamma^\uparrow} = \kappa_{\gamma^\uparrow, \gamma^\downarrow} \times \frac{\gamma^\downarrow}{\tilde{\gamma}^\downarrow}.$$

### 5.3 No-transaction region

We next turn to the no-transaction region in terms of real (as opposed to shadow) quantities. We shall see that the relationship between the real no-transaction region and the shadow no-transaction region is neatly described by real Möbius transformations.

**Definition 5.5.** Denote by  $\mathbb{R} \cup \{\overline{\infty}\}$  the compactified real line. For  $c \in (0, \infty)$ , define the real Möbius transformation  $\tau_c : \mathbb{R} \cup \{\overline{\infty}\} \rightarrow \mathbb{R} \cup \{\overline{\infty}\}$  by

$$\tau_c(q) := \begin{cases} \frac{cq}{1+(c-1)q} & \text{if } q \in \mathbb{R} \setminus \{\frac{1}{1-c}\}, \\ \overline{\infty} & \text{if } q = \frac{1}{1-c}, \\ \frac{c}{c-1} & \text{if } q = \overline{\infty}. \end{cases}$$

Note that the pair  $(\{\tau_c, c \in (0, \infty)\}, \circ)$  forms an Abelian group with the additional property

$$\tau_c \circ \tau_d = \tau_{cd}, \quad c, d \in (0, \infty).$$

**Remark 5.6.** We write  $\overline{\infty}$  to distinguish this ‘‘point at infinity’’ from the points  $-\infty$  and  $\infty$  of the extended real line. We also agree that any real number  $x$  satisfies both  $x > \overline{\infty}$  and  $x < \overline{\infty}$ , and therefore interpret  $(\overline{\infty}, x)$  as  $(-\infty, x)$  and  $(x, \overline{\infty})$  as  $(x, \infty)$ . We finally agree that  $c/0 := \overline{\infty}$  for  $c \in \mathbb{R} \setminus \{0\}$  so that  $\tau_1 = \text{id}$ .

With the help of Möbius transformations, we may define the boundary points of the real no-trade region (which depends on the transaction cost parameters  $\gamma^\uparrow$  and  $\gamma^\downarrow$  separately) in terms of the boundary points of the shadow no-trade region (which depends on the transaction cost parameters  $\gamma^\uparrow$  and  $\gamma^\downarrow$  only through their product  $\xi = \gamma^\uparrow \gamma^\downarrow$ ).

**Definition 5.7.** Let  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$  with  $\xi := \gamma^\uparrow \gamma^\downarrow > 1$  be such that  $\xi$  lies in the well-posedness range of transaction costs. Let  $q_*(\xi), q^*(\xi)$  be as in Proposition 5.3. Define  $p_*(\gamma^\uparrow, \gamma^\downarrow), p^*(\gamma^\uparrow, \gamma^\downarrow) \in \mathbb{R}$  by

$$p_*(\gamma^\uparrow, \gamma^\downarrow) := \tau_{\frac{1}{\gamma^\uparrow}}(q_*(\xi)) = \frac{q_*(\xi)}{\gamma^\uparrow + (1 - \gamma^\uparrow)q_*(\xi)}, \quad (5.10)$$

$$p^*(\gamma^\uparrow, \gamma^\downarrow) := \tau_{\gamma^\downarrow}(q^*(\xi)) = \frac{\gamma^\downarrow q^*(\xi)}{1 + (\gamma^\downarrow - 1)q^*(\xi)}. \quad (5.11)$$

Note that  $\tau_{\gamma^\uparrow}(p^*(\gamma^\uparrow, \gamma^\downarrow)) = \tau_{\gamma^\uparrow} \circ \tau_{\frac{1}{\gamma^\uparrow}}(q_*(\xi)) = \tau_1(q_*(\xi)) = q_*(\xi)$  and similarly  $\tau_{\frac{1}{\gamma^\downarrow}}(p^*(\gamma^\uparrow, \gamma^\downarrow)) = q^*(\xi)$ . The following result gives some estimates for the lower and upper bound of the no-transaction region.

**Lemma 5.8.** *Let  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$  with  $\xi := \gamma^\uparrow \gamma^\downarrow > 1$  be such that  $\xi$  lies in the well-posedness range of transaction costs. Then*

$$\begin{aligned} 0 < p_*(\gamma^\uparrow, \gamma^\downarrow) &< q_M < p^*(\gamma^\uparrow, \gamma^\downarrow) < 1, & \text{if } q_M \in (0, 1) \\ 0 < p_*(\gamma^\uparrow, \gamma^\downarrow) &< p^*(\gamma^\uparrow, \gamma^\downarrow) = 1, & \text{if } q_M = 1 \\ 0 < p_*(\gamma^\uparrow, \gamma^\downarrow) &< p^*(\gamma^\uparrow, \gamma^\downarrow) < \tau_{\gamma^\downarrow}(2q_M - 1) & \text{if } q_M > 1, \\ \tau_{\frac{1}{\gamma^\uparrow}}(2q_M) &< p_*(\gamma^\uparrow, \gamma^\downarrow) < p^*(\gamma^\uparrow, \gamma^\downarrow) < 0, & \text{if } q_M < 0. \end{aligned}$$

Moreover,  $1 \in [p_*(\gamma^\uparrow, \gamma^\downarrow), p^*(\gamma^\uparrow, \gamma^\downarrow)]$  if and only if  $1 \in [q_*(\xi), q^*(\xi)]$ .

**Remark 5.9.** (a) It follows from Lemma 5.8 (see also Davis and Norman [10, p704]) that if the Merton line lies in the first quadrant, then the no-transaction region also lies in the first quadrant (i.e., it is never optimal to have a short position in either the riskless asset or the risky asset) and contains the Merton line. A numerical illustration of this case will be given in Section 6.1. However, if  $q_M < 0$  or if  $q_M > 1$ , then it may be the case that  $q_M$  lies outside the interval  $[p_*(\gamma^\uparrow, \gamma^\downarrow), p^*(\gamma^\uparrow, \gamma^\downarrow)]$  and that the Merton line lies *outside* the no-transaction region. As discussed in Remark 4.4 this has to be the case if  $\gamma^\downarrow > \frac{q_M}{q_M - 1}$ . If  $q_M > 1$ , then Davis and Norman [10, p704] conjectured that the no-transaction region lies in the second quadrant, but this need not be the case (although it must always intersect the second quadrant, and indeed we must have  $p^*(\gamma^\uparrow, \gamma^\downarrow) > 1$ ). Conversely, Shreve and Soner [34, p675] conjectured that  $q_M > p^*(\gamma^\uparrow, \gamma^\downarrow)$ , whenever  $q_M > 1$ , but this is not true for small transaction costs.

(b) If  $q_M > 1$  (and  $m_M, m(1) > 0$ ), a new phenomenon arises. It follows from Proposition 5.3 that if  $\xi$  is large enough so that  $q_*(\xi) < 1$ , then  $n_\xi$  passes through the singular point  $(1, m(1))$  and on the interval  $(1, q^*(\xi))$ ,  $n_\xi$  does not depend on  $\xi$  (and hence also  $q^*(\xi)$  does not depend on  $\xi$ ). It follows that once  $\xi$  is sufficiently large,  $p^*(\gamma^\uparrow, \gamma^\downarrow)$  is independent of  $\gamma^\downarrow$ . In the additive case, Hobson et al. [22, Corollary 5.4] give a financial explanation behind this observation, which also applies to the case of stochastic differential utility. In particular, if  $1 \in (q_*(\xi), q^*(\xi))$ , then it is possible (indeed inevitable) that at some point the agent, who finances consumption from cash wealth, will exhaust holdings of the riskless asset. At this point they continue consuming, and finance consumption by borrowing. Subsequently they will sell units of the risky asset in order to stop their short position in the riskless asset from getting too large, but they will always have a short position in cash. Put simply the process  $Q$  never returns to the set  $(q_*(\xi), 1]$  once it has left it; see also Theorem 5.13(a) below. It follows that if the initial holdings involve a short cash position, then the trajectory of the process  $(X_t, \Phi_t)$  will be such that  $Q_t$  never hits the lower threshold  $q_*(\xi)$ , the agent will never purchase units of the risky asset, and the transaction costs on purchases is irrelevant.

We proceed to describe existence, uniqueness and further properties of the shadow fraction of wealth in terms of the real fraction of wealth.

**Proposition 5.10.** *Let  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$  with  $\xi := \gamma^\uparrow \gamma^\downarrow > 1$  be such that  $\xi$  lies in the well-posedness range of transaction costs. Then there exists a unique, increasing, continuously differentiable function  $q_{\gamma^\uparrow, \gamma^\downarrow} : [p_*(\gamma^\uparrow, \gamma^\downarrow), p^*(\gamma^\uparrow, \gamma^\downarrow)] \rightarrow [q_*(\xi), q^*(\xi)]$  with the following properties:*

(a) *on  $[p_*(\gamma^\uparrow, \gamma^\downarrow), p^*(\gamma^\uparrow, \gamma^\downarrow)] \setminus \{1\}$ ,  $q_{\gamma^\uparrow, \gamma^\downarrow}$  satisfies the ODE*

$$q'(p) = \frac{\ell(q(p)) - n_\xi(q(p))}{\ell(p) - m(p)}. \quad (5.12)$$

(b)  $q_{\gamma^\uparrow, \gamma^\downarrow}$  satisfies the boundary conditions

$$q_{\gamma^\uparrow, \gamma^\downarrow}(p_*(\gamma^\uparrow, \gamma^\downarrow)) = \tau_{\gamma^\uparrow}(p_*(\gamma^\uparrow, \gamma^\downarrow)) = q_*(\xi) \quad \text{and} \quad q_{\gamma^\uparrow, \gamma^\downarrow}(p^*(\gamma^\uparrow, \gamma^\downarrow)) = \tau_{\frac{1}{\gamma^\downarrow}}(p^*(\gamma^\uparrow, \gamma^\downarrow)) = q^*(\xi),$$

$$q'_{\gamma^\uparrow, \gamma^\downarrow}(p_*(\gamma^\uparrow, \gamma^\downarrow)) = \tau'_{\gamma^\uparrow}(p_*(\gamma^\uparrow, \gamma^\downarrow)) \quad \text{and} \quad q'_{\gamma^\uparrow, \gamma^\downarrow}(p^*(\gamma^\uparrow, \gamma^\downarrow)) = \tau'_{\frac{1}{\gamma^\downarrow}}(p^*(\gamma^\uparrow, \gamma^\downarrow)).$$

(c) If  $1 \in [p_*(\gamma^\uparrow, \gamma^\downarrow), p^*(\gamma^\uparrow, \gamma^\downarrow)]$ , then  $q_{\gamma^\uparrow, \gamma^\downarrow}(1) = 1$

Moreover, if  $\tilde{\gamma}^\downarrow, \tilde{\gamma}^\uparrow \in [1, \infty)$  are such that  $\tilde{\gamma}^\downarrow \tilde{\gamma}^\uparrow = \xi$ , then

$$q_{\tilde{\gamma}^\downarrow, \tilde{\gamma}^\uparrow} = q_{\gamma^\uparrow, \gamma^\downarrow} \circ \tau_{\frac{\tilde{\gamma}^\downarrow}{\gamma^\downarrow}} = q_{\gamma^\uparrow, \gamma^\downarrow} \circ \tau_{\frac{\tilde{\gamma}^\uparrow}{\gamma^\uparrow}}.$$

## 5.4 The candidate value function and candidate optimal strategy

In order to define the candidate value function on the whole of  $\mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ$  (and not just inside the no-trade region), we need to extend  $n$ ,  $\kappa$  and  $q$ . The idea is to extend  $n$  and  $\kappa$  so that they are constant outside the no-transaction region. This in turn determines the extension of  $q$ .

**Definition 5.11.** Let  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$  with  $\xi := \gamma^\uparrow \gamma^\downarrow > 1$  be such that  $\xi$  lies in the well-posedness range of transaction costs.

(a) Define the function  $\bar{n}_\xi : (-\infty, \infty) \rightarrow (0, \infty)$  by

$$\bar{n}_\xi(q) := \begin{cases} n_\xi(q_*(\xi)) = m(q_*(\xi)) & \text{if } q \in (-\infty, q_*(\xi)), \\ n_\xi(q) & \text{if } q \in [q_*(\xi), q^*(\xi)], \\ n_\xi(q^*(\xi)) = m(q^*(\xi)) & \text{if } q \in (q^*(\xi), \infty). \end{cases}$$

(b) Define the function  $\bar{\kappa}_{\gamma^\uparrow, \gamma^\downarrow} : (-\infty, \infty) \rightarrow [\frac{1}{\gamma^\downarrow}, \gamma^\uparrow]$  by

$$\bar{\kappa}_{\gamma^\uparrow, \gamma^\downarrow}(q) := \begin{cases} \kappa_{\gamma^\uparrow, \gamma^\downarrow}(q_*(\xi)) = \gamma^\uparrow & \text{if } q \in (-\infty, q_*(\xi)), \\ \kappa_{\gamma^\uparrow, \gamma^\downarrow}(q) & \text{if } q \in [q_*(\xi), q^*(\xi)], \\ \kappa_{\gamma^\uparrow, \gamma^\downarrow}(q^*(\xi)) = \frac{1}{\gamma^\downarrow} & \text{if } q \in (q^*(\xi), \infty). \end{cases}$$

(c) Define the function  $\bar{q}_{\gamma^\uparrow, \gamma^\downarrow} : (\tau_{\frac{1}{\gamma^\downarrow}}(\bar{\infty}), \tau_{\gamma^\downarrow}(\bar{\infty})) = (\frac{-1}{\gamma^\uparrow - 1}, \frac{\gamma^\downarrow}{\gamma^\uparrow - 1}) \rightarrow \mathbb{R}$  by

$$\bar{q}_{\gamma^\uparrow, \gamma^\downarrow}(p) := \begin{cases} \tau_{\gamma^\uparrow}(p) & \text{if } p \in (\tau_{\frac{1}{\gamma^\downarrow}}(\bar{\infty}), p_*(\gamma^\uparrow, \gamma^\downarrow)), \\ q_{\gamma^\uparrow, \gamma^\downarrow}(p) & \text{if } p \in [p_*(\gamma^\uparrow, \gamma^\downarrow), p^*(\gamma^\uparrow, \gamma^\downarrow)], \\ \tau_{\frac{1}{\gamma^\downarrow}}(p) & \text{if } p \in (p^*(\gamma^\uparrow, \gamma^\downarrow), \tau_{\gamma^\downarrow}(\bar{\infty})), \end{cases}$$

We can now define candidate value function on the whole of  $\mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ$ .

**Definition 5.12.** Let  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$  with  $\xi := \gamma^\uparrow \gamma^\downarrow > 1$  be such that  $\xi$  lies in the well-posedness range of transaction costs. Define the value function  $\hat{V}^{\gamma^\uparrow, \gamma^\downarrow} : [0, \infty) \times \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ \rightarrow (1 - R)(0, \infty)$  by

$$\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi) := e^{-\delta\theta t} \frac{\left( x + \phi y \bar{\kappa}_{\gamma^\uparrow, \gamma^\downarrow} \left( \bar{q}_{\gamma^\uparrow, \gamma^\downarrow} \left( \frac{\phi y}{x + \phi y} \right) \right) \right)^{1-R}}{1 - R} \bar{n}_\xi \left( \bar{q}_{\gamma^\uparrow, \gamma^\downarrow} \left( \frac{\phi y}{x + \phi y} \right) \right)^{-\theta S}. \quad (5.13)$$

We proceed to turn to the candidate optimal strategy. Since we consider all parameter combinations (in particular,  $1 \in [p_*(\gamma^\uparrow, \gamma^\downarrow), p^*(\gamma^\uparrow, \gamma^\downarrow)]$  is allowed), this is quite delicate. Even in the additive case, the following result is not covered by standard results in the extant literature (see e.g. [34, 28]). For this reason, we provide a self-contained proof in Appendix A.3. Note that our proof only uses results on reflected SDEs in one dimension and is therefore technically substantially easier than the results in the extant literature (including [34, 28]), which rely on reflection result in a two dimensional domain.

**Theorem 5.13.** *Let  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$  with  $\xi := \gamma^\uparrow \gamma^\downarrow > 1$  be such that  $\xi$  is in the well-posedness range of transaction costs. Let  $(x, y, \phi) \in \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ$ .*

(a) *There exist unique processes  $(\hat{Q}, G = G^\uparrow - G^\downarrow)$  with continuous paths and with initial values  $\hat{Q}_0 = \min(\max(q(\frac{\phi y}{x+\phi y}), q_*(\xi)), q^*(\xi))$ ,  $G_0^\uparrow = 0 = G_0^\downarrow$ , such that the process  $\hat{Q} = (\hat{Q}_t)_{t \geq 0}$  defined by*

$$d\hat{Q}_t = \frac{2S}{\sigma(1-S)}(\ell(\hat{Q}_t) - n_\xi(\hat{Q}_t))dB_t + \hat{Q}_t \left( n_\xi(\hat{Q}_t) - 2\theta S\{\ell(\hat{Q}_t) - n_\xi(\hat{Q}_t)\} \right) dt + dG_t^\uparrow - dG_t^\downarrow \quad (5.14)$$

*takes values in  $[q_*(\xi), q^*(\xi)]$ , and  $G^\uparrow, G^\downarrow$  are nondecreasing and increase only when  $\hat{Q}_t = q_*(\xi)$  and  $\hat{Q}_t = q^*(\xi)$  respectively. Moreover, if  $1 \in [q_*(\xi), q^*(\xi)]$ , let  $\tau := \inf\{t \geq 0 : \hat{Q}_t \geq 1\}$ . Then  $\hat{Q} \in [q^*(\xi), 1)$  on  $[0, \tau]$ , and  $\hat{Q} = 1$  on  $[\tau, \infty]$  if  $q^*(\xi) = 1$  and  $\hat{Q} \in (1, q_*(\xi)]$  on  $[\tau, \infty]$  if  $q^*(\xi) > 1$ .*

(b) *Set*

$$\hat{\Phi}_0 := \begin{cases} q_*(\xi) \left( \phi + \frac{x}{y\gamma^\uparrow} \right) & \text{if } \phi y \gamma^\uparrow (1 - q_*(\xi)) < q_*(\xi)x, \\ \phi & \text{if } \phi y \gamma^\uparrow (1 - q_*(\xi)) \geq q_*(\xi)x \text{ and } \frac{\phi y}{\gamma^\downarrow} (1 - q^*(\xi)) \leq q^*(\xi)x, \\ q^*(\xi) \left( \phi + \frac{x\gamma^\downarrow}{y} \right) & \text{if } \frac{\phi y}{\gamma^\downarrow} (1 - q^*(\xi)) > q^*(\xi)x, \end{cases}$$

*and the processes  $\hat{\Phi} = (\hat{\Phi}_t)_{t \geq 0}$ ,  $\hat{C} = (\hat{C}_t)_{t \geq 0}$  and  $\hat{X} = (\hat{X}_t)_{t \geq 0}$  by*

$$\begin{aligned} \hat{\Phi}_t &:= \hat{\Phi}_0 \exp \left( \frac{G_t^\uparrow}{q_*} - \frac{G_t^\downarrow}{q^*} \right), \\ \hat{C}_t &:= \frac{Y_t \kappa(\hat{Q}_t) \hat{\Phi}_t}{\hat{Q}_t} n_\xi(\hat{Q}_t), \\ \hat{X}_t &:= \frac{1 - \hat{Q}_t}{\hat{Q}_t} Y_t \kappa(\hat{Q}_t) \hat{\Phi}_t. \end{aligned} \quad (5.15)$$

*Then*

- (i)  $(\hat{\Phi}, \hat{C}) \in \mathcal{A}_{\gamma^\uparrow, \gamma^\downarrow}(x, y, \phi)$ ,
- (ii)  $\hat{X} = X^{x, y, \phi, \hat{\Phi}, \hat{C}}$  and
- (iii)  $\hat{Q}$  satisfies

$$\hat{Q}_t = \frac{Y_t \kappa_{\gamma^\uparrow, \gamma^\downarrow}(\hat{Q}_t) \hat{\Phi}_t}{\hat{X}_t + Y_t \kappa_{\gamma^\uparrow, \gamma^\downarrow}(\hat{Q}_t) \hat{\Phi}_t}. \quad (5.16)$$

**Remark 5.14.** (a) The case  $\phi y \gamma^\uparrow (1 - q_*(\xi)) < q_*(\xi)x$  corresponds to starting in the region where the trading process  $\Phi$  involves an initial purchase of the risky asset, and the case  $\frac{\phi y}{\gamma^\downarrow} (1 - q^*(\xi)) > q^*(\xi)x$  corresponds to the case where the trading process involves an initial sale of the risky asset, and in each case the expression for  $\hat{X}_0$  given by (5.15) evaluated at  $t = 0$  reflects the impact of the trade. Implicit in the statement is the fact that we cannot have both  $\phi y \gamma^\uparrow (1 - q_*(\xi)) < q_*(\xi)x$  and  $\frac{\phi y}{\gamma^\downarrow} (1 - q^*(\xi)) > q^*(\xi)x$ .

(b) We can have  $1 \in (q_*(\xi), q^*(\xi))$ . Then the diffusion coefficient in (5.14) vanishes and the drift coefficient is positive for  $\hat{Q} = 1$ . In particular, the diffusion  $\hat{Q}$  is not regular for the domain  $(q_*(\xi), q^*(\xi))$ .

If  $\hat{Q}_s \geq 1$  then  $\hat{Q}_t > 1$  for  $t > s$ . This is equivalent to the fact that if  $\hat{X}_s \leq 0$  then for  $t > s$ , we have  $\hat{X}_t < 0$ , and the agent always takes a short position in the risky asset, see also Remark 5.9(b).

(c) We could instead start with dynamics for the candidate for  $P = \frac{Y\Phi}{X+Y\Phi}$ . These are given by:

$$\frac{d\hat{P}_t}{\hat{P}_t(1-\hat{P}_t)} = \left( (\mu - r) - \sigma^2 \hat{P}_t + \frac{n_\xi(q_{\gamma^\uparrow, \gamma^\downarrow}(\hat{P}_t))}{1 - q_{\gamma^\uparrow, \gamma^\downarrow}(\hat{P}_t)} \right) dt + \sigma dB_t + \frac{dG_t}{q_{\gamma^\uparrow, \gamma^\downarrow}(\hat{P}_t)(1 - q_{\gamma^\uparrow, \gamma^\downarrow}(\hat{P}_t))}.$$

However, these dynamics are more complicated than those of  $\hat{Q}$ , not least because they involve expressions like  $n_\xi(q_{\gamma^\uparrow, \gamma^\downarrow}(p))$ . Moreover, for  $\hat{P} = 1$ , the right hand side of the SDE is not well-defined.

We now link the candidate value function with the candidate optimal strategy.

**Theorem 5.15.** *Let  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$  with  $\xi := \gamma^\uparrow \gamma^\downarrow > 1$  be such that  $\xi$  is in the well-posedness range of transaction costs. Let  $(x, y, \phi) \in \mathcal{S}_{(\gamma^\uparrow, \gamma^\downarrow)}^\circ$  and  $(\hat{\Phi}, \hat{C}) \in \mathcal{A}_{(\gamma^\uparrow, \gamma^\downarrow)}(x, y, \phi)$  be the corresponding candidate optimal strategy from Theorem 5.13. Then  $\hat{C}$  is uniquely evaluable if  $\theta \leq 1$  and uniquely proper with right-continuous paths if  $\theta > 1$ . Moreover,*

$$\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, X_t^{x, y, \phi, \hat{\Phi}, \hat{C}}, Y_t, \hat{C}_t) = V_t^{\hat{C}} \quad \mathbb{P}\text{-a.s.}, \quad t \geq 0,$$

where  $V^{\hat{C}}$  denotes the unique utility process if  $\theta \leq 1$  and the unique proper utility process if  $\theta > 1$ .

## 5.5 Verification argument

In order to complete the verification, we would like to use something like the martingale optimality principle, which in the additive case ( $R = S$ ,  $\rho = 0$ ) says that  $M$  given by  $M_t = \int_0^t e^{-\delta s} \frac{C_s^{1-S}}{1-S} ds + \hat{V}(t, X_t, Y_t, \Phi_t)$  is always a supermartingale and a martingale for the optimal strategy. In the general case with  $R \neq S$ , classical martingale optimality arguments break down due to the fact that the value process enters the integrand. Instead we need a comparison theorem for solutions of BSDEs.

To this end, we recall the notions of a subsolution and a supersolution of a BSDE. Definitions of sub- and supersolutions of BSDEs in the literature can vary slightly depending on the context. The definition we use here is the one used in Herdegen et al. [18, Definition 5.3], specialised to the case of Epstein-Zin SDU.

**Definition 5.16.** Let  $C \in \mathcal{P}_+$  be a consumption stream. A  $(1-R)\mathbb{R}_+$ -valued, càd, optional process  $V$  is called

- a *subsolution* for  $C$  if  $\limsup_{t \rightarrow \infty} \mathbb{E}[V_{t+}] \leq 0$  and for all bounded stopping times  $\tau_1 \leq \tau_2$ ,

$$V_{\tau_1} \leq \mathbb{E} \left[ V_{\tau_2+} + \int_{\tau_1}^{\tau_2} e^{-\delta s} \frac{C_s^{1-S}}{1-S} ((1-R)V_s)^\rho ds \middle| \mathcal{F}_{\tau_1} \right].$$

- a *supersolution* for  $C$  if  $\liminf_{t \rightarrow \infty} \mathbb{E}[V_{t+}] \geq 0$  and for all bounded stopping times  $\tau_1 \leq \tau_2$ ,

$$V_{\tau_1} \geq \mathbb{E} \left[ V_{\tau_2+} + \int_{\tau_1}^{\tau_2} e^{-\delta s} \frac{C_s^{1-S}}{1-S} ((1-R)V_s)^\rho ds \middle| \mathcal{F}_{\tau_1} \right].$$

- a *solution* for  $C$  if it is both a subsolution and a supersolution and  $\mathbb{E}[\int_0^\infty C_s^{1-S} ((1-R)V_s)^\rho ds] < \infty$ .

The next result essentially shows that for any  $(\hat{\Phi}, \hat{C}) \in \mathcal{A}_{\gamma^\uparrow, \gamma^\downarrow}(x, y, \phi)$  the process  $(\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, X_t, Y_t, \Phi_t))_{t \geq 0}$  is a supersolution for  $C$ . However, for technical reasons, this is more delicate and a stochastic perturbation argument (as in the classical Merton case, see [17]) is needed.

**Proposition 5.17.** *Let  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$  and suppose that  $\xi := \gamma^\uparrow \gamma^\downarrow \in (1, \infty)$  lies in the well-posedness range of transaction costs. Let  $(x, y, \phi) \in \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ$  and  $(\hat{\Phi}, \hat{C}) \in \mathcal{A}_{\gamma^\uparrow, \gamma^\downarrow}(x, y, \phi)$  be the corresponding candidate optimal strategy from Theorem 5.13. Let  $(\Phi, C) \in \mathcal{A}_{\gamma^\uparrow, \gamma^\downarrow}(x, y, \phi)$  and set  $X := X^{x, y, \phi, \Phi, C}$  and  $\hat{X} := X^{x, y, \phi, \hat{\Phi}, \hat{C}}$ . Then for every  $\varepsilon > 0$ ,  $(\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, X_t + \varepsilon \hat{X}_t, Y_t, \Phi_t + \varepsilon \hat{\Phi}_t))_{t \geq 0}$  is a supersolution for  $C + \varepsilon \hat{C}$ .*

We now can formulate a rigorous verification theorem.

**Theorem 5.18.** *Let  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$  and suppose that  $\xi := \gamma^\uparrow \gamma^\downarrow \in (1, \infty)$  lies in the well-posedness range of transaction costs. Let  $(x, y, \phi) \in \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ$  and  $\hat{C} \in \mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}^*(x, y, \phi)$  be the corresponding candidate optimal strategy. Then*

$$\sup_{C \in \mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}^*(x, y, \phi)} V_0^C = V_0^{\hat{C}} = \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(0, x, y, \phi),$$

where  $\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}$  is the candidate value function from (5.13).

## 5.6 The ill-posed case

We end this section by showing that if  $\gamma^\uparrow \gamma^\downarrow \in (1, \infty)$  does not lie in the well-posedness range of transaction costs, the problem (4.3) is ill posed, in the sense that either there exists a sequence of admissible strategies generating infinite utility or every strategy leads to utility of  $-\infty$ .

**Proposition 5.19.** *Fix  $(x, y, \phi) \in \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ$  and let  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$  be such that  $\xi := \gamma^\uparrow \gamma^\downarrow > 1$ . Suppose that  $\xi$  does not lie in the well-posedness range of transaction costs given in Definition 5.1.*

- (a) *If  $S \leq R < 1$  or  $R < S < 1$ , then there exists a sequence of consumption streams  $(C^n)_{n \in \mathbb{N}} \in \mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}^*(x, \phi, y)$  such that  $\lim_{n \rightarrow \infty} V_0^{C^n} = +\infty$ .*
- (b) *If  $S \geq R > 1$  or  $R > S > 1$ , then  $V_0^C = -\infty$  for all  $C \in \mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}^*(x, \phi, y)$ .*

## 6 Diagrammatic representations and numerical illustrations

In this section, we provide two numerical examples to illustrate some important qualitative properties of the solution, as represented by the function  $n_\xi$  and the limits of the no-transaction region  $\{q_*(\xi), q^*(\xi)\}$  (in the shadow-quantity of wealth coordinates) and  $\{p_*(\gamma^\uparrow, \gamma^\downarrow), p^*(\gamma^\uparrow, \gamma^\downarrow)\}$  in the original coordinates. The examples are not meant to be exhaustive, and indeed several further canonical cases can arise depending on the signs of  $\{1 - R, q_M, 1 - q_M\}$ , and the locations of the quadratics  $m(q)$  and  $\ell(q)$ . Some of the cases which are omitted here are treated more comprehensively in Hobson et al. [22, 21] in the special case of CRRA utility.

### 6.1 First Numerical Example: $R, S < 1$ and $q_M \in (0, 1)$

The model parameters are chosen such that  $R, S < 1$ ,  $m(0) > 0$ ,  $m(1) > 0$ ,  $\min_{q \in (0, 1)} m(q) < 0$  and  $q_M \in (0, 1)$ .<sup>5</sup> See the caption of Figure 1 for the precise values of each parameter used in the numerics. In this case, the problem is only well-posed for sufficiently large transaction costs (recall Definition 5.1). The solutions to the family of ODEs  $(n_{(z)}(q))_{q \geq z}$  introduced in part (a) of Proposition A.1 are only defined for  $z \in (0, q_-^m)$ , where  $q_-^m$  is the smaller root of  $m(q) = 0$ . Each solution  $(n_{(z)}(q))_{q \geq z}$  starts at  $(z, m(z))$  and crosses  $m$  again at some  $(\zeta^+(z), m(\zeta^+(z)))$ . Different choices of the left-starting point  $z$  correspond to different levels of total transaction costs  $\xi = \gamma^\uparrow \gamma^\downarrow$ . The correct choice of  $z$  can be identified from

<sup>5</sup>Economically this means that the investment-consumption problem is well-posed when only investment in the riskless asset is allowed, and also when only investment in the risky-asset is allowed, but the problem is ill-posed when transaction costs are zero and investment in both assets is permitted. Further, in the frictionless case, the optimal investment strategy involves a long position in both the risky asset and in cash.

the equation  $\Sigma^+(z) = \gamma^\uparrow \gamma^\downarrow$ , where  $\Sigma^+(\cdot)$  is defined in Proposition A.1. Then the purchase and sale boundaries (in shadow portfolio weight) are given by  $q_* = z$  and  $q^* = \zeta^+(z)$ . For example, Figure 1a shows two possible solutions to the ODE for  $n$  with different left-starting points and in turn different points of crossing with  $m$ . (Note, such solutions cannot cross, as they solve a first-order ODE with regular co-efficients, at least away from 1.) They correspond to the total transaction cost levels of  $\xi = 1.0609$  and  $\xi = 1.08742$  respectively. Note that if the transaction costs are such that  $\xi \leq 1.01903$  then the problem is ill-posed.

Consider three combinations of transaction costs  $(\gamma^\uparrow, \gamma^\downarrow)$  given by  $(1.03, 1.03)$ ,  $(1.05575, 1.03)$  and  $(1.05575, 1.00488)$  respectively. They are constructed such that Set 1 and Set 3 give the same value of  $\xi = 1.0609$  whereas Set 2 gives  $\xi = 1.08742$ . Fundamental quantities of the optimal solution under each set of transaction costs are shown in Figure 1. The dotted lines in each plot indicate the extension of the fundamental quantities beyond the no-transaction region (see Definition 5.11). The cross, circle and square markers locate the boundaries of the no-transaction region under Set 1, 2 and 3 of transaction costs respectively.

In Figure 1a, the transaction costs parameters of Set 1 and Set 3 yield identical optimal consumption rates per unit shadow wealth since  $n(\cdot)$  and  $q^*$  only depend on the total transaction costs  $\xi$ . Set 2 carries a higher total transaction cost and then it follows that the optimal consumption rate (in shadow wealth terms) is higher (and the no-transaction interval  $(q_*, q^*)$  is wider). In contrast, Figure 1b and 1d show that the three sets of transaction costs result in different ratios of shadow-to-real price as well as mappings between real and shadow portfolio weights. This is because they are affected by  $\gamma^\uparrow$  and  $\gamma^\downarrow$  individually, not just via the total transaction costs  $\xi = \gamma^\uparrow \gamma^\downarrow$ . These two quantities are monotonic with respect to the shadow portfolio weight.

Figure 2 shows how the no-transaction region (in both real and shadow portfolio weights) changes with the transaction costs. In this case,  $p_*$  and  $q_*$  (resp.  $p^*$  and  $q^*$ ) are decreasing (resp. increasing) in both  $\gamma^\uparrow$  and  $\gamma^\downarrow$ . But notably, Set 1 and 3 of the transaction costs will result in the same levels of  $q_*$  and  $q^*$ , as indicated by the same levels of the cross and square markers in Figure 2a and 2b. In terms of real portfolio weights, Set 1 has a larger  $p_*$  and  $p^*$  than Set 3. This is due to the fact that  $p_*$  and  $p^*$  depend on  $\gamma^\uparrow$  and  $\gamma^\downarrow$  individually. Note as well that in this example (with Merton ratio below one, i.e.  $q_M \in (0, 1)$ ) the no-transaction region  $[p_*, p^*]$  contains the Merton ratio for all levels of transaction costs, and is always contained in the first quadrant (i.e.  $0 < p_* < q_M < p^* < 1$ ), recall Remark 5.9.

Several qualitative features are also apparent from Figures 1 and 2 which may be traced back to the fact that  $n$  is decreasing and solutions  $n$  with different starting points get closer as  $q$  increases. In particular we observe that  $q_*$  varies more than  $q^*$  as  $\xi$  changes, and whilst  $\gamma^\uparrow$  affects both  $p_*$  and  $p^*$ ,  $\gamma^\downarrow$  mainly affects  $p_*$ .

## 6.2 Second Numerical Example: $R, S > 1$ and $q_M > 1$

In this example, we consider a combination of model parameters<sup>6</sup> such that  $R, S > 1$ ,  $m(q) > 0$  for  $q \in [0, 1]$  and  $q_M > 1$ . Specifically, the Merton ratio is (slightly) larger than one and the problem is well-posed for any level of transaction costs. An interesting feature in the case of the Merton ratio being above one is that all solutions to the family of ODEs in part (c) of Proposition A.2 with left-starting points below one will pass through the singular point  $(1, m(1))$ . Consequently, for any  $z < 1$ , we must have  $n_{[z]}(q) = n_{[1]}(q)$  on  $q \in (1, \zeta^+(1))$  and they all cross  $m$  at the same coordinate  $(\zeta^+(1), m(\zeta^+(1)))$ . In the shadow fraction of wealth coordinates, the sale boundary  $q^*(\xi)$  is constant as a function of  $\xi$  for sufficiently high transaction costs. See Figure 3a for an example where the two plots of  $n$  correspond to  $\xi = 1.0609$  and  $\xi = 1.08742$ . On  $q < 1$ , the solution  $n$  corresponding to smaller  $\xi$  dominates the solution corresponding to higher  $\xi$ , although the magnitude of the difference is very small numerically. The two

<sup>6</sup>The parameters are such that the problem is well-posed for any value of transaction costs, and both when only investments in the riskless asset are allowed and when only investments in the risky asset are allowed. Further, in the frictionless case, the optimal investment strategy involves borrowing to finance a leveraged position in the risky asset.

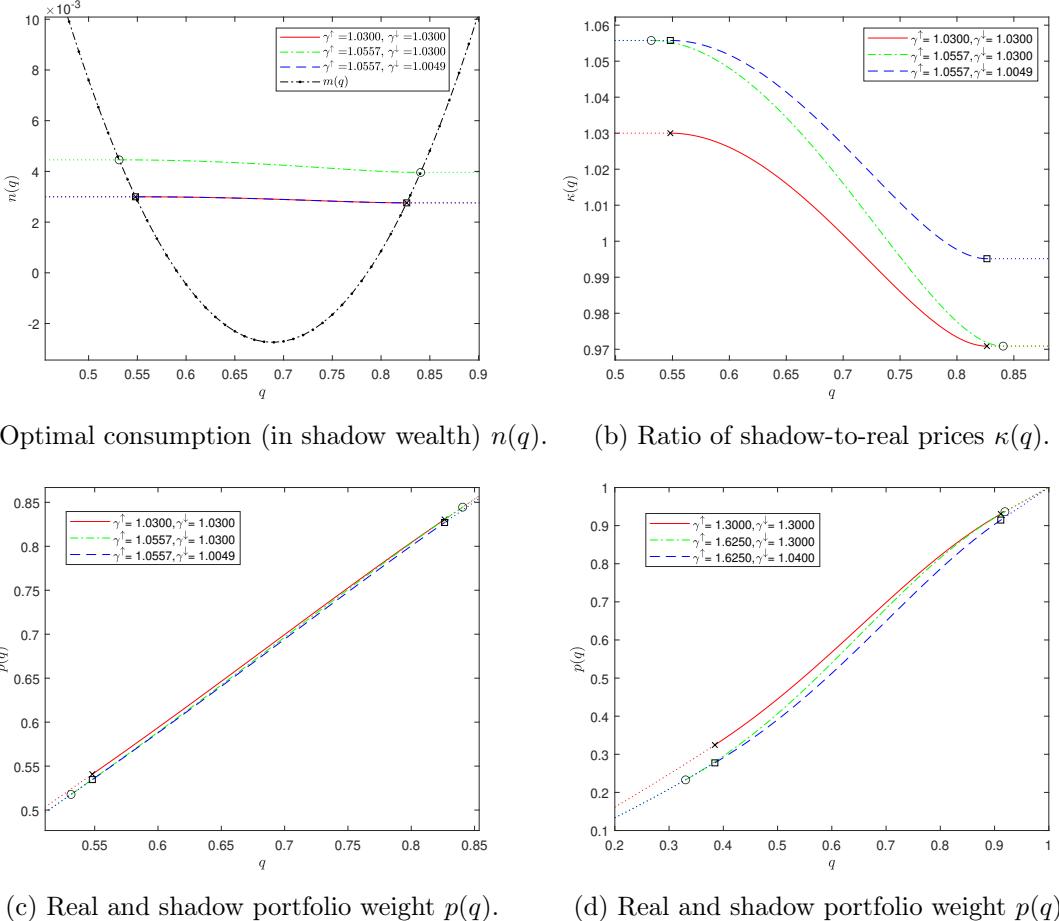


Figure 1: The fundamental quantities as a function of shadow portfolio weights in Example 1. Parameters used:  $R = 2/3$ ,  $S = 1/3$ ,  $\delta = 0.045$ ,  $r = 0$ ,  $\mu = 0.2$ ,  $\sigma = 0.66$  (such that  $\lambda = 0.3030$  and  $\alpha = 0.045$ ). The cross, circle and square markers indicate the boundary points of the no-transaction region (in  $p^*$  or  $q^*$ ) under the set of transaction costs  $(\gamma^\uparrow, \gamma^\downarrow)$  given by  $(1.03, 1.03)$ ,  $(1.05575, 1.03)$  and  $(1.05575, 1.00488)$  respectively. Panel (d) is identical to Panel (c) except the three sets of transaction costs  $(\gamma^\uparrow, \gamma^\downarrow)$  are now chosen to be  $(1.3, 1.3)$ ,  $(1.625, 1.3)$  and  $(1.625, 1.04)$ . The purpose of Panel (d) is to illustrate features of the relationships between the portfolio weights as functions of the transaction costs which may be difficult to ascertain in Panel (c) given that the curves lie almost on top of each other.

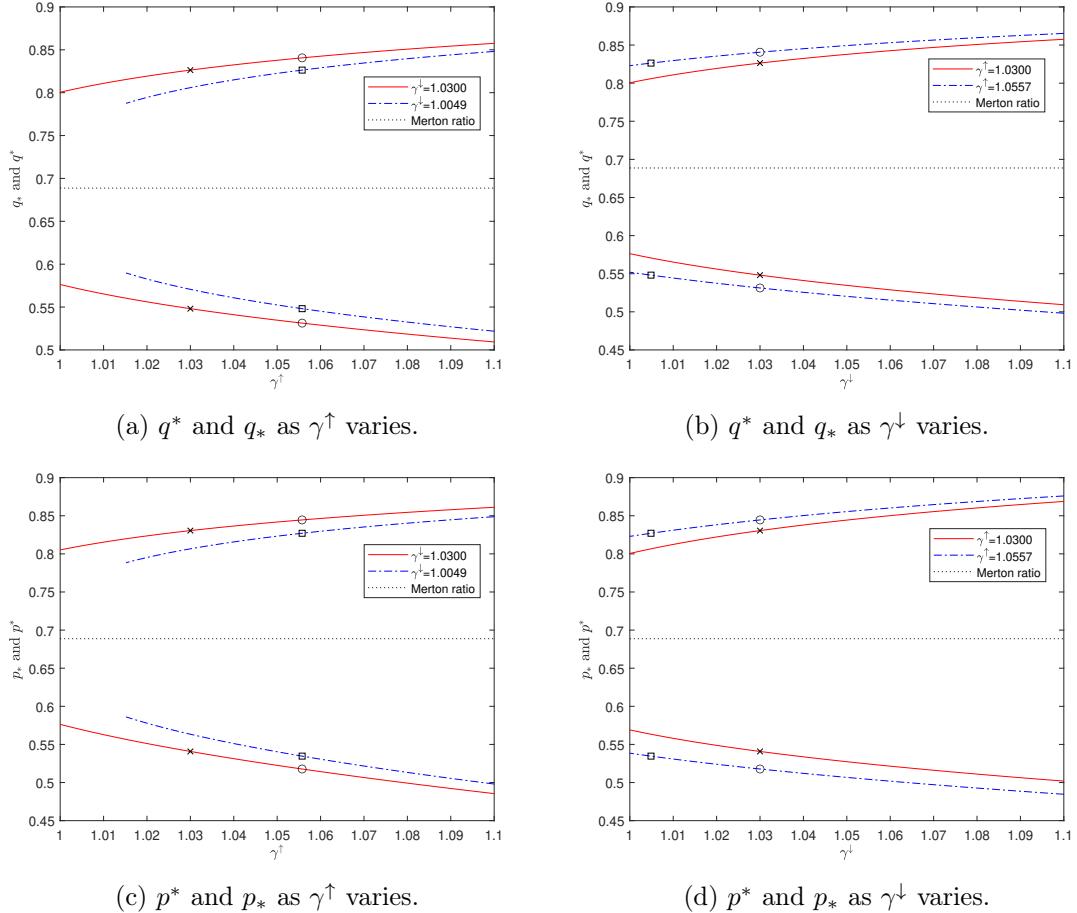


Figure 2: Behaviours of the no-transaction region as a function of transaction costs in Example 1. Parameters used:  $R = 2/3$ ,  $S = 1/3$ ,  $\delta = 0.045$ ,  $r = 0$ ,  $\mu = 0.2$ ,  $\sigma = 0.66$  (such that  $\lambda = 0.3030$  and  $\alpha = 0.045$ ). The cross, circle and square markers indicate the boundary points of the no-transaction region (in  $p^*$  or  $q^*$ ) under the set of transaction costs  $(\gamma^+, \gamma^-)$  given by  $(1.03, 1.03)$ ,  $(1.05575, 1.03)$  and  $(1.0557, 1.00488)$  respectively. We only display the results when the problem is well-posed where the transaction costs satisfy  $\gamma^+ \gamma^- = \xi > 1.01903$ .

functions pass through the same singular point  $(1, m(1))$  and coincide on  $q > 1$  until they both cross  $m(q)$  again on  $q > 1$  at the point indicated by the markers (which  $q$ -coordinate is just slightly above the Merton level in this example).

We consider the same three sets of transaction costs as in Example 1. Set 1 and Set 3 again result in identical optimal consumption rates (as functions of shadow wealth) as represented by  $n$  because they share the same level of  $\xi = \gamma^\uparrow \gamma^\downarrow$ . Furthermore, the shadow consumption rates are the same across all the three sets on  $q > 1$ . In Figure 3b,  $\kappa(q)$  is identical on  $q > 1$  under Sets 1 and 2 of transaction costs (whilst the difference is numerically very small on  $q < 1$ ). This is not surprising since  $\kappa(q)$  over  $q > 1$  only depends on  $(\gamma^\uparrow, \gamma^\downarrow)$  via  $q^*(\xi)$ ,  $(n_\xi(\cdot))_{q \leq q^*(\xi)}$  and  $\gamma^\downarrow$ ,<sup>7</sup> and Set 1 and 2 share the same value of  $\gamma^\downarrow$ . Further, when we move to the transaction cost parameters in Set 3 we find that on  $[1, q^*]$  the corresponding  $\kappa$  is a multiplicative scaling of the  $\kappa$  from Sets 1 and 2 (and this multiplicative scaling extends to the interval  $[q_*(\xi), q^*(\xi)]$  when we compare the functions  $\kappa$  that arise in Sets 1 and 3). These properties are inherited by the map between the real and shadow portfolio weight  $p = p(q)$ , where comparing the results from Sets 1 and 2, the difference in cost of purchase has no impact on the function  $p(q)$  on  $q > 1$ , and only a numerically small difference below 1), see Figure 3d.

The fact that the Merton ratio  $q_M$  is above 1 brings new phenomena, especially concerning the location of the no-transaction region and how it varies with transaction costs. Figure 4a and 4b confirm that the sale boundary  $q^*$  in shadow fractions of wealth units is totally insensitive to transaction costs (under our baseline setup with sufficiently large  $\xi$ ), while  $q_*$  is decreasing in transaction costs. Moreover, since  $p^*$  only depends on transaction costs via  $q^*(\xi)$  and  $\gamma^\downarrow$ ,  $p^*$  is totally insensitive to  $\gamma^\uparrow$  as shown in Figure 4c (recall Remark 5.9. Note also that  $p^*$  is decreasing in  $\gamma^\downarrow$  and this monotonicity is opposite to that in Example 1. Further, Figure 4c and 4d show that for some parameter values  $q_M \notin [p_*, p^*]$ , i.e. the Merton line is not necessarily contained in the no-transaction region, again recall Remark 5.9. For example, Figure 4d shows that  $p^*$  is above (resp. below) the Merton line when  $\gamma^\downarrow$  is small (resp. large).

## 7 Small transaction costs

In this section we discuss the small transaction cost regime, and how the boundaries of the no-transaction region depend on the transaction costs  $\gamma^\uparrow$  and  $\gamma^\downarrow$  when the transaction costs are small. In particular we look for an expansion in terms of  $\varepsilon^\uparrow = \gamma^\uparrow - 1$  and  $\varepsilon^\downarrow = 1 - \frac{1}{\gamma^\downarrow}$ . Let  $\varepsilon = \gamma^\uparrow \gamma^\downarrow - 1 = \frac{\varepsilon^\uparrow \varepsilon^\downarrow}{1 - \varepsilon^\downarrow} = \varepsilon^\uparrow + \varepsilon^\downarrow + \mathcal{O}(\varepsilon^\downarrow(\varepsilon^\uparrow + \varepsilon^\downarrow))$ .

Historically, the Merton investment-consumption problem with transaction costs has proved to be very challenging (even for the case of additive utility), and a natural response is to look for approximate solutions in the case of small transaction costs.<sup>8</sup> For additive utility, Rogers [33] argued that it is natural to expect the width of the no-transaction region to be of the order of the size of the transaction cost to the power one-third. This result was formalised by Janeček and Shreve [23]. They showed (under an assumption that the transaction costs on buying and selling are identical, or in our setting that  $\gamma^\downarrow - 1 = 1 - \frac{1}{\gamma^\uparrow}$ ), and under a (almost necessary) assumption that the problem without transaction costs problem is well-posed) that  $p_* = q_M - \Delta^{1/3} \varepsilon^{1/3} + \mathcal{O}(\varepsilon^{2/3})$  and that  $p^* = q_M + \Delta^{1/3} \varepsilon^{1/3} + \mathcal{O}(\varepsilon^{2/3})$ , where  $\Delta$  is defined below in (7.1). In general, further progress was difficult because characterisations of the value function depended on finding a solution to a non-linear second order free-boundary problem, with two free boundaries. One case that is slightly simpler is the case of logarithmic utility (which formally may be considered as the case  $R = 1 = S$  in our setting). In the case of logarithmic utility Gerhold et al. [15] focus on the dual problem to give an expansion for  $p_*$ ,  $p^*$  to order  $\mathcal{O}(\varepsilon^{2/3})$  and an expansion for the optimal consumption which is valid up to order  $\mathcal{O}(\varepsilon)$ .

Still in the additive case, Choi et al. [8] and Hobson et al. [22], showed how the problem for general

<sup>7</sup>We have  $\kappa(q) = \gamma^\uparrow \exp\left(\frac{S}{1-S} \int_{q_*(\xi)}^q \frac{n'_\xi(v)}{vn_\xi(v)} dv\right) = \frac{1}{\gamma^\downarrow} \exp\left(-\frac{S}{1-S} \int_q^{q^*(\xi)} \frac{n'_\xi(v)}{vn_\xi(v)} dv\right)$ .

<sup>8</sup>It is important to note that our method gives a solution which is valid for all values of transaction costs, and then we look in the small transaction cost regime. Many other studies in the small transaction cost regime only construct solutions as expansions about the frictionless case.

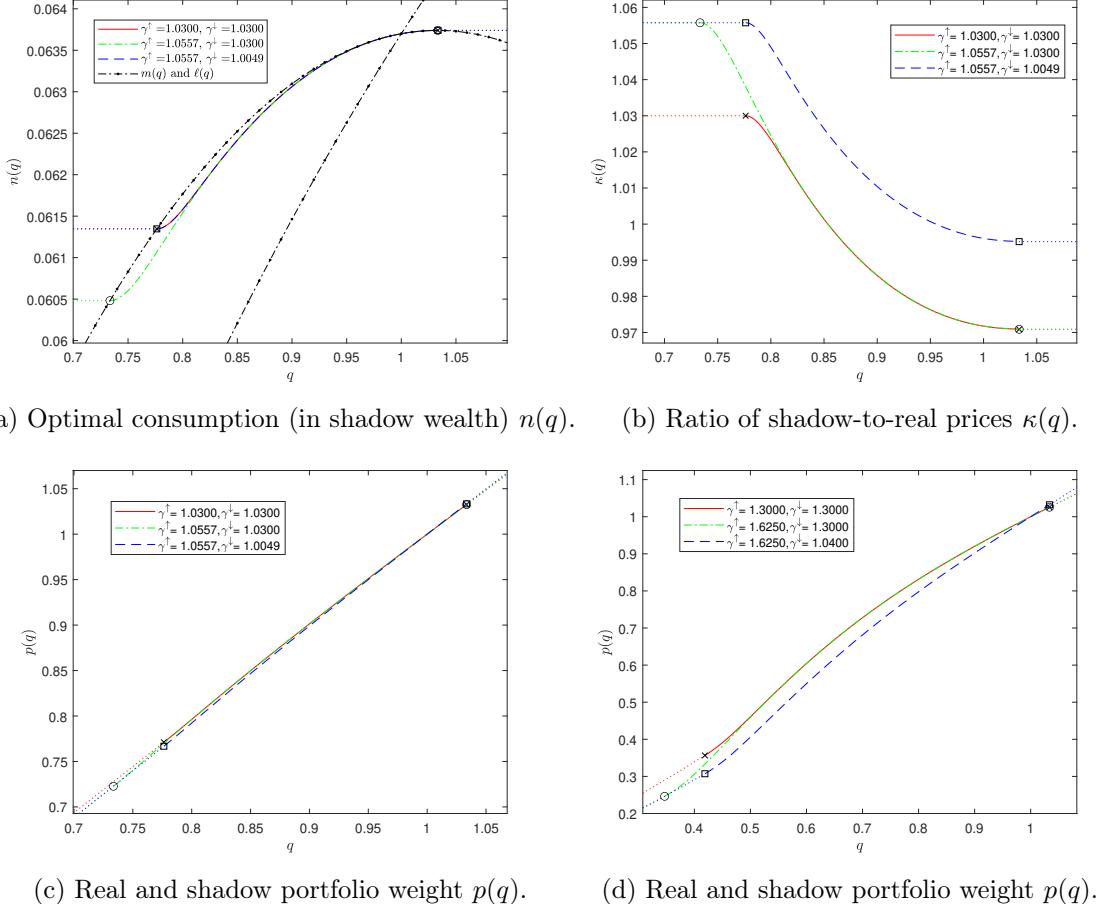


Figure 3: The fundamental quantities as a function of shadow portfolio weights in Example 2. Parameters used:  $R = 2$ ,  $S = 4$ ,  $\delta = 0.1$ ,  $r = 0$ ,  $\mu = 0.1$ ,  $\sigma = 0.22$  (such that  $\lambda = 0.45455$  and  $\alpha = 0.1$ ). The cross, circle and square markers indicate the boundary points of the no-transaction region (in  $p_*^*$  or  $q_*^*$ ) under the set of transaction costs  $(\gamma^+, \gamma^-)$  given by  $(1.03, 1.03)$ ,  $(1.05575, 1.03)$  and  $(1.05575, 1.00488)$  respectively. Panel (d) is identical to Panel (c) except the three sets of transaction costs  $(\gamma^+, \gamma^-)$  are now chosen to be  $(1.3, 1.3)$ ,  $(1.625, 1.3)$  and  $(1.625, 1.04)$ . The purpose of Panel (d) is to illustrate features of the relationships between the portfolio weights as functions of the transaction costs which may be difficult to ascertain in Panel (c) given that the curves lie almost on top of each other.

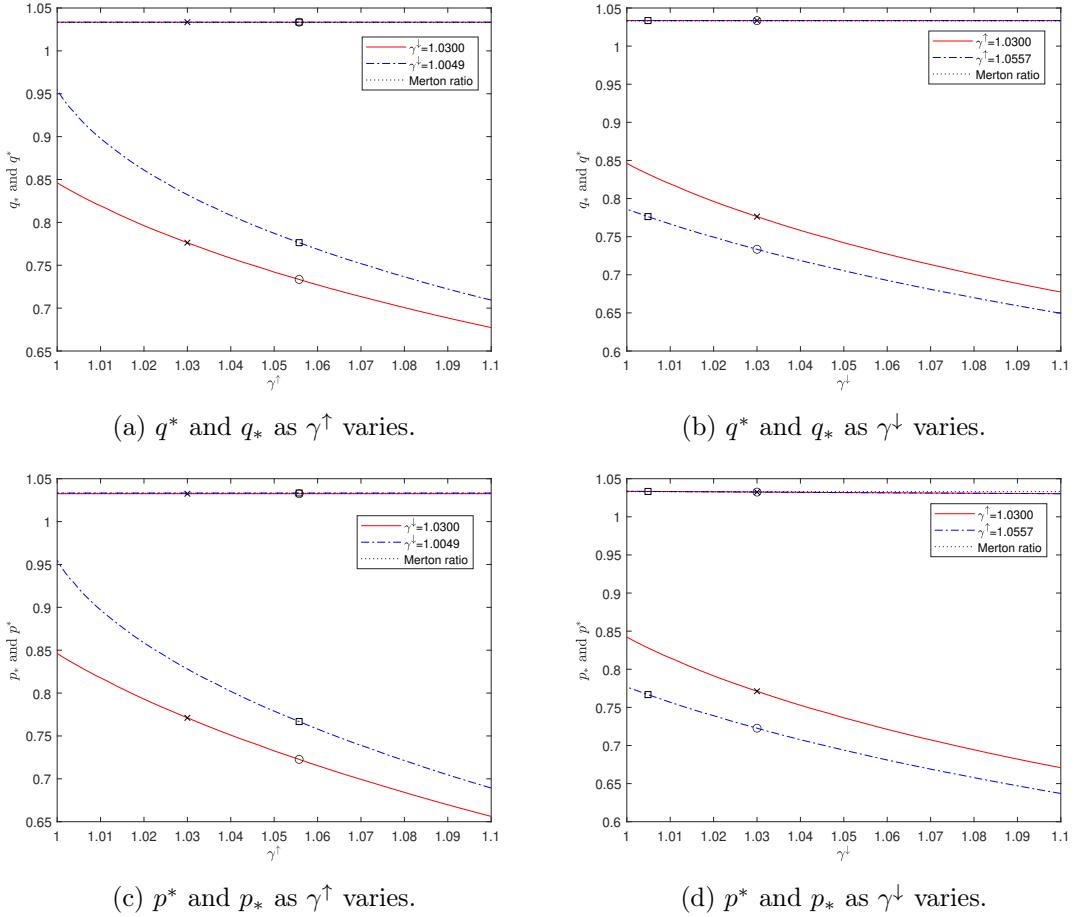


Figure 4: Behaviours of the no-transaction region as a function of transaction costs in Example 2. Parameters used:  $R = 2$ ,  $S = 4$ ,  $\delta = 0.1$ ,  $r = 0$ ,  $\mu = 0.1$ ,  $\sigma = 0.22$  (such that  $\lambda = 0.45455$  and  $\alpha = 0.1$ ). The cross, circle and square markers indicate the boundary points of the no-transaction region (in  $p_*$  or  $q_*$ ) under the set of transaction costs  $(\gamma^{\uparrow}, \gamma^{\downarrow})$  given by  $(1.03, 1.03)$ ,  $(1.05575, 1.03)$  and  $(1.0557, 1.00488)$  respectively.

transaction costs can be reduced to a first order equation. This facilitated a simpler derivation of the expansion for small transaction costs, see Choi [6] and Hobson et al. [22], and these papers calculated the second order term. Choi [6] assumed zero transaction costs on selling (i.e. that  $\gamma^\downarrow = 1$ ), but his results can be translated to the more general case. Hobson et al. [22] additionally considered the expansion in the case  $q_M = 1$ , in which case the width of the no transaction wedge is of order  $\varepsilon^{1/2}$ . (We could extend our results to this case similarly).

Melynyk et al. [28] considered the small transaction cost case for Epstein-Zin SDU. Under some restrictions on the parameters (which go beyond well-posedness of the problem) and under the assumption that the transaction costs on buying and selling are identical, (i.e.  $\gamma^\downarrow - 1 = 1 - \frac{1}{\gamma^\uparrow}$ ) they derived an expansion to order  $\varepsilon^{2/3}$  for the value function and thence find expansions (again accurate to order  $\varepsilon^{2/3}$ ) for the boundaries of the optimal no-transaction region and the optimal consumption.

Define

$$\begin{aligned}\Delta &= \frac{3q_M^2(1-q_M)^2}{4R}; \\ \Sigma &= \frac{4m_M}{3\sigma^2q_M(1-q_M)^2} = \frac{m_Mq_M}{R\sigma^2\Delta}; \\ \Psi &= \frac{1}{5}\Sigma^2 + \frac{1}{5}\frac{(5q_M-3)}{q_M(1-q_M)}\Sigma - \frac{(3-10q_M+10q_M^2)+4Rq_M(1-q_M)}{15q_M^2(1-q_M)^2}.\end{aligned}\tag{7.1}$$

**Proposition 7.1.** *Suppose that  $R, S < 1$  and  $m_M \geq 0$  or  $R, S > 1$  and  $m_M > 0$ , and suppose that  $q_M \notin \{0, 1\}$ . Then for all  $\varepsilon > 0$  sufficiently small*

$$\begin{aligned}q_*(1+\varepsilon) &= q_M - \Delta^{\frac{1}{3}}\varepsilon^{\frac{1}{3}} - \Sigma\Delta^{\frac{2}{3}}\varepsilon^{\frac{2}{3}} - \Psi\Delta\varepsilon + \mathcal{O}\left(\varepsilon^{\frac{4}{3}}\right) \\ q^*(1+\varepsilon) &= q_M + \Delta^{\frac{1}{3}}\varepsilon^{\frac{1}{3}} - \Sigma\Delta^{\frac{2}{3}}\varepsilon^{\frac{2}{3}} + \Psi\Delta\varepsilon + \mathcal{O}\left(\varepsilon^{\frac{4}{3}}\right)\end{aligned}$$

**Remark 7.2.** (a) The assumption of Proposition 7.1 that either  $R, S < 1$  and  $m_M \geq 0$  or  $R, S > 1$  and  $m_M > 0$  is equivalent to an assumption that the problem (4.3) is well-posed for all sufficiently small transaction costs.

- (b) In the shadow-fraction of wealth co-ordinates, the width of no-transaction region is  $\mathcal{O}(\varepsilon^{1/3})$  and is symmetric about the Merton proportion  $q_M$ . To order  $\varepsilon^{1/3}$  the width of the no-transaction region depends on risk aversion  $R$  (both directly, and indirectly through  $q_M = \frac{\lambda}{\sigma R}$ ), but not on  $S$ , the reciprocal of the EIS.
- (c) The  $\mathcal{O}(\varepsilon^{2/3})$  term is same for both  $q_*$  and  $q^*$  and moves the no-transaction region in the direction of smaller positions in the risky asset. This term does not depend directly on  $S$ , although it does depend indirectly on  $S$  via the frictionless consumption rate  $m_M$ .
- (d) The  $\mathcal{O}(\varepsilon^{2/3})$  term is same for both  $q_*$  and  $q^*$ , except that the signs are opposite, so that rather than moving the no-transaction region it makes it wider or narrower. This term does not depend directly on  $S$ , although it does depend indirectly on  $S$  via the frictionless consumption rate  $m_M$ .
- (e) It is straightforward to extend the arguments to higher order as required, or to the case  $q_M = 1$ . If  $\lambda = 0 = q_M$  then the problem is degenerate and optimal strategy for the agent is to instantly eliminate any position (long or short) in the risky asset, and thereafter to keep all wealth in the bank account.

**Corollary 7.3.** *Suppose that  $R, S < 1$  and  $m_M \geq 0$  or  $R, S > 1$  and  $m_M > 0$ , and suppose that  $q_M \notin \{0, 1\}$ . Then for all  $\varepsilon^\uparrow, \varepsilon^\downarrow > 0$  sufficiently small*

$$\begin{aligned}p_*(1+\varepsilon^\uparrow, 1+\varepsilon^\downarrow) &= q_M - \Delta^{\frac{1}{3}}(\varepsilon^\uparrow + \varepsilon^\downarrow)^{\frac{1}{3}} - \Sigma\Delta^{\frac{2}{3}}(\varepsilon^\uparrow + \varepsilon^\downarrow)^{\frac{2}{3}} - \Psi\Delta(\varepsilon^\uparrow + \varepsilon^\downarrow) - \varepsilon^\uparrow q_M(1-q_M) + \mathcal{O}(\varepsilon^{\frac{4}{3}}) \\ p^*(1+\varepsilon^\uparrow, 1+\varepsilon^\downarrow) &= q_M + \Delta^{\frac{1}{3}}(\varepsilon^\uparrow + \varepsilon^\downarrow)^{\frac{1}{3}} - \Sigma\Delta^{\frac{2}{3}}(\varepsilon^\uparrow + \varepsilon^\downarrow)^{\frac{2}{3}} + \Psi\Delta(\varepsilon^\uparrow + \varepsilon^\downarrow) + \varepsilon^\downarrow q_M(1-q_M) + \mathcal{O}(\varepsilon^{\frac{4}{3}})\end{aligned}$$

**Remark 7.4.** (a) In the original coordinates, the width of the no-trade region is  $\mathcal{O}(\varepsilon^{1/3})$  and to leading order is symmetric about the Merton proportion  $q_M$ . Also to leading order, the width of the no-transaction region depends on risk aversion  $R$  (both directly, and indirectly through  $q_M = \frac{\lambda}{\sigma R}$ ), but not on the reciprocal of the EIS  $S$ . See also Janeček and Shreve [23, Remark 1] in the additive case, and Melnyk et al. [28] for the corresponding result under EZ-SDU.

- (b) The  $\mathcal{O}(\varepsilon^{2/3})$  term is same for both  $p_*$  and  $p^*$  and moves the no-transaction region in the direction of smaller positions in the risky asset. This term does not depend directly on  $S$ , although it does depend indirectly on  $S$  via  $m_M$ . See also [23, 28].
- (c) The individual transaction costs  $\varepsilon^\uparrow$  and  $\varepsilon^\downarrow$  only enter at linear order. The linear order term does not depend on the  $S$ , except indirectly through  $m_M$ . This is the final term with this property.

**Proposition 7.5.** Suppose that  $R, S < 1$  and  $m_M \geq 0$  or  $R, S > 1$  and  $m_M > 0$ , and suppose that  $q_M \notin \{0, 1\}$ . Let  $h_3(v) = (1+v)^2(2-v)$ . Then for all  $\varepsilon^\uparrow, \varepsilon^\downarrow > 0$  sufficiently small

$$\begin{aligned} C_{1+\varepsilon^\uparrow, 1+\varepsilon^\downarrow}^*(t, x, y, \phi) &= C_{1,1}^*(t, x, y, \phi) \left( 1 + \frac{R(1-S)}{S} \frac{\sigma^2}{2m_M} \Delta^{2/3} (\varepsilon^\uparrow + \varepsilon^\downarrow)^{2/3} \right. \\ &\quad \left. + \frac{q_M}{2S} \left( 1 - \frac{1}{2} h_3 \left( \frac{p - q_M}{\Delta^{1/3} \varepsilon^{1/3}} \right) \right) (\varepsilon^\uparrow + \varepsilon^\downarrow) + \frac{q_M}{2} (\varepsilon^\uparrow - \varepsilon^\downarrow) \right) + \mathcal{O}(\varepsilon^{4/3}) \end{aligned}$$

where  $p = \frac{y\phi}{x+y\phi}$  and  $C_{1,1}^*(t, x, y, \phi) := (x + y\phi)m_M$  is the optimal consumption under zero transaction costs.

**Remark 7.6.** (a) To zeroth order consumption is equal to optimal consumption without transaction costs, and there is no term of order  $\varepsilon^{1/3}$ . (See also Janeček and Shreve [23, Remark 2] in the additive case and Melnyk et al. [28, Corollary 6.1] in the case of EZ-SDU.)

- (b) The first non-zero correction term is independent of  $p = \frac{y\phi}{x+y\phi}$  and is of order  $\varepsilon^{2/3}$ , ([23] and [28] make a similar observation.) It does not depend on the individual transaction costs. Mathematically, the fact that the consumption is constant arises from the fact that solutions  $n_{(q_M-u)}^+$  are (approximately) horizontal for small  $u$ , (since  $n_{(q_M-u)}^+$  has zero derivative at both endpoints where it intersects with  $m$ ). Here  $n_{(q_M-u)}^+$  is the solution to  $n = O(n, q)$  started at  $(q_M - u, n(q_M - u) = m(q_M - u))$ , and is defined fully in Appendix A.1.
- (c) The sign of the first non-zero correction term depends on the sign of  $(1 - S)$ . If  $S < 1$  then the agent consumes more than in the zero-transaction case; if  $S > 1$  then the agent consumes less.<sup>9</sup> At its heart, the explanation for this relationship is that given in Remark 3.5 about the impact of investing a sub-optimal fraction of wealth in the risky asset, and the effect this has on the certainty equivalent wealth of the agent. The impact of transaction costs is to cause the agent to invest a sub-optimal fraction of wealth in the risky asset. If  $S < 1$  the fact that the certainty equivalent value of the

<sup>9</sup>Janeček and Shreve [23, Remark 2] and Melnyk et al. [28] also make this observation. Janeček and Shreve write ‘...the existence of transaction costs increases the size of consumption for  $R \in (0, 1)$ , while the consumption is decreased for  $R > 1$ . This is explained by the fact that the elasticity of inter-temporal substitution,  $1/R$ , is high for small  $R$ ’ – note that in their additive model  $R \equiv S$ . However, whilst the two facts (existence of transaction costs increases (respectively decreases) consumption for  $R \in (0, 1)$  (respectively  $R \in (1, \infty)$ ), and the elasticity of inter-temporal substitution  $1/R$  is high for small  $R$ ) are correct, no evidence is offered as to why the second fact is an explanation of the first, or why the sign of  $R - 1$  is critical. In the case of zero-transaction costs (and therefore also, approximately in the case of small transaction costs) the optimal consumption is  $m_M = \left( r + \frac{\lambda^2}{2R} \right) + \frac{1}{S} \left( \delta - r - \frac{\lambda^2}{2R} \right)$ . Assuming that  $\delta - r - \frac{\lambda^2}{2R} > 0$ , the primary effect of increasing  $S$  is to reduce the optimal consumption, but this does *not* give an explanation of why the sign of consumption changes relative to the zero-transaction cost case depend on the sign of  $(1 - S)$ . For this a more subtle argument is needed. Melnyk et al. [28, Page 1147] point out that in the case of stochastic differential utility the reciprocal of the EIS is captured by  $S$  (and is separate to the risk aversion  $R$ ) but defer to Janeček and Shreve for an explanation. We return to this issue in Section 8.3.

agent's holdings goes down (relative to the zero transaction cost case) leads the agent to increase their instantaneous consumption. When  $S > 1$ , again the presence of transaction costs causes the agent to invest a sub-optimal fraction of wealth in the risky asset. This time this leads the agent to reduce instantaneous consumption. See Section 8.3 for further development of this argument.

- (d) We expect that when  $R = S = 1$  we recover the expansion for logarithmic utility, as studied by Gerhold et al. [15]. Indeed, in this case Gerhold et al. argue that the leading order correction to the optimal consumption is of order  $\varepsilon$  (and not of order  $\varepsilon^{2/3}$ ). Our results are consistent with this claim.
- (e) Individual transaction costs enter at  $\mathcal{O}(\varepsilon)$ . This is also the first term at which the fraction  $p$  of wealth invested in the risky asset affects the consumption.

## 8 Comparative statics

### 8.1 Comparative statics of the boundaries in $S$

**Proposition 8.1.** *Fix all the model parameters except  $S$ . Suppose  $\hat{S} > \tilde{S}$  and the problems under  $S = \hat{S}$  and  $S = \tilde{S}$  are well-posed where the optimal purchase and sale boundaries are  $(\hat{p}_*, \hat{p}^*)$  and  $(\tilde{p}_*, \tilde{p}^*)$  respectively.*

- (a) *If*

$$\delta - r - \frac{\lambda^2}{2R} \geq 0, \quad (8.1)$$

*then  $\hat{p}_* \geq \tilde{p}_*$  and  $\hat{p}^* \geq \tilde{p}^*$ .*

- (b) *Suppose instead*

$$\delta - r - \frac{\lambda^2}{2R} < 0. \quad (8.2)$$

*If the problems under  $S = \hat{S}$  and  $S = \tilde{S}$  are well-posed for all sufficiently small transaction costs, then there exists  $\varepsilon > 0$  (which may depend on  $\hat{S}, \tilde{S}$ ) such that  $\hat{p}_* \leq \tilde{p}_*$  and  $\hat{p}^* \leq \tilde{p}^*$  for all  $\xi = \gamma^\uparrow \gamma^\downarrow \in [1, 1 + \varepsilon]$ .*

The proof of Proposition 8.1 is given in Appendix C.1.

Note that although we have made no assumptions about the sign of  $r$  it is natural to assume that  $r > 0$  and then  $r + \frac{\lambda^2}{2R} > 0$ . Then it is easily seen that the condition for the problem with (sufficiently) small transaction costs to be well-posed, namely  $S \left( r + \frac{\lambda^2}{2R} \right) > r + \frac{\lambda^2}{2R} - \delta$ , is equivalent to  $S > 1 - \frac{\delta}{r + \frac{\lambda^2}{2R}}$ .

**Corollary 8.2.** *Fix all the model parameters except  $S$ .*

*Suppose that  $\delta \geq r + \frac{\lambda^2}{2R} > 0$ . Suppose  $S < 1$ . Then the problem is always well-posed and  $p_*$  and  $p^*$  are increasing in  $S$  for all levels of transaction costs. Conversely, suppose  $S > 1$ . Then  $p_*$  and  $p^*$  are increasing in  $S$  over the range of transaction costs for which the problem is well-posed. In particular  $p_*$  and  $p^*$  are increasing in  $S$  for small transaction costs.*

*Suppose that  $r + \frac{\lambda^2}{2R} > \max\{\delta, 0\}$ . Then the problem is well-posed for small transaction costs for  $S > \max\{1 - \frac{\delta}{r + \frac{\lambda^2}{2R}}, 0\}$  and then, still for small transaction costs,  $p_*$  and  $p^*$  are decreasing in  $S$  over this range.*

Each of Figure 5 and 6 shows the two cases described in Proposition 8.1 where the monotonicity behaviours of  $p_*$ ,  $p^*$  can be different. To understand why different cases can arise, recall that in the frictionless case the optimal consumption rate is given by

$$m_M = m(q_M) = r + \frac{\lambda^2}{2R} + \frac{\delta - r - \frac{\lambda^2}{2R}}{S}.$$

The sign of  $\delta - r - \frac{\lambda^2}{2R}$  therefore governs whether, in the absence of transaction costs, the (constant) optimal consumption rate increases or decreases in  $S$ . Suppose  $\delta - r - \frac{\lambda^2}{2R}$  is positive. Then the agent with lower  $S$  wants to consume at a higher rate whilst maintaining a constant fraction of wealth invested in the risky asset given by the Merton ratio. As trading does not incur transaction costs, the higher level of consumption can be supported interchangeably by sale of the risky asset and/or withdrawal of cash from the riskless account.

With transaction costs, we expect that the sign of  $\delta - r - \frac{\lambda^2}{2R}$  affects the monotonicity of the optimal consumption rate with  $S$  in a similar fashion (at least when transaction costs are small). If  $\delta - r - \frac{\lambda^2}{2R}$  is positive, the agent with lower  $S$  again wants to consume more, but now such an agent will prefer to finance consumption from cash wealth rather than from sales of the risky asset because of the market frictions. Hence the agent with lower  $S$  in general has a stronger incentive to hold cash in anticipation of the need to consume more, resulting in a downward shift of the no-transaction wedge  $[p_*, p^*]$ . The opposite will happen when  $\delta - r - \frac{\lambda^2}{2R}$  is negative. In general, the variation of  $p^*$  in  $S$  is more prominent under higher level of transaction costs.

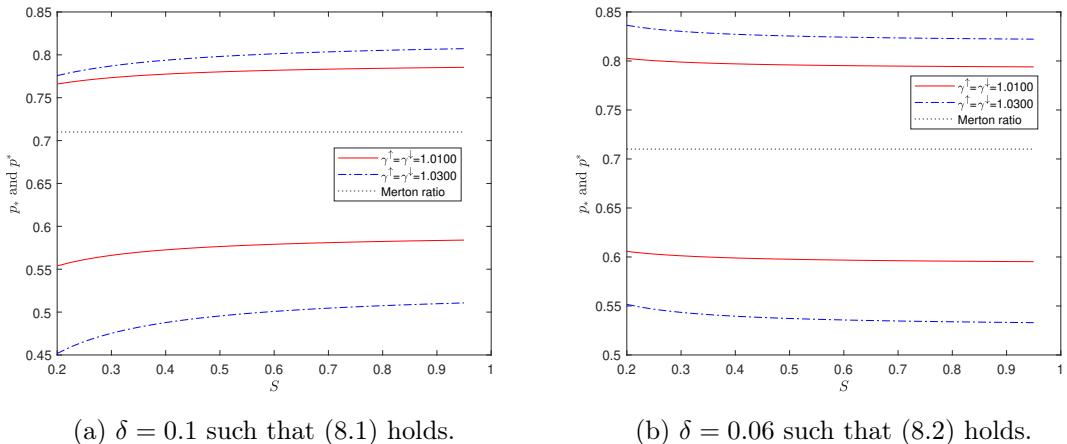


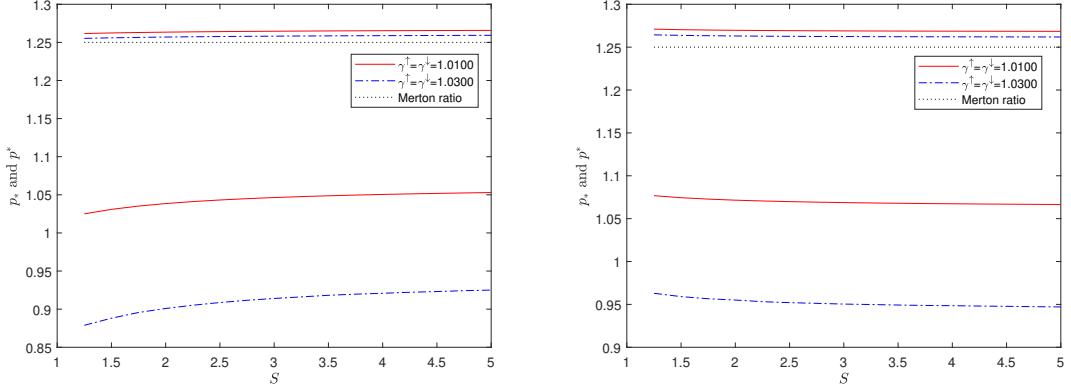
Figure 5:  $p_*$  and  $p^*$  as  $S$  varies when  $R, S < 1$  under  $\gamma^+ = \gamma^- = 1.01$  and  $\gamma^+ = \gamma^- = 1.03$ . Base parameters used:  $R = 2/3$ ,  $r = 0$ ,  $\mu = 0.2$ ,  $\sigma = 0.65$ .

## 8.2 Comparative statics of the boundaries in $R$

The proof of Proposition 8.3 is given in Appendix C.2.

**Proposition 8.3.** *Suppose  $\lambda > 0$ . Fix all the model parameters except  $R$ . Let  $\hat{R} > \tilde{R}$  be such that the problems under  $R = \hat{R}$  and  $R = \tilde{R}$  are well-posed where the optimal purchase and sale boundaries are  $(\hat{p}_*, \hat{p}^*)$  and  $(\tilde{p}_*, \tilde{p}^*)$  respectively. If any of the conditions in Lemma C.1 holds for  $R \in \{\hat{R}, \tilde{R}\}$ , then  $\hat{p}_* \leq \tilde{p}_*$  and  $\hat{p}^* \leq \tilde{p}^*$ . Consequently, subject to the well-posedness of the problem:*

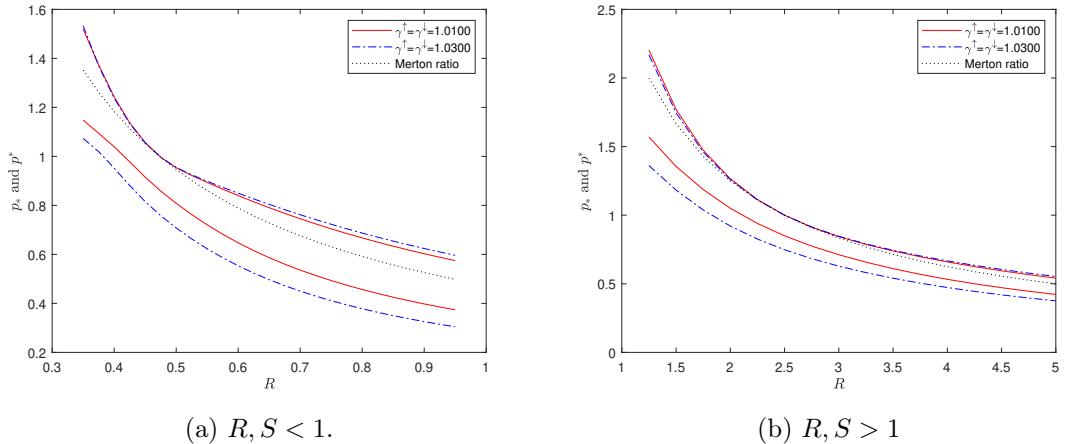
- (a) *If  $S < 1$  then  $p_*$  and  $p^*$  are decreasing in  $R$  for any  $R \in (0, 1)$ .*
- (b) *If  $S > 1$  and  $m(0) > 0$  then  $p_*$  and  $p^*$  are decreasing in  $R$  for  $R \in [2, \infty)$ . If also  $\lambda \geq \frac{\sigma}{2}$  then  $p_*$  and  $p^*$  are decreasing in  $R$  for  $R \in (1, \infty)$ .*



(a)  $\delta = 0.1$  such that (8.1) holds.

(b)  $\delta = 0.05$  such that (8.2) holds.

Figure 6:  $p_*$  and  $p^*$  as  $S$  varies when  $R, S > 1$  under  $\gamma^+ = \gamma^- = 1.01$  and  $\gamma^+ = \gamma^- = 1.03$ . Base parameters used:  $R = 2$ ,  $r = 0$ ,  $\mu = 0.1$ ,  $\sigma = 0.2$ .



(a)  $R, S < 1$ .

(b)  $R, S > 1$

Figure 7:  $p_*$  and  $p^*$  as  $R$  varies under  $\gamma^+ = \gamma^- = 1.01$  and  $\gamma^+ = \gamma^- = 1.03$ . Base parameters used in Panel (a):  $S = 1/3$ ,  $\delta = 0.1$ ,  $r = 0$ ,  $\mu = 0.2$ ,  $\sigma = 0.65$ . Base parameters used in Panel (b):  $S = 4$ ,  $r = 0$ ,  $\delta = 0.1$ ,  $\mu = 0.1$ ,  $\sigma = 0.2$ . All parameters (and the values of  $R$  considered) are such that at least one of the sufficient conditions in Lemma C.1 holds.

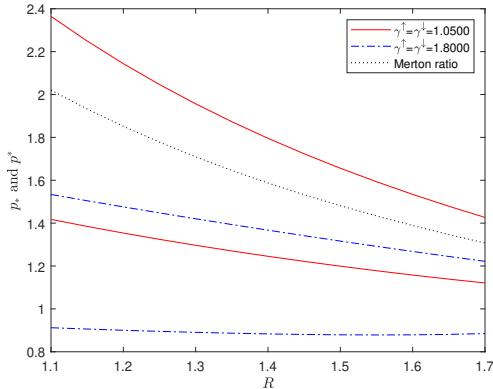


Figure 8:  $p_*$  and  $p^*$  as  $R$  varies when all the stated conditions in Lemma C.1 do not hold. Transaction costs considered are  $\gamma^+ = \gamma^- = 1.05$  and  $\gamma^+ = \gamma^- = 1.8$ . Base parameters used:  $S = 2$ ,  $r = 0$ ,  $\delta = -0.11$ ,  $\mu = 0.2$ ,  $\sigma = 0.3$ . In this example,  $p_*$  is not monotonic in  $R$  under  $\gamma^+ = \gamma^- = 1.8$ .

The main conclusion from Proposition 8.3 is that in the typical case  $p_*$  and  $p^*$  are decreasing in  $R$ . See Figure 7. This result is proven to hold over the regime where the problem is well-posed whenever  $R < 1$  and also when  $R > 1$  unless either  $m(0) \leq 0$  or  $R \in (1, 2)$  and  $\lambda < \frac{\sigma}{2}$ . The condition  $m(0) > 0$  is equivalent to  $\delta + r(S-1) < 0$ , which is automatically satisfied if  $R, S > 1$  and  $\delta, r > 0$ . Such monotonicity behaviour of  $p_*$  and  $p^*$  in  $R$  is perhaps not too surprising in view of the fact that the Merton ratio  $q_M = \lambda/(R\sigma)$  is decreasing in  $R$ . Nonetheless, it is useful to stress that the consumption motive will also affect  $p_*$  and  $p^*$  as  $R$  varies. Note that the optimal consumption rate in the frictionless case can be rewritten as

$$m_M = m(q_M) = r + \frac{\delta - r}{S} + \frac{\lambda^2}{2R} \frac{S-1}{S}. \quad (8.3)$$

Then for a fixed  $S < 1$ , the (frictionless) consumption rate is increasing in  $R$ . Following similar intuition to that used in Section 8.1, in the case with frictions we expect that the elevated consumption due to an increase in  $R$  will make the agent more inclined to hold cash (i.e. a smaller fraction of wealth is invested in the risky asset). This effect is in the same direction as the impact of  $R$  on the Merton ratio (when  $S < 1$ ).

However, if  $S > 1$  then (8.3) suggests the optimal frictionless consumption rate is decreasing in  $R$ . The (lack of) consumption motive will then encourage the agent to hold less cash and hence the fraction of wealth invested in the risky asset becomes larger. The effect of  $R$  on the investment level through the consumption motive is therefore opposite in sign to the direct effect of  $R$  on the Merton ratio, and the net effect on the investment level becomes ambiguous. This is the reason why in the case  $S > 1$  we require further conditions in Proposition 8.3 to be able to conclude that  $p_*$  and  $p^*$  are decreasing in  $R$ . Indeed, there exist counterexamples such that the boundaries are not decreasing in  $R$ , although usually they require somewhat extreme model parameters such as a negative discount rate in conjunction with very high transaction costs. See Figure 8.

### 8.3 What determines whether consumption increases or decreases with the introduction of transaction costs?

In this section we give three related arguments which together justify the comments made in Remarks 3.5 and 7.6(c).

Consider the continuous-time investment-consumption problem for EZ-SDU in the frictionless case. Suppose the agent uses constant proportional strategies so that the value function is as given in (3.6) where  $H$  is as given in (3.5), i.e.

$$H(q, m) = \delta + (S-1) \left( r - m + \lambda \sigma q - \frac{1}{2} q^2 \sigma^2 R \right).$$

Then, provided  $H(q, m) > 0$ , and with  $z = x + y\theta$  denoting initial wealth,  $V_0 = \frac{z^{1-R}}{1-R} \left( \frac{m^{1-S}}{H(q, m)} \right)^\theta$ . Maximising  $V_0$  is equivalent to maximising  $((1-R)V_0)^{\frac{1}{1-R}} = z \frac{m}{H(q, m)^{\frac{1}{1-S}}}$ . Note that this quantity is expressed in economically meaningful units as it represents the certainty equivalent value.

As in Section 3.1, suppose that the fraction of wealth invested in the risky asset is constrained to be equal to  $q$ , where  $q$  is fixed and given, and possibly suboptimal, and consider choosing the optimal consumption rate. The maximum certainty equivalent can be found by differentiation and we get that the optimal consumption rate  $m$  is given by  $m = m(q)$  and solves  $\frac{1}{m} = \frac{H_m(q, m)}{(1-S)H(q, m)}$ . Since  $H_m(q, m) = (1-S)$  this simplifies to  $m(q) = H(q, m(q))$ , so that  $m(q)$  is the root of the decreasing (in  $m$ ) function  $h_q$  where  $h_q(m) = H(q, m) - m$ .

Suppose  $S < 1$ . Then  $H(q, m)$  has a minimum in  $q$  at  $q_M$  (independent of the value of  $m$ ). In particular, for  $q \neq q_M$ ,  $0 = h_q(m(q)) = H(q, m(q)) - m(q) > H(q_M, m(q)) - m(q) = h_{q_M}(m(q))$  (where the strict inequality follows from the strict convexity of  $H$  in  $q$ ). Then, since  $h_{q_M}$  is decreasing we must

have  $m(q_M) < m(q)$ . In particular,  $m(\cdot)$  has a minimum at  $q_M$ . Conversely, suppose  $S > 1$  so that  $H(q, m)$  has a maximum in  $q$  at  $q_M$ . Then,  $0 = h_q(m(q)) = H(q, m(q)) - m(q) \geq H(q_M, m(q)) - m(q) = h_{q_M}(m(q))$  and since  $h_{q_M}$  is decreasing we must have  $m(q_M) \geq m(q)$ . In particular,  $m(\cdot)$  has a maximum at  $q_M$  (and the uniqueness of the minimiser follows from the uniqueness of the maximiser of  $H$  in  $q$  as before).

It follows that (in the zero-transaction cost case) when  $S < 1$  the impact of using a non-optimal investment strategy, and then optimising over consumption is to increase the immediate consumption rate relative to the optimal consumption rate under an optimal investment strategy. When  $S > 1$ , the impact of using a non-optimal investment strategy, and then optimising over consumption is to decrease the immediate consumption rate.

One impact of transaction costs is to force the agent to use a suboptimal investment strategy when compared with the frictionless world. The above analysis then explains whether the impact of transaction costs on the consumption rate is to increase or decrease instantaneous consumption (relative to the frictionless case) and why this depends exactly on the shape of  $H(q, m(q))$  (does it have a minimum or a maximum at  $q_M$ ) which in turn is directly related to the sign of  $(S - 1)$ .

The above argument might be considered a mathematical justification of the role of  $(S - 1)$  in determining the impact of the introduction of transaction costs on consumption, but it does not directly relate to the economic explanation of the fundamental role of  $S - 1$  given in Remark 3.5 in terms of the interplay between two impacts of reducing the certainty equivalent value of future consumption. To understand what is happening at a more fundamental level, we consider recursive utility over one time period.

Consider an agent in a one-period economy (with times 0 and 1) containing a risk-less asset and a risky asset. The agent has initial wealth  $x$  and recursive preferences. The time-0 controls available to the agent are to decide the amount  $c \in (0, 1)$  to consume at  $t = 0$ , and the number of units  $\phi$  of the risky asset to hold between times 0 and 1. We assume that the price of the risk-less asset is normalised so that it is constant over time, that the initial price of the risky asset is 1 and that the return on the risky asset is given by the time-1 measurable random variable  $Z$  (so that 1 invested in the risky asset at time 0 yields  $(1 + Z)$  at time-1). It follows that the wealth of the agent at time 1 (we assume no frictions) is  $x - c + \phi Z$ . We assume that all wealth at time 1 is consumed at time 1.

Under quasi-arithmetic recursive preferences, the aim of the agent is to maximise the certainty equivalent value

$$f^{-1}(f(c) + f \circ g^{-1}(\mathbb{E}[g(x - c + \phi Z)])) \quad (8.4)$$

where  $f$  and  $g$  are increasing, concave functions. The example to consider is  $f(z) = \frac{z^{1-S}}{1-S}$  and  $g(z) = \frac{z^{1-R}}{1-R}$  in which case the recursive utility is precisely the discrete-time (one-period) analogue of the EZ-SDU we are considering in the main body of the paper.

We specialise to the case of  $g(z) = \frac{z^{1-R}}{1-R}$ , but leave  $f$  general. Then, with  $w = \frac{\phi}{x-c}$  we find  $g^{-1}(\mathbb{E}[g(x - c + \phi Z)]) = (x - c)D(w)$  where  $D(w) = g^{-1}(\mathbb{E}[g(1 + wZ)]) = (\mathbb{E}[(1 + wZ)^{1-R}])^{\frac{1}{1-R}}$ . Then maximising (8.4) simplifies to first maximising  $D$  with respect to  $w$ , and then with  $\hat{w} = \text{argmax } D(w)$ , maximising

$$f^{-1}(f(c) + f((x - c)D(\hat{w})))$$

We can also consider the problem of maximising the certainty equivalent value for a given investment strategy. Suppose  $w$  (and therefore  $D = D(w)$ ) is fixed. Then, since  $f$  is increasing, it is sufficient to maximise  $f(c) + f(D(x - c))$ . Assuming differentiability of  $f$  (and an interior maximum, for which a sufficient condition is  $f'(0) = \infty$ ) the maximiser  $\hat{c} = \hat{c}(x, D)$  solves

$$f'(c) = Df'(D(x - c)) = \frac{1}{(x - c)} [yf'(y)]|_{y=D(x-c)}. \quad (8.5)$$

We want to consider the impact of changing  $D$ , and it is clear that this depends on whether  $yf'(y)$  is increasing or decreasing in  $y$ . By the concavity of  $f$ , the left expression of (8.5) is decreasing in  $c$  and the middle expression is increasing in  $c$ . Therefore, if  $yf'(y)$  is increasing in  $y$  then  $\hat{c}(x, D)$  is decreasing

in  $D$ ; conversely, if  $yf'(y)$  is decreasing in  $y$  then  $\hat{c}(x, D)$  is increasing in  $D$ . For  $f(z) = \frac{z^{1-S}}{1-S}$ ,  $zf'(z)$  is decreasing precisely when  $S > 1$ , so that  $\hat{c}(x, D)$  is increasing in  $D$  if and only if  $S > 1$ <sup>10</sup>.

Finally we combine this argument with the impact of using a sub-optimal investment strategy as in the continuous-time case: if  $w^*$  is optimal then  $D(w) < D(w^*)$  and optimal consumption in markets with frictions is lower than in the frictionless case if and only if  $S > 1$ .

The third argument is related to Remark 3.5 and connects the above analysis to the impact of  $1/S$  as a measure of the inter-temporal substitutability of consumption – high  $S$  indicates a strong preference for smooth consumption over time. Consider the first equality in (8.5).  $D$  enters into the expression  $Df'(D(x - c))$  twice. First, the factor  $D$  means that decreasing  $D$  lowers the expression, resulting in higher optimal consumption. This role for  $D$  is related to the impact of changes of future wealth on consumption and is independent of the preferences expressed via  $f$ . Second  $D$  is an argument of the factor  $f'(D(x - c))$  – here lowering  $D$  increases the marginal time-1 utility and therefore, in isolation, decreases time-0 consumption. This factor depends on the preferences via  $f'$  and is strongest when  $S$  is large. Thus, when  $S$  is small, the first effect dominates and using suboptimal investment strategies leads to increased consumption. When  $S$  is large, the desire for smooth consumption increases, the second effect dominates, and time-0 consumption falls.

## 9 Concluding remarks

In this paper we studied the optimal investment-consumption problem over the infinite horizon in a Black–Scholes–Merton financial market. Relative to the standard frictionless case first investigated by Merton [29, 30], the innovations are that we consider proportional transaction costs and non-additive preferences, as captured by Epstein–Zin stochastic differential utility. Our results determine when the problem is well-posed, and in this case, we give a complete description of the optimal investment and consumption strategies. A major innovation is our focus on the *shadow fraction of wealth in the risky asset*: it turns out that with this as the primary variable, the problem reduces to the study of a free-boundary problem for one-dimensional, singular ODE.

Even for additive (CRRA) preferences (studied first by Magill and Constantinides [26] and Davis and Norman [10]) our approach brings substantial simplifications and insights, and we are able give new (financial) interpretations to many of the quantities which arise.

The non-additive case brings extra technical challenges (especially as regards existence and uniqueness of the utility process associated with a given consumption stream, and the comparison theorem used to facilitate the verification lemma); we overcome these challenges by making use of previous results in the frictionless case, see [19].

Study of the non-additive case is increasingly important as there is mounting evidence that the additive case fails to accurately describe investor behaviour. This paper is the first to give a full solution of the problem in the non-additive case (previous results only covered the small transaction cost case and only for an incomplete set of parameter combinations) and therefore is the first to make it possible to properly study the comparative statics of the problem and to distinguish the separate roles played by the risk aversion (which primarily affects the width and location of the no-transaction wedge), and the elasticity of intertemporal substitution (which primarily affects the consumption rate).

The methods of this paper rely crucially on scaling properties of the problem (including the dynamics of the price process, the proportional transaction costs and the form of the recursive utility) to reduce the dimensionality, and it looks difficult to extend them beyond this case. Notwithstanding, looking at the extant literature in the additive case, it seems that there are many natural and important extensions of our problem—we give outlines for four—which would become tractable (including for SDU) using the methods of this paper.

---

<sup>10</sup>Alternatively for  $f(z) = \frac{z^{1-S}}{1-S}$  of power law form,  $\hat{c}$  solves  $c^{-S} = D^{1-S}(x - c)^{1-S}$ . It follows that  $\hat{c} = D^{1-\frac{1}{S}}x$  which is increasing in  $D$  if and only if  $S > 1$ .

The first extension is the problem with infinite transaction costs (and then it is important to distinguish between infinite transaction costs on selling and infinite transaction costs on buying), which is a precursor to an expansion solution for problems with a large transaction parameter. The case  $R, S > 1$  and  $m(0) < 0 < m(1)$  seems especially interesting as the point where  $\ell$  crosses zero becomes a singular point for the solution  $n_\infty$ .

The second extension is the problem with multiple assets. At the general level, with transaction costs to be paid on all exchanges of assets (perhaps also requiring all transactions to be made to or from the cash account) the intuition is that the no-transaction region becomes a no-transaction cone, see for example Chen and Dai [5]. However, whereas in the one-risky asset case, the no-trade region is an interval which can be characterised by its endpoints, in the multi-asset case, the cross-section of the no-transaction region will be convex, but otherwise may be complicated in shape and difficult to characterise. However, results in the additive case (especially Hobson et al. [21] and Choi [7]) give hope that it is possible to find a complete solution in a problem with multiple risky assets in which transaction costs are levied on purchases and sales of one asset only. Following the approach of this paper, the aim is to reduce the problem via the shadow fraction of wealth to the solution of a free-boundary problem for a first order differential equation. We even expect that apart from special cases, the resulting ODE is simpler than that in the single asset case studied in this paper in that there is no singularity at 1.

The third extension is to a problem in which there are different interest rates for lending and borrowing. In the frictionless case this impacts on  $m(q)$ , and the different interest rates lead to a discontinuity in the derivative of  $m$  at  $q = 1$ . Then, it can very easily occur that the turning point in  $m$  is at  $q_M = 1$ . Then, when transaction costs are introduced it follows that  $q_M = 1$  is no longer a special case, but rather a case of particular interest. Most papers exclude this case, not least because the expansions in Section 7 break down, but it can easily be covered by our methods.

The fourth extension, which is ongoing work, is to a problem with transaction costs and proportional dividends. In a problem without transaction costs, a model with dividends can be reduced to one without by allowing the investor the use the dividends to purchase additional units of stock. But, with transaction costs such repurchases would incur costs. Instead, if dividend rates are paid at proportional rate  $\beta$  the dynamics of  $Y$  become  $dY_t = (\mu - \beta)Y_t dt + \sigma Y_t dB_t$  and the wealth equation (4.1) becomes  $dX_t = (rX_t + \beta Y_t) dt - C_t dt - \gamma^\uparrow Y_t d\Phi_t^\uparrow + \frac{Y_t}{\gamma^\downarrow} d\Phi_t^\downarrow$ . The approach of the paper still applies with the proviso that  $m$  and  $\ell$  become different quadratics and that  $\kappa$  and  $n$  now solve a pair of coupled first-order ODEs. The presence of dividends provides a natural way for the portfolio to re-balance from the risky asset to cash without trading in stocks, so that we expect different behaviour to the case without dividends.

These extensions and other related problems are partly ongoing work and partly left to future research.

## A Additional results and proofs for the problem with frictions

### A.1 The free boundary problem

The following two results lie at the heart of Proposition 5.3.

**Proposition A.1.** *Let  $R < 1$ . Assume that  $m(0), m(1) > 0$ .*

(a) *Let  $q_M \in (0, 1)$ . Then for each  $z \in \{m > 0\} \cap (0, q_M)$ , there exist a unique  $\zeta^+(z) \in \{m > 0\} \cap (q_M, 1)$  and a unique continuously differentiable function  $n_{(z)}^+ : [z, \zeta^+(z)] \rightarrow [m(z), m(\zeta^+(z))]$  with  $n_{(z)}^+(q) < \ell(q)$  for  $q \in [z, \zeta^+(z)]$  that satisfies on  $[z, \zeta^+(z)]$  the ODE*

$$\frac{d}{dq} n_{(z)}^+(q) = O(n_{(z)}^+(q), q).$$

*as well as the boundary conditions*

$$n_{(z)}^+(z) = m(z) \quad \text{and} \quad n_{(z)}^+(\zeta^+(z)) = m(\zeta^+(z)).$$

Moreover, the map  $\Sigma^+ : \{m > 0\} \cap (0, q_M) \rightarrow (1, \infty)$  given by

$$\Sigma^+(z) := \exp \left( \int_z^{\zeta^+(z)} -\frac{1}{q(1-q)} \frac{m(q) - n_{(z)}^+(q)}{\ell(q) - n_{(z)}^+(q)} dq \right)$$

is continuous and decreasing, and satisfies

$$\begin{aligned} \lim_{z \downarrow 0 = \inf(\{m > 0\} \cap (0, q_M))} \Sigma^+(z) &= \infty \\ \lim_{z \uparrow \sup(\{m > 0\} \cap (0, q_M))} \Sigma^+(z) &= \begin{cases} \exp \left( \int_{q_-^m}^{q_+^m} -\frac{1}{q(1-q)} \frac{m(q)}{\ell(q)} dq \right) & \text{if } m(q_M) < 0 \\ 1 & \text{if } m(q_M) \geq 0. \end{cases} \end{aligned}$$

(b) Let  $q_M = 1$ . Then for each  $z \in \{m > 0\} \cap (0, 1) = (0, 1)$ , there exist a unique continuously differentiable decreasing function  $n_{(z)}^+ : [z, 1] \rightarrow [m(z), m(1)]$  with  $n_{(z)}^+(q) < \ell(q)$  for  $q \in [z, 1)$  that satisfies on  $[z, 1)$  the ODE

$$\frac{d}{dq} n_{(z)}^+(q) = O(n_{(z)}^+(q), q).$$

as well as the boundary conditions

$$n_{(z)}^+(z) = m(z) \quad \text{and} \quad n_{(z)}^+(1) = m(1) \quad \text{and} \quad \frac{d}{dq} n_{(z)}^+(1) = m'(1) = 0.$$

Moreover, the map  $\Sigma^+ : \{m > 0\} \cap (0, 1) \rightarrow (1, \infty)$  given by

$$\Sigma^+(z) := \exp \left( \int_z^1 -\frac{1}{q(1-q)} \frac{m(q) - n_{(z)}^+(q)}{\ell(q) - n_{(z)}^+(q)} dq \right)$$

is continuous and decreasing, and satisfies

$$\begin{aligned} \lim_{z \downarrow 0 = \inf(\{m > 0\} \cap (0, 1))} \Sigma^+(z) &= \infty \\ \lim_{z \uparrow 1 = \sup(\{m > 0\} \cap (0, 1))} \Sigma^+(z) &= 1. \end{aligned}$$

(c) Let  $q_M > 1$ . Then for each  $z \in \{m > 0\} \cap (0, q_M)$ , there exist a unique  $\zeta^+(z) \in \{m > 0\} \cap (q_M, 2q_M - 1)$  and a unique continuously differentiable decreasing function  $n_{(z)}^+ : [z, \zeta^+(z)] \rightarrow [m(z), m(\zeta^+(z))]$  with  $n_{(z)}^+(q) \neq \ell(q)$  for  $q \in [z, \zeta^+(z)] \setminus \{1\}$  that satisfies on  $[z, \zeta^+(z)] \setminus \{1\}$  the ODE

$$\frac{d}{dq} n_{(z)}^+(q) = O(n_{(z)}^+(q), q).$$

as well as the boundary conditions

$$n_{(z)}^+(z) = m(z) \quad \text{and} \quad n_{(z)}^+(\zeta^+(z)) = m(z)$$

and, if  $z < 1$ , the crossing condition

$$n_{(z)}^+(1) = m(1) \quad \text{and} \quad \frac{d}{dq} n_{(z)}^+(1) = m'(1).$$

Moreover, the map  $\Sigma^+ : \{m > 0\} \cap (0, q_M) \rightarrow (1, \infty)$  given by

$$\Sigma^+(z) := \exp \left( \int_z^{\zeta^+(z)} -\frac{1}{q(1-q)} \frac{m(q) - n_{(z)}^+(q)}{\ell(q) - n_{(z)}^+(q)} dq \right)$$

is continuous and decreasing, and satisfies

$$\lim_{z \downarrow 0 = \inf(\{m > 0\} \cap (0, q_M))} \Sigma^+(z) = \infty$$

$$\lim_{z \uparrow \sup(\{m > 0\} \cap (0, q_M))} \Sigma^+(z) = \begin{cases} \exp\left(\int_{q_-^m}^{q_+^m} -\frac{1}{q(1-q)} \frac{m(q)}{\ell(q)} dq\right) & \text{if } m(q_M) < 0 \\ 1 & \text{if } m(q_M) \geq 0. \end{cases}$$

(d) Let  $q_M < 0$ . Then for each  $z \in \{m > 0\} \cap (q_M, 0)$ , there exist a unique  $\zeta^-(z) \in \{m > 0\} \cap (-\infty, q_M)$  and a unique continuously differentiable decreasing function  $n_{(z)}^- : [\zeta^-(z), z] \rightarrow [m(\zeta^-(z)), m(z)]$  with  $n_{(z)}^-(q) \neq \ell(q)$  for  $q \in [\zeta^-(z), z] \setminus \{1\}$  that satisfies on  $[\zeta^-(z), z]$  the ODE

$$\frac{d}{dq} n_{(z)}^-(q) = O(n_{(z)}^-(q), q).$$

as well as the boundary conditions

$$n_{(z)}^-(z) = m(z) \quad \text{and} \quad n_{(z)}^-(\zeta^-(z)) = m(z).$$

Moreover, the map  $\Sigma^- : \{m > 0\} \cap (q_M, 0) \rightarrow (1, \infty)$  given by

$$\Sigma^-(z) := \exp\left(\int_{\zeta^-(z)}^z -\frac{1}{q(1-q)} \frac{m(q) - n_{(z)}^+(q)}{\ell(q) - n_{(z)}^+(q)} dq\right)$$

is continuous and decreasing, and satisfies

$$\lim_{z \uparrow 0 = \sup(\{m > 0\} \cap (q_M, 0))} \Sigma^-(z) = \infty$$

$$\lim_{z \downarrow \inf(\{m > 0\} \cap (q_M, 0))} \Sigma^-(z) = \begin{cases} \exp\left(\int_{q_-^m}^{q_+^m} -\frac{1}{q(1-q)} \frac{m(q)}{\ell(q)} dq\right) & \text{if } m(q_M) < 0 \\ 1 & \text{if } m(q_M) \geq 0. \end{cases}$$

**Proposition A.2.** Let  $R > 1$ . Assume that  $m(q_M) > 0$ .

(a) Let  $q_M \in (0, 1)$ . Then for each  $z \in \{m > 0\} \cap (0, q_M)$ , there exist a unique  $\zeta^+(z) \in \{m > 0\} \cap (q_M, 1)$  and a unique continuously differentiable increasing function  $n_{(z)}^+ : [z, \zeta^+(z)] \rightarrow [m(z), m(\zeta^+(z))]$  with  $n_{(z)}^+(q) \neq \ell(q)$  for  $q \in [z, \zeta^+(z)]$  that satisfies on  $[z, \zeta^+(z)]$  the ODE

$$\frac{d}{dq} n_{(z)}^+(q) = O(n_{(z)}^+(q), q)$$

with the boundary conditions

$$n_{(z)}^+(z) = m(z) \quad \text{and} \quad n_{(z)}^+(\zeta^+(z)) = m(z).$$

Moreover, the map  $\Sigma^+ : \{m > 0\} \cap (0, q_M) \rightarrow (1, \infty)$  given by

$$\Sigma^+(z) := \exp\left(\int_z^{\zeta^+(z)} -\frac{1}{q(1-q)} \frac{m(q) - n_{(z)}^+(q)}{\ell(q) - n_{(z)}^+(q)} dq\right)$$

is continuous and decreasing, and satisfies

$$\lim_{z \uparrow q_M = \sup(\{m > 0\} \cap (0, q_M))} \Sigma^+(z) = 1,$$

$$\lim_{z \downarrow \inf(\{m > 0\} \cap (0, q_M))} \Sigma^+(z) = \begin{cases} \exp\left(\int_{q_-^m}^{q_+^m} -\frac{1}{q(1-q)} \frac{m(q)}{\ell(q)} dq\right) & \text{if } m(0) < 0, \\ \infty & \text{if } m(0) \geq 0. \end{cases}$$

(b) Let  $q_M = 1$ . Then for each  $z \in \{m > 0\} \cap (0, q_M)$ , there exist a unique continuously differentiable increasing function  $n_{(z)}^+ : [z, 1] \rightarrow [m(z), m(1)]$  with  $n_{(z)}^+(q) \neq \ell(q)$  for  $q \in [z, 1)$  that satisfies on  $[z, 1)$  the ODE

$$\frac{d}{dq} n_{(z)}^+(q) = O(n_{(z)}^+(q), q)$$

with the boundary conditions

$$n_{(z)}^+(z) = m(z) \quad \text{and} \quad n_{(z)}^+(1) = m(1) \quad \text{and} \quad \frac{d}{dq} n_{(z)}^+(1) = m'(1) = 0.$$

Moreover, the map  $\Sigma^+ : \{m > 0\} \cap (0, 1) \rightarrow (1, \infty)$  given by

$$\Sigma^+(z) := \exp \left( \int_z^1 -\frac{1}{q(1-q)} \frac{m(q) - n_{(z)}^+(q)}{\ell(q) - n_{(z)}^+(q)} dq \right)$$

is continuous and decreasing, and satisfies

$$\begin{aligned} \lim_{z \uparrow 1 = q_M = \sup(\{m > 0\} \cap (0, q_M))} \Sigma^+(z) &= 1, \\ \lim_{z \downarrow \inf(\{m > 0\} \cap (0, q_M))} \Sigma^+(z) &= \begin{cases} \exp \left( \int_{q_m^m}^{q_m^m} -\frac{1}{q(1-q)} \frac{m(q)}{\ell(q)} dq \right) & \text{if } m(0) < 0, \\ \infty & \text{if } m(0) \geq 0. \end{cases} \end{aligned}$$

(c) Let  $q_M > 1$ . Then for each  $z \in \{m > 0\} \cap (0, q_M)$ , there exist a unique  $\zeta^+(z) \in \{m > 0\} \cap (q_M, 2q_M - 1)$  and a unique continuously differentiable increasing function  $n_{(z)}^+ : [z, \zeta^+(z)] \rightarrow [m(z), m(\zeta^+(z))]$  with  $n_{(z)}^+(q) \neq \ell(q)$  for  $q \in [z, \zeta^+(z)] \setminus \{1\}$  that satisfies on  $[z, \zeta^+(z)] \setminus \{1\}$  the ODE

$$\frac{d}{dq} n_{(z)}^+(q) = O(n_{(z)}^+(q), q)$$

with the boundary conditions

$$n_{(z)}^+(z) = m(z) \quad \text{and} \quad n_{(z)}^+(\zeta^+(z)) = m(z)$$

and, if  $z < 1$ , the crossing condition at 1,

$$n_{(z)}^+(1) = m(1) \quad \text{and} \quad \frac{d}{dq} n_{(z)}^+(1) = m'(1).$$

Moreover, the map  $\Sigma^+ : \{m > 0\} \cap (0, q_M) \rightarrow (1, \infty)$  given by

$$\Sigma^+(z) := \exp \left( \int_z^{\zeta^+(z)} -\frac{1}{q(1-q)} \frac{m(q) - n_{(z)}^+(q)}{\ell(q) - n_{(z)}^+(q)} dq \right)$$

is continuous and decreasing, and satisfies

$$\begin{aligned} \lim_{z \uparrow q_M = \sup(\{m > 0\} \cap (0, q_M))} \Sigma^+(z) &= 1, \\ \lim_{z \downarrow \inf(\{m > 0\} \cap (0, q_M))} \Sigma^+(z) &= \begin{cases} \exp \left( \int_{q_m^m}^{q_m^m} -\frac{1}{q(1-q)} \frac{m(q)}{\ell(q)} dq \right) & \text{if } m(0) < 0, \\ \infty & \text{if } m(0) \geq 0. \end{cases} \end{aligned}$$

(d) Let  $q_M < 0$ . Then for each  $z \in \{m > 0\} \cap (q_M, 0)$ , there exist a unique  $\zeta^-(z) \in \{m > 0\} \cap (-\infty, q_M)$  and a unique, continuously differentiable, increasing function  $n_{(z)}^- : [\zeta^-(z), z] \rightarrow [m(\zeta^-(z)), m(z)]$  with  $n_{(z)}^-(q) \neq \ell(q)$  for  $q \in [\zeta^-(z), z] \setminus \{1\}$  that satisfies on  $[\zeta^-(z), z]$  the ODE

$$\frac{d}{dq} n_{(z)}^-(q) = O(n_{(z)}^-(q), q)$$

as well as the boundary conditions

$$n_{(z)}^-(z) = m(z) \quad \text{and} \quad n_{(z)}^-(\zeta^-(z)) = m(z).$$

Moreover, the map  $\Sigma^- : \{m > 0\} \cap (q_M, 0) \rightarrow (1, \infty)$  given by

$$\Sigma^-(z) := \exp \left( \int_{\zeta^-(z)}^z -\frac{1}{q(1-q)} \frac{m(q) - n_{(z)}^+(q)}{\ell(q) - n_{(z)}^+(q)} dq \right)$$

is continuous and decreasing, and satisfies

$$\begin{aligned} \lim_{z \downarrow q_M = \inf(\{m > 0\} \cap (q_M, 0))} \Sigma^-(z) &= 1, \\ \lim_{z \uparrow \sup(\{m > 0\} \cap (q_M, 0))} \Sigma^-(z) &= \begin{cases} \exp \left( \int_{q_-^m}^{q_+^m} -\frac{1}{q(1-q)} \frac{m(q)}{\ell(q)} dq \right) & \text{if } m(0) < 0, \\ \infty & \text{if } m(0) \geq 0. \end{cases} \end{aligned}$$

*Proof of Proposition A.1 and A.2.* Complete proofs are omitted but a structurally identical family of first order ODEs is studied in [22], Section 3. See also Section 3.5 of [38] and Section 5 of [21] for the analysis of a similar problem in the context of optimal investment and consumption under proportional transaction costs with multiple assets.  $\square$

*Proof of Proposition 5.3.* We only establish (a) – (f) for the case  $q_M \in (0, 1)$ , the argument for the other three cases is similar.

First, we show existence of  $q_*(\xi)$ ,  $q^*(\xi)$  and  $n_\xi$ . For  $z \in \{m > 0\} \cap (0, q_M)$ , define  $\zeta^+(z) \in \{m > 0\} \cap (q_M, 1)$  and the function  $n_{(z)}^+ : [z, \zeta^+(z)] \rightarrow [m(z), m(\zeta^+(z))]$  as in Propositions A.1(a) or A.2(a), i.e.  $n_{(z)}^+$  solves  $n' = O(n, q)$  subject to  $n_{(z)}^+(z) = m(z)$ , and then crosses  $m$  at  $\zeta^+(z)$ . Further, define the map  $\Sigma^+ : \{m > 0\} \cap (0, q_M) \rightarrow (1, \infty)$  by

$$\Sigma^+(z) := \exp \left( \int_z^{\zeta^+(z)} -\frac{1}{q(1-q)} \frac{m(q) - n_{(z)}^+(q)}{\ell(q) - n_{(z)}^+(q)} dq \right)$$

Since  $\xi$  is in the well-posedness range of transaction costs and the function  $\Sigma^+$  is increasing and continuous, it follows from Propositions A.1(a) or A.2(a) that there exists a unique  $z \in \{m > 0\} \cap (0, q_M)$  such that  $\Sigma^+(z) = \xi$ . Set  $q_*(\xi) := z$ ,  $q^*(\xi) := \zeta^+(z)$  and  $n_\xi := n_{(z)}^+$ . Then it follows from Propositions A.1(a) or A.2(a) that  $n_\xi$  satisfies properties (a) – (f). Uniqueness follows since  $\Sigma^+$  is strictly increasing.

Finally, we establish (5.5) and (5.6). We only provide the argument for (5.5) and for the case  $q_M > 0$ : the argument for (5.6) or for  $q_M < 0$  is similar.

Suppose  $0 < q^* \leq 1$ . Then  $(\xi - 1)(q^*)^2 \leq (\xi - 1)q^*$  which can be rewritten as  $q^*(1 + (\xi - 1)q^*) \leq \xi q^*$ . Then  $q_* < q_M \leq q^* \leq \frac{\xi q^*}{(1 + (\xi - 1)q^*)}$ . Similarly if  $q_* \leq 1 < q^*$  then  $1 + (\xi - 1)q^* < q^* + (\xi - 1)q^* = \xi q^*$  so that  $q_* \leq 1 < \frac{\xi q^*}{(1 + (\xi - 1)q^*)}$ . Finally, the non-trivial case is when  $1 < q_* < q_M < q^*$ . Then on  $(q_*(\xi), q^*(\xi))$ ,  $n_\xi \geq m \geq \ell$  if  $R < 1$  and  $n_\xi \leq m \leq \ell$  if  $R > 1$ . In either case, this implies that  $0 < \frac{m(q) - n_\xi(q)}{\ell(q) - n_\xi(q)} < 1$  on  $(q_*(\xi), q^*(\xi))$ . This together with (5.4) gives

$$\ln(\xi) < \int_{q_*(\xi)}^{q^*(\xi)} \frac{1}{q(q-1)} dq = \log \frac{q^*(\xi) - 1}{q^*(\xi)} - \log \frac{q_*(\xi) - 1}{q_*(\xi)}$$

Rearranging terms yields the first inequality in (5.5). The second inequality is trivial.  $\square$

The following corollary is useful in the study of the ill-posed case.

**Corollary A.3.** *Suppose  $\bar{\xi} \in (1, \infty)$ .*

*Suppose  $R < 1$ . Then  $\lim_{\xi \downarrow \bar{\xi}} \inf_{q \in [q_*(\xi), q^*(\xi)]} n_\xi(q) = 0$ .*

*Suppose  $R < 1$ . Then  $\lim_{\xi \downarrow \bar{\xi}} \inf_{q \in [q_*(\xi), q^*(\xi)]} n_\xi(q) = 0$ .*

*Proof.* This follows from the fact that  $\Sigma^+$  is increasing and continuous.  $\square$

*Proof of Corollary 5.4.* Uniqueness of  $\kappa_{\gamma^\uparrow, \gamma^\downarrow}$  follows from standard ODE theory, using the (double) boundary condition in (5.8).

To establish existence and argue the remaining claims, define  $\kappa_{\gamma^\uparrow, \gamma^\downarrow} : [q_*(\xi), q^*(\xi)] \rightarrow (0, \infty)$  by

$$\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q) := \gamma^\uparrow \exp \left( \frac{S}{1-S} \int_{q_*(\xi)}^q \frac{n'_\xi(v)}{vn_\xi(v)} dv \right) = \frac{1}{\gamma^\downarrow} \exp \left( -\frac{S}{1-S} \int_q^{q^*(\xi)} \frac{n'_\xi(v)}{vn_\xi(v)} dv \right), \quad (\text{A.1})$$

where the equality in (A.1) follows from the integral constraint (5.4) for  $n_\xi$  and the fact that  $\gamma^\uparrow \gamma^\downarrow = \xi$ . Then  $\kappa_{\gamma^\uparrow, \gamma^\downarrow}$  is well-defined (since  $0 \notin [q_*(\xi), q^*(\xi)]$ ) and satisfies (5.8). Moreover,  $\kappa_{\gamma^\uparrow, \gamma^\downarrow}$  is continuously differentiable with derivative

$$\kappa'_{\gamma^\uparrow, \gamma^\downarrow}(q) = \frac{S}{1-S} \frac{n'_\xi(q)}{qn_\xi(q)} \kappa_{\gamma^\uparrow, \gamma^\downarrow}(q). \quad (\text{A.2})$$

Plugging in  $q_*(\xi)$  and  $q^*(\xi)$  and using (5.3) gives (5.9). Moreover, using the ODE (5.2) for  $n$ , we obtain that  $\kappa_{\gamma^\uparrow, \gamma^\downarrow}$  satisfies the ODE (5.7) on  $[q_*(\xi), q^*(\xi)] \setminus \{1\}$ . Finally, using that  $(1-R)$  and  $(1-S)$  have the same sign and that  $(1-R)n_\xi$  is decreasing by Proposition 5.3(b), (A.2) directly yields (a).

The additional claim is straightforward, given (A.1).  $\square$

## A.2 No-transaction region

We proceed to list some properties of the Möbius transformations  $\tau_c(q)$  for  $c \in (0, \infty)$  and  $q \in \mathbb{R} \cup \{\bar{\infty}\}$ . First, for fixed  $c$ ,  $\tau_c$  is continuously differentiable and strictly increasing on  $(\bar{\infty}, \frac{1}{1-c} = \tau_{\frac{1}{c}}(\bar{\infty}))$  and  $(\frac{1}{1-c} = \tau_{\frac{1}{c}}(\bar{\infty}), \bar{\infty})$ . Moreover, for fixed  $q$ ,  $\tau_c$  is also continuously differentiable with respect to the parameter  $c$ . For future reference, we note that for  $q \in \mathbb{R} \setminus \{\frac{1}{1-c}, 0, 1\}$ ,

$$\tau'_c(q) = \frac{\tau_c(q)(1 - \tau_c(q))}{q(1 - q)}, \quad (\text{A.3})$$

$$\frac{\partial}{\partial c} \tau_c(q) = \frac{1}{c} \tau_c(q)(1 - \tau_c(q)). \quad (\text{A.4})$$

*Proof of Lemma 5.8.* If, for example,  $q_M \in (0, 1)$  then the inequalities follow easily since, by the properties of  $\tau_c$  on  $(0, 1)$ ,  $p_*(\gamma^\uparrow, \gamma^\downarrow) = \tau_{\frac{1}{\gamma^\uparrow}}(q_*(\xi)) \leq q_*(\xi) < q_M < q^*(\xi) \leq \tau_{\gamma^\downarrow}(q^*(\xi)) = p_*(\gamma^\uparrow, \gamma^\downarrow)$ . The result for  $q_M = 1$  is very similar.

For other values of  $q_M$  the result is a little more subtle. We consider the case when  $q_M > 1$ , the case  $q_M < 0$  is similar and omitted. If  $q_* \leq 1$  we have  $p_*(\gamma^\uparrow, \gamma^\downarrow) \leq 1 < \tau_{\gamma^\downarrow}(q^*(\xi)) = p^*(\gamma^\uparrow, \gamma^\downarrow)$ . So, suppose  $1 < q_* < q_M$ . By the final result of Proposition 5.3 we have

$$q_*(\xi) < \frac{\xi q^*(\xi)}{1 + (\xi - 1)q^*(\xi)} = \tau_\xi(q^*(\xi)).$$

Then  $p_*(\gamma^\uparrow, \gamma^\downarrow) = \tau_{\frac{1}{\gamma^\uparrow}}(q^*(\xi)) < \tau_{\frac{1}{\gamma^\uparrow}} \circ \tau_\xi(q^*(\xi)) = \tau_{\gamma^\downarrow}(q^*(\xi)) = p^*(\gamma^\uparrow, \gamma^\downarrow)$ .

For the final inequality in the case  $q_M > 1$  note that in the case  $R < 1$ , since  $n_\xi$  is decreasing on  $(q_*(\xi) \vee 1, q^*(\xi))$  and  $m$  is a quadratic with minimum at  $q_M$ , we must have  $q^*(\xi) - q_M < q_M - (1 \vee q_*(\xi))$ . In particular,  $q^*(\xi) < 2q_M - 1$  and  $p^*(\gamma^\uparrow, \gamma^\downarrow) = \tau_{\gamma^\downarrow}(q^*(\xi)) < \tau_{\gamma^\downarrow}(2q_M - 1)$ . The argument for  $R > 1$  is very similar.  $\square$

*Proof of Proposition 5.10.* Uniqueness of  $q_{\gamma^\uparrow, \gamma^\downarrow}$  follows from standard ODE theory, using the (double) boundary conditions in (b).

To establish existence of  $q_{\gamma^\uparrow, \gamma^\downarrow}$ , we first define its inverse function. Motivated by (4.5), we let  $p_{\gamma^\uparrow, \gamma^\downarrow} : [q_*(\xi), q^*(\xi)] \rightarrow \mathbb{R}$  be given by

$$p_{\gamma^\uparrow, \gamma^\downarrow}(q) := \frac{q}{\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q) + (1 - \kappa_{\gamma^\uparrow, \gamma^\downarrow}(q))q} = \tau_{\frac{1}{\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q)}}(q).$$

The boundary conditions (5.8) of  $\kappa_{\gamma^\uparrow, \gamma^\downarrow}$  give

$$p_{\gamma^\uparrow, \gamma^\downarrow}(q_*(\xi)) = \tau_{\frac{1}{\gamma^\uparrow}}(q_*(\xi)) = p_*(\gamma^\uparrow, \gamma^\downarrow) \quad \text{and} \quad p_{\gamma^\uparrow, \gamma^\downarrow}(q^*(\xi)) = \tau_{\gamma^\downarrow}(q^*(\xi)) = p^*(\gamma^\uparrow, \gamma^\downarrow),$$

and the fact that  $\tau_c(1) = 1$  for all  $c \in (0, \infty)$ , gives  $p_{\gamma^\uparrow, \gamma^\downarrow}(1) = 1$  if  $1 \in [q_*(\xi), q^*(\xi)]$ . Moreover,  $p_{\gamma^\uparrow, \gamma^\downarrow}$  is continuously differentiable because  $\kappa_{\gamma^\uparrow, \gamma^\downarrow}$  is. Using (A.3) and (A.4), the ODE (5.7) for  $\kappa$ , and the simple identity

$$\frac{p(1-p)}{q(1-q)} = \frac{m(p) - \ell(p)}{m(q) - \ell(q)}, \quad p \in \mathbb{R}, q \in \mathbb{R} \setminus \{0, 1\},$$

we obtain for  $q \in [q_*(\xi), q^*(\xi)] \setminus \{1\}$ ,

$$\begin{aligned} p'_{\gamma^\uparrow, \gamma^\downarrow}(q) &= \frac{p_{\gamma^\uparrow, \gamma^\downarrow}(q)(1 - p_{\gamma^\uparrow, \gamma^\downarrow}(q))}{q(1-q)} - \frac{p_{\gamma^\uparrow, \gamma^\downarrow}(q)(1 - p_{\gamma^\uparrow, \gamma^\downarrow}(q))}{q(1-q)} \frac{m(q) - n_\xi(q)}{\ell(q) - n_\xi(q)} \\ &= \frac{p_{\gamma^\uparrow, \gamma^\downarrow}(q)(1 - p_{\gamma^\uparrow, \gamma^\downarrow}(q))}{q(1-q)} \frac{\ell(q) - m(q)}{\ell(q) - n_\xi(q)} \\ &= \frac{\ell(p_{\gamma^\uparrow, \gamma^\downarrow}(q)) - m(p_{\gamma^\uparrow, \gamma^\downarrow}(q))}{\ell(q) - n_\xi(q)}. \end{aligned} \tag{A.5}$$

Looking at (A.5) shows that  $p'_{\gamma^\uparrow, \gamma^\downarrow}$  is positive. Indeed, the first factor in (A.5) coincides by (A.3) with  $\tau'_c(q)$  for  $c = \frac{1}{\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q)}$  and is therefore positive; the second factor is also positive because  $n_\xi$  and  $\ell$  both lie on the same side of  $m$  on  $(q_*(\xi), q^*(\xi)) \setminus \{1\}$ . Thus,  $p_{\gamma^\uparrow, \gamma^\downarrow}$  is strictly increasing and maps  $[q_*(\xi), q^*(\xi)]$  onto  $[p_*(\gamma^\uparrow, \gamma^\downarrow), p^*(\gamma^\uparrow, \gamma^\downarrow)]$ . If we now define the function  $q_{\gamma^\uparrow, \gamma^\downarrow} : [q_*(\xi), q^*(\xi)] \rightarrow [q_*(\xi), q^*(\xi)]$  by  $q_{\gamma^\uparrow, \gamma^\downarrow} := p_{\gamma^\uparrow, \gamma^\downarrow}^{-1}$ , then  $q_{\gamma^\uparrow, \gamma^\downarrow}$  satisfies (a) to (c). (Note that since  $\tau_c(1) = 1$  for all  $c$  we immediately get  $p_{\gamma^\uparrow, \gamma^\downarrow} = 1$  whence  $q_{\gamma^\uparrow, \gamma^\downarrow}(1) = 1$  also.)

Finally, if  $\tilde{\gamma}^\downarrow, \tilde{\gamma}^\uparrow \in [1, \infty)$  are such that  $\tilde{\gamma}^\downarrow \tilde{\gamma}^\uparrow = \xi$ , then  $\frac{\tilde{\gamma}^\downarrow}{\tilde{\gamma}^\uparrow} = \frac{\gamma^\downarrow}{\gamma^\uparrow}$  and it suffices to show that the function  $q_{\gamma^\uparrow, \gamma^\downarrow} \circ \tau_{\frac{\tilde{\gamma}^\downarrow}{\tilde{\gamma}^\uparrow}}$  satisfies properties (a) - (c). For (a), this follows from the ODE (5.12) for  $q_{\gamma^\uparrow, \gamma^\downarrow}$  and (A.3)  $\square$

### A.3 The candidate value function and candidate optimal strategy

We first collect some properties of the extended functions  $\bar{n}_\xi$  and  $\bar{\kappa}_{\gamma^\uparrow, \gamma^\downarrow}$  and  $\bar{q}_{\gamma^\uparrow, \gamma^\downarrow}$ . The following result follows directly from Proposition 5.3, Corollary 5.4 and Proposition 5.10 together with Definition 5.11.

**Proposition A.4.** *Let  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$  with  $\xi := \gamma^\uparrow \gamma^\downarrow > 1$  be such that  $\xi$  lies in the well-posedness range of transaction costs.*

(a) *The functions  $\bar{n}_\xi$  and  $\bar{\kappa}_{\gamma^\uparrow, \gamma^\downarrow}$  are continuously differentiable and satisfy on  $(-\infty, \infty)$  the relationship*

$$q \frac{\bar{\kappa}'_{\gamma^\uparrow, \gamma^\downarrow}(q)}{\bar{\kappa}_{\gamma^\uparrow, \gamma^\downarrow}(q)} = \frac{S}{1-S} \frac{\bar{n}'_\xi(q)}{\bar{n}_\xi(q)}.$$

(b) *The function  $\bar{q}_{\gamma^\uparrow, \gamma^\downarrow}$  is continuously differentiable and on  $[p_*(\gamma^\uparrow, \gamma^\downarrow), p^*(\gamma^\uparrow, \gamma^\downarrow)]$  satisfies,*

$$p(1-p) \bar{q}'_{\gamma^\uparrow, \gamma^\downarrow}(p) = \frac{S}{1-S} \frac{2}{\sigma^2} (\ell(\bar{q}_{\gamma^\uparrow, \gamma^\downarrow})(p) - n_\xi(\bar{q}_{\gamma^\uparrow, \gamma^\downarrow}(p)))$$

and off  $[p_*(\gamma^\uparrow, \gamma^\downarrow), p^*(\gamma^\uparrow, \gamma^\downarrow)]$  we have  $p(1-p) \bar{q}'_{\gamma^\uparrow, \gamma^\downarrow}(p) = \frac{S}{1-S} \frac{2}{\sigma^2} (\ell(\bar{q}_{\gamma^\uparrow, \gamma^\downarrow})(p) - m(\bar{q}_{\gamma^\uparrow, \gamma^\downarrow}(p)))$

(c) Let  $(x, y, \phi) \in \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ$ . Then

$$x + \phi y \bar{\kappa}_{\gamma^\uparrow, \gamma^\downarrow} \left( \bar{q}_{\gamma^\uparrow, \gamma^\downarrow} \left( \frac{\phi y}{x + y \phi} \right) \right) > 0,$$

and

$$\frac{\phi y \bar{\kappa}_{\gamma^\uparrow, \gamma^\downarrow} \left( \bar{q}_{\gamma^\uparrow, \gamma^\downarrow} \left( \frac{\phi y}{x + y \phi} \right) \right)}{x + \phi y \bar{\kappa}_{\gamma^\uparrow, \gamma^\downarrow} \left( \bar{q}_{\gamma^\uparrow, \gamma^\downarrow} \left( \frac{\phi y}{x + y \phi} \right) \right)} = \bar{q}_{\gamma^\uparrow, \gamma^\downarrow} \left( \frac{\phi y}{x + y \phi} \right).$$

We proceed to study properties of the candidate value function.

**Proposition A.5.** Let  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$  with  $\xi := \gamma^\uparrow \gamma^\downarrow > 1$  be such that  $\xi$  lies in the well-posedness range of transaction costs. Then  $\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}$  is continuously differentiable with respect to  $t, x, \phi$  and twice continuously differentiable with respect to  $y$ . Moreover, fix  $(t, x, y, \phi) \in [0, \infty) \times \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ$ . Then

(a)  $\hat{V}_x^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi) > 0$  and  $\hat{V}_\phi^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi) > 0$ , we have the identity

$$\frac{\hat{V}_\phi^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi)}{y \hat{V}_x^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi)} = \bar{\kappa}_{\gamma^\uparrow, \gamma^\downarrow} \left( \bar{q}_{\gamma^\uparrow, \gamma^\downarrow} \left( \frac{\phi y}{x + \phi y} \right) \right), \quad (\text{A.6})$$

and (A.6) equals  $\gamma^\uparrow$  if  $\frac{\phi y}{x + \phi y} \in (\tau_{\frac{1}{\gamma^\uparrow}(\infty)} = \frac{-1}{\gamma^\uparrow - 1}, p_*(\gamma^\uparrow, \gamma^\downarrow)]$  and  $\frac{1}{\gamma^\downarrow}$  if  $\frac{\phi y}{x + \phi y} \in [p^*(\gamma^\uparrow, \gamma^\downarrow), \tau_{\gamma^\downarrow}(\infty) = \frac{\gamma^\downarrow}{\gamma^\downarrow - 1})$ , respectively.

(b)

$$\begin{aligned} & g_{EZ}(t, c, \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi)) - c \hat{V}_x^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi) \\ & \leq \theta S \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi) \bar{n}_\xi \left( \bar{q}_{\gamma^\uparrow, \gamma^\downarrow} \left( \frac{\phi y}{x + \phi y} \right) \right) \\ & \leq -\hat{V}_t^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi) - rx \hat{V}_x^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi) - \mu y \hat{V}_y^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi) - \frac{1}{2} \sigma^2 y^2 \hat{V}_{yy}^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi), \end{aligned}$$

where the inequality in (A.8) is an equality if  $c = (x + \phi y \bar{\kappa}_{\gamma^\uparrow, \gamma^\downarrow}(\bar{q}_{\gamma^\uparrow, \gamma^\downarrow}(\frac{\phi y}{x + \phi y}))) \bar{n}_\xi(\bar{q}_{\gamma^\uparrow, \gamma^\downarrow}(\frac{\phi y}{x + \phi y}))$ , and the inequality in (A.9) is an equality if  $\frac{\phi y}{x + \phi y} \in [p_*(\gamma^\uparrow, \gamma^\downarrow), p^*(\gamma^\uparrow, \gamma^\downarrow)]$ .

*Proof.* First, it follows from the continuous differentiability of  $\bar{n}_\xi$ ,  $\bar{\kappa}_{\gamma^\uparrow, \gamma^\downarrow}$  and  $\bar{q}_{\gamma^\uparrow, \gamma^\downarrow}$  that  $\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}$  is continuously differentiable with respect to all its arguments. Moreover, dropping the references to  $\gamma^\uparrow$ ,  $\gamma^\downarrow$  and  $\xi$  and the arguments for simplicity and using Proposition A.4(a), we obtain

$$\hat{V}_t = -\delta \theta \hat{V}, \quad (\text{A.10})$$

$$\hat{V}_x = (1 - R) \hat{V} \frac{1}{x + \phi y \bar{\kappa}(\bar{q})} \quad (\text{A.11})$$

$$\hat{V}_\phi = (1 - R) \hat{V} \frac{y \bar{\kappa}(\bar{q})}{x + \phi y \bar{\kappa}(\bar{q})}, \quad (\text{A.12})$$

$$\hat{V}_y = (1 - R) \hat{V} \frac{\phi \bar{\kappa}(\bar{q})}{x + \phi y \bar{\kappa}(\bar{q})} = (1 - R) \hat{V} \frac{\bar{q}}{y} \quad (\text{A.13})$$

Next, it follows from the continuous differentiability of  $\bar{n}_\xi$  and  $\bar{\kappa}_{\gamma^\uparrow, \gamma^\downarrow}$  and  $\bar{q}_{\gamma^\uparrow, \gamma^\downarrow}$  and (A.13) that  $\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}$  is twice continuously differentiable with respect to  $y$ . Again dropping the references to  $\gamma^\uparrow$ ,  $\gamma^\downarrow$  and  $\xi$  and the

arguments, and using that  $\hat{V}_q = 0$  we obtain

$$\begin{aligned} y^2 \hat{V}_{yy} &= \hat{V}(1-R)\bar{q}((1-R)\bar{q}-1) + (1-R)\hat{V}y \frac{\partial p}{\partial y} \frac{\partial \bar{q}}{\partial p} \\ &= -R(1-R)\hat{V}\bar{q}^2 + \hat{V}(1-R) \left( p(1-p) \frac{\partial \bar{q}}{\partial p} - \bar{q}(1-\bar{q}) \right) \end{aligned}$$

In summary, we have

$$y^2 \hat{V}_{yy} = \begin{cases} -R(1-R)\hat{V}\bar{q}^2 & \text{if } \frac{\phi y}{x+\phi y} \in (\tau_{\frac{1}{\gamma^\uparrow}}(\bar{\infty}), p_*(\gamma^\uparrow, \gamma^\downarrow)) \cup (p^*(\gamma^\uparrow, \gamma^\downarrow), \tau_{\gamma^\downarrow}(\bar{\infty})), \\ -R(1-R)\hat{V}\bar{q}^2 + \frac{2\theta S}{\sigma^2} \hat{V}(m(\bar{q}) - \bar{n}(\bar{q})) & \text{if } \frac{\phi y}{x+\phi y} \in [p_*(\gamma^\uparrow, \gamma^\downarrow), p^*(\gamma^\uparrow, \gamma^\downarrow)]. \end{cases} \quad (\text{A.14})$$

(a) This follows directly from (A.11) and (A.12).

(b) Using (A.11) and maximising  $g_{EZ}(t, c, \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi)) - cV_x^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi)$  over  $c$ , we obtain that the maximum is attained at

$$\hat{c} = \left( x + \phi y \bar{\kappa}_{\gamma^\uparrow, \gamma^\downarrow} \left( \bar{q}_{\gamma^\uparrow, \gamma^\downarrow} \left( \frac{\phi y}{x + \phi y} \right) \right) \right) \bar{n}_\xi \left( \bar{q}_{\gamma^\uparrow, \gamma^\downarrow} \left( \frac{\phi y}{x + \phi y} \right) \right).$$

and that the corresponding maximal value is given by

$$g_{EZ}(t, \hat{c}, \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi)) - \hat{c}V_x^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi) = \theta S \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi) \bar{n}_\xi \left( \bar{q}_{\gamma^\uparrow, \gamma^\downarrow} \left( \frac{\phi y}{x + \phi y} \right) \right).$$

Thus, we have (A.8)

Next, using (A.10) – (A.12) with Proposition A.4(c) and (A.14) together with the definition of  $m$  in (3.7), we obtain

$$-\hat{V}_t - rx\hat{V}_x - \mu y\hat{V}_y - \frac{1}{2}\sigma^2 y^2 \hat{V}_{yy} = \begin{cases} \theta S \hat{V} m(\bar{q}) & \text{if } \frac{\phi y}{x+\phi y} \in (\frac{1}{\gamma^\uparrow}(\bar{\infty}), p_*(\gamma^\uparrow, \gamma^\downarrow)) \cup (p^*(\gamma^\uparrow, \gamma^\downarrow), \tau_{\gamma^\downarrow}(\bar{\infty})), \\ \theta S \hat{V} \bar{n}(\bar{q}) & \text{if } \frac{\phi y}{x+\phi y} \in [p_*(\gamma^\uparrow, \gamma^\downarrow), p^*(\gamma^\uparrow, \gamma^\downarrow)]. \end{cases}$$

Now, in case that  $\frac{\phi y}{x+\phi y} \in [p_*(\gamma^\uparrow, \gamma^\downarrow), p^*(\gamma^\uparrow, \gamma^\downarrow)]$ , we obtain an equality in (A.9). By contrast, in case that  $\frac{\phi y}{x+\phi y} \in (\frac{1}{\gamma^\uparrow}(\bar{\infty}), p_*(\gamma^\uparrow, \gamma^\downarrow)) \cup (p^*(\gamma^\uparrow, \gamma^\downarrow), \tau_{\gamma^\downarrow}(\bar{\infty}))$ , we obtain an inequality in (A.9) noting that  $\theta S(1-R)\hat{V}$  is positive and  $\frac{m}{1-R}$  is a convex quadratic function so that  $\frac{m(\bar{q})}{1-R} \geq \frac{m(q_*(\xi))}{1-R} = \frac{\bar{n}(q_*(\xi))}{1-R}$  if  $\bar{q} \leq q_*(\xi) \leq q_M$  and  $\frac{m(\bar{q})}{1-R} \geq \frac{m(q^*(\xi))}{1-R} = \frac{\bar{n}(q^*(\xi))}{1-R} = \frac{n(\bar{q})}{1-R}$  if  $q \geq q^*(\xi) \geq q_M$ , where we have used Proposition 5.3 for the inequalities  $q_*(\xi) \leq q_M \leq q^*(\xi)$ .  $\square$

**Corollary A.6.** *Let  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$  with  $\xi := \gamma^\uparrow \gamma^\downarrow > 1$  be such that  $\xi$  is in the well-posedness range of transaction costs.*

(a) *Fix  $(x, y, \phi) \in \mathcal{S}_{(\gamma^\uparrow, \gamma^\downarrow)}^\circ$  and  $(x', y, \phi') \in \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}$ . Then  $(x + x', y, \phi + \phi') \in \mathcal{S}_{(\gamma^\uparrow, \gamma^\downarrow)}^\circ$  and*

$$\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(x + x', y, \phi + \phi') \geq \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(x, y, \phi).$$

(b) *If  $(x', y, \phi') \in \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ$  is such that  $x' := x - \gamma^\uparrow y(\phi' - \phi)^+ + \frac{y}{\gamma^\uparrow}(\phi' - \phi)^-$  (corresponding to an initial bulk trade), then  $\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi) \geq \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, x', y, \phi')$ . Moreover, the inequality is an equality if either  $\frac{\phi y}{x+\phi y} < p_*(\gamma^\uparrow, \gamma^\downarrow)$  and  $\frac{\phi' y}{x+\phi' y} = p_*(\gamma^\uparrow, \gamma^\downarrow)$  or  $\frac{\phi y}{x+\phi y} > p_*(\gamma^\uparrow, \gamma^\downarrow)$  and  $\frac{\phi' y}{x+\phi' y} = p_*(\gamma^\uparrow, \gamma^\downarrow)$ .*

*Proof.* (a) First, it follows from the definition of  $\mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}$  and  $\mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ$  that  $x + \varepsilon x', y, \phi + \varepsilon \phi' \in \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}^\circ$  for all  $\varepsilon \in [0, 1]$ . Hence, it suffices to show that  $x' \hat{V}_x^{\gamma^\uparrow, \gamma^\downarrow}(x + \varepsilon x', y, \phi + \varepsilon \phi') + \phi' \hat{V}_\phi^{\gamma^\uparrow, \gamma^\downarrow}(x + \varepsilon x', y, \phi + \varepsilon \phi') \geq 0$  for all  $\varepsilon \in [0, 1]$ . So fix  $\varepsilon \in [0, 1]$ . Then by Proposition A.5(a) and the definition of  $\mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}$ ,

$$\begin{aligned} & x' \hat{V}_x^{\gamma^\uparrow, \gamma^\downarrow}(x + \varepsilon x', y, \phi + \varepsilon \phi') + \phi' \hat{V}_\phi^{\gamma^\uparrow, \gamma^\downarrow}(x + \varepsilon x', y, \phi + \varepsilon \phi') \\ &= V_x^{\gamma^\uparrow, \gamma^\downarrow}(x + \varepsilon x', s, \phi + \varepsilon \phi') \left( x' + \phi' y \bar{\kappa}_{\gamma^\uparrow, \gamma^\downarrow} \left( \bar{q}_{\gamma^\uparrow, \gamma^\downarrow} \left( \frac{(\phi + \varepsilon \phi') y}{x + \varepsilon x' + (\phi + \varepsilon \phi') y} \right) \right) \right) \geq 0. \end{aligned}$$

where we use that if  $\phi' > 0$  then  $x' + \phi' y \bar{\kappa}(\bar{q}) \geq x' + \frac{\phi' y}{\gamma^\uparrow} \geq 0$  since  $(x', y, \phi') \in \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}$  and if  $\phi' < 0$  then  $x' + \phi' y \bar{\kappa}(\bar{q}) \geq x' + \phi' y \gamma^\downarrow \geq 0$  again since  $(x', y, \phi') \in \mathcal{S}_{\gamma^\uparrow, \gamma^\downarrow}$ .

(b) We only consider the case that  $\phi' > \phi$ , the case  $\phi' < \phi$  is similar. Then  $x' = x - \gamma^\uparrow(\phi' - \phi)y$ . Define the function  $\varepsilon : [0, 1] \rightarrow (0, \infty)$  by  $f(\varepsilon) = V^{\gamma^\uparrow, \gamma^\downarrow}(t, (1 - \varepsilon)x + \varepsilon x', y, (1 - \varepsilon)\phi + \varepsilon \phi')$ . Then  $f(0) = \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, x, y, \phi)$  and  $f(1) = \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, x', y, \phi')$ . It suffices to show that  $f'(\varepsilon) \leq 0$  and  $f'(\varepsilon) = 0$  if  $\frac{\phi y}{x + \phi y} < \frac{\phi' y}{x + \phi' y} = p_*(\gamma^\uparrow, \gamma^\downarrow)$ . Dropping the arguments for simplicity, the claim follows from part (a) via

$$\begin{aligned} f'(\varepsilon) &= \hat{V}_x^{\gamma^\uparrow, \gamma^\downarrow}(x' - x) + \hat{V}_\phi^{\gamma^\uparrow, \gamma^\downarrow}(\phi' - \phi) = -\hat{V}_x^{\gamma^\uparrow, \gamma^\downarrow} y \gamma^\uparrow(\phi' - \phi) + \hat{V}_\phi^{\gamma^\uparrow, \gamma^\downarrow}(\phi' - \phi) \\ &= \hat{V}_x^{\gamma^\uparrow, \gamma^\downarrow} y(\phi' - \phi) \left( \bar{\kappa}_{\gamma^\uparrow, \gamma^\downarrow} \left( \bar{q}_{\gamma^\uparrow, \gamma^\downarrow} \left( \frac{((1 - \varepsilon)\phi + \varepsilon \phi') y}{(1 - \varepsilon)x + \varepsilon x' + ((1 - \varepsilon)\phi + \varepsilon \phi') y} \right) \right) - \gamma^\uparrow \right). \end{aligned}$$

□

*Proof of Theorem 5.13.* For better readability, we are going to drop the subscripts and arguments  $\xi$  and  $\gamma^\uparrow, \gamma^\downarrow$ . For future reference define

$$\begin{aligned} a_Q(q) &= \frac{2S}{\sigma(1 - S)}(\ell(q) - n(q)) = \sigma \frac{\ell(q) - n_\xi(q)}{D}; \\ b_Q(q) &= q(n(q) - 2\theta S\{\ell(q) - n(q)\}), \end{aligned}$$

so that  $d\hat{Q}_t = a_Q(\hat{Q}_t) dB_t + b_Q(\hat{Q}_t) dt + dG_t^\uparrow - dG_t^\downarrow$ .

(a) Existence and uniqueness of a solution to (5.14) follows from standard results on reflecting one-dimensional diffusions, see, for example, [37, Theorem 4.1]. The additional claim is trivial on  $\llbracket 0, \tau \rrbracket$ . For the argument on  $\llbracket \tau, \infty \rrbracket$ , first note that  $a_Q(1) = 0$  and  $b_Q(1) = n_\xi(1) > 0$  by Proposition 5.3(e) and the fact that  $m(1) = \ell(1)$ .

If  $q = 1$ , then  $\hat{Q}_\tau = 1$  on  $\{\tau < \infty\}$  since  $\hat{Q} \in [q_*, q^*]$ . Moreover, it is not difficult to check that  $dG_t = b_Q(1) dt$  if  $t \geq \tau$ , and the process  $\hat{Q}$  remains constant on  $\llbracket \tau, \infty \rrbracket$ .

By contrast, if  $q^* > 1$ , then  $\hat{Q}_t > 1$  on  $\{\tau < t\}$  by Feller's boundary test for diffusions (see e.g. [24, Proposition 5.5.22 (a)]) since  $b_Q(q) > 0$  and  $a_Q(q) = O(1 - q)$  in a neighbourhood of 1 because  $\ell'(1) - n'_\xi(1) = \ell'(1) - m'(1) \neq 0$  by Proposition (5.14)(e) and the fact that  $q_M \neq 1$ .

(b) Let  $(\hat{\Phi}, \hat{C}, \hat{X})$  be defined as in the statement of the proposition. First, we argue that we cannot have both  $\phi y \gamma^\uparrow(1 - q_*) < q_* x$  and  $\frac{\phi y}{\gamma^\downarrow}(1 - q^*) > q^* x$ . We argue this in the case  $q_*, q^* > 0$ ; the case  $q_*, q^* < 0$  is similar, but easier.

If  $q_*, q^* > 0$  then it is sufficient to show that  $\gamma^\uparrow \frac{(1 - q_*)}{q_*} > \frac{1}{\gamma^\downarrow} \frac{(1 - q^*)}{q^*}$ . If  $0 < q_* \leq 1 < q^*$  then this is immediate, and if  $0 < q_* < q^* \leq 1$  it follows from the fact that  $\frac{1 - q}{q}$  is decreasing on  $(0, 1]$  and then  $\frac{1}{\gamma^\downarrow} \frac{(1 - q^*)}{q^*} \leq \frac{(1 - q^*)}{q^*} < \frac{(1 - q_*)}{q_*} \leq \gamma^\uparrow \frac{(1 - q_*)}{q_*}$ . Finally, if  $1 \leq q_* < q^*$  the result follows from  $\xi < \frac{(q^* - 1)}{q^*} \frac{q^*}{(q_* - 1)}$  as shown in the proof of Proposition 5.3.

Now consider  $(\hat{\Phi}_0, \hat{X}_0)$ . If both  $\frac{\phi y}{\gamma^\downarrow}(1 - q^*) \leq q^* x$  and  $q_* x \leq \phi y \gamma^\uparrow(1 - q_*)$  then  $\hat{\Phi}_0 = \phi$  and using  $\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q) = \frac{q(1 - p)}{(1 - q)p}$  and  $p = \frac{\phi x}{x + \phi y}$  we find  $\hat{X}_0 = \frac{1 - \hat{Q}_0}{\hat{Q}_0} Y_0 \kappa_{\gamma^\uparrow, \gamma^\downarrow}(\hat{Q}_0) \hat{\Phi}_0 = \frac{(1 - p)}{p} y \phi = x$ . In particular,  $(\hat{\Phi}_0, \hat{X}_0) = (\phi, x)$  and  $\hat{X}_0$  trivially satisfies the initial condition (2nd unlabelled equation in Section 4).

If  $\frac{\phi y}{\gamma^\downarrow}(1 - q^*) > q^*x$  or if  $q_*x > \phi y \gamma^\uparrow(1 - q_*)$  then we have to account for the initial re-balancing of the portfolio. We describe what happens in the latter case, the details for the former case are similar.

Suppose  $\phi y \gamma^\uparrow(1 - q_*) < q_*x$ . In this case  $\hat{\Phi}_0 = q_* \left( \phi + \frac{x}{y\gamma^\uparrow} \right) > \phi$  and  $\hat{Q}_0 = q_*$ . Then from  $\hat{X}_0 = \frac{1 - q_*}{q_*} y \gamma^\uparrow q_* \left( \phi + \frac{x}{y\gamma^\uparrow} \right)$  we find  $\hat{X}_0 = x - q_* \left( \hat{\Phi}_0 - \phi \right)$ . In particular  $\hat{X}_0$  satisfies the initial condition (4.2).

Rewriting (5.15) we find that  $\hat{Q}$  solves

$$\hat{Q}_t = \frac{Y_t \kappa_{\gamma^\uparrow, \gamma^\downarrow}(\hat{Q}_t) \hat{\Phi}_t}{\hat{X}_t + Y_t \kappa_{\gamma^\uparrow, \gamma^\downarrow}(\hat{Q}_t) \hat{\Phi}_t},$$

and hence that, provided  $\hat{X}$  represents the cash wealth of the agent (and  $\kappa_{\gamma^\uparrow, \gamma^\downarrow}(\hat{Q})$  is the relative shadow price),  $\hat{Q}$  solves (5.16) and has the interpretation as the shadow fraction of wealth. Since  $\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q_*) = \gamma^\uparrow$  and  $\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q^*) = \frac{1}{\gamma^\downarrow}$ , the claim that  $\hat{X} = X^{x, y, \phi, \hat{\Phi}, \hat{C}}$  will follow if we can show that

$$d\hat{X}_t = \left( r\hat{X}_t - \hat{C}_t \right) dt - Y_t \kappa_{\gamma^\uparrow, \gamma^\downarrow}(\hat{Q}_t) d\hat{\Phi}_t, \quad (\text{A.17})$$

so that  $\hat{X}$  is the wealth process associated with consumption  $\hat{C}$  and trading strategy  $\hat{\Phi}$ .

We look for an expression for  $d\hat{X}_t$  of the form  $d\hat{X}_t = a_t dB_t + b_t dt + c_t dG_t$  where  $a_t = a_X(\hat{X}_t, Y_t, \hat{Q}_t, \hat{\Phi}_t)$  and similarly for  $b_t$  and  $c_t$ .

In the sequel, we only consider the case  $0 < q^*(\xi) < q^*(\xi) < 1$  so that  $\hat{Q} \in (0, 1)$ . The other cases in Proposition 5.3 can be treated in a similar way, replacing  $\log(1 - \hat{Q})$  by  $\log(\hat{Q} - 1)$  if  $\hat{Q} > 1$  and replacing  $\log \hat{Q}$  and  $\log \hat{\Phi}$  by  $\log(-\hat{Q})$  and  $\log(-\hat{\Phi})$ , respectively, if  $\hat{Q} < 0$ . Special care is needed if  $1 \in [q^*(\xi) < q^*(\xi)]$ . In that case, by the final part of (a), one has to argue separately on  $\llbracket 0, \tau \rrbracket$  and  $\llbracket \tau, \infty \rrbracket$ . We leave the details to the reader.

From (5.15) we have  $\ln \hat{X}_t = \ln(1 - \hat{Q}_t) - \ln \hat{Q}_t + \ln Y_t + \ln \kappa_{\gamma^\uparrow, \gamma^\downarrow}(\hat{Q}_t) + \ln \hat{\Phi}_t$  and then

$$\begin{aligned} \frac{d\hat{X}_t}{\hat{X}_t} - \frac{1}{2} \frac{d[\hat{X}]_t}{\hat{X}_t^2} &= -\frac{d\hat{Q}_t}{(1 - \hat{Q}_t)} - \frac{1}{2} \frac{d[\hat{Q}]_t}{(1 - \hat{Q}_t)^2} - \frac{d\hat{Q}_t}{\hat{Q}_t} + \frac{1}{2} \frac{d[\hat{Q}]_t}{\hat{Q}_t^2} + \frac{dY_t}{Y_t} - \frac{1}{2} \frac{d[Y]_t}{Y_t^2} \\ &\quad + \frac{\kappa'_{\gamma^\uparrow, \gamma^\downarrow}(\hat{Q}_t)}{\kappa_{\gamma^\uparrow, \gamma^\downarrow}(\hat{Q}_t)} d\hat{Q}_t + \frac{1}{2} \left( \frac{\kappa''_{\gamma^\uparrow, \gamma^\downarrow}(\hat{Q}_t)}{\kappa_{\gamma^\uparrow, \gamma^\downarrow}(\hat{Q}_t)} - \left( \frac{\kappa'_{\gamma^\uparrow, \gamma^\downarrow}(\hat{Q}_t)}{\kappa_{\gamma^\uparrow, \gamma^\downarrow}(\hat{Q}_t)} \right)^2 \right) d[\hat{Q}]_t + \frac{d\hat{\Phi}_t}{\hat{\Phi}_t}. \end{aligned}$$

Equating co-efficients, and using  $\frac{\kappa'_{\gamma^\uparrow, \gamma^\downarrow}(q)}{\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q)} - \frac{1}{q(1-q)} = -\frac{D}{\ell(q)-n_\xi(q)}$  and the definition of  $a_Q(q)$  we find

$$\frac{a_X(x, y, q, \phi)}{x} = \left( a_Q(q) \left( \frac{\kappa'_{\gamma^\uparrow, \gamma^\downarrow}(q)}{\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q)} - \frac{1}{q(1-q)} \right) + \sigma \right) = 0$$

In particular,  $\hat{X}$  has no quadratic variation term. Further, using  $\frac{d\hat{\Phi}_t}{\hat{\Phi}_t} = \frac{dG_t^\uparrow}{\hat{Q}_t} - \frac{dG_t^\downarrow}{\hat{Q}_t} = \frac{dG}{\hat{Q}_t}$  and the fact that  $\kappa'_{\gamma^\uparrow, \gamma^\downarrow} = 0$  at  $q_*$  and  $q^*$ ,

$$\frac{c_X(x, y, q, \phi)}{x} = \frac{\kappa'_{\gamma^\uparrow, \gamma^\downarrow}(q)}{\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q)} - \frac{1}{q(1-q)} + \frac{1}{q} = \frac{-1}{1-q}.$$

Then  $c_t dG_t = -\frac{\hat{X}_t}{1 - \hat{Q}_t} dG_t = -\frac{\hat{X}_t \hat{Q}_t}{1 - \hat{Q}_t} \frac{d\hat{\Phi}_t}{\hat{\Phi}_t} = -Y_t \kappa_{\gamma^\uparrow, \gamma^\downarrow}(\hat{Q}_t) d\hat{\Phi}_t$ . Finally,

$$\begin{aligned} \frac{b_X(x, y, q, \phi)}{x} &= \left( \frac{\kappa'_{\gamma^\uparrow, \gamma^\downarrow}(q)}{\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q)} - \frac{1}{(1-q)} - \frac{1}{q} \right) b_Q(q) + \mu - \frac{\sigma^2}{2} \\ &\quad + \frac{a_Q(q)^2}{2} \left( -\frac{1}{(1-q)^2} + \frac{1}{q^2} + \frac{\kappa''_{\gamma^\uparrow, \gamma^\downarrow}(q)}{\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q)} - \left( \frac{\kappa'_{\gamma^\uparrow, \gamma^\downarrow}(q)}{\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q)} \right)^2 \right). \end{aligned}$$

Then, differentiating  $(1-q)q^{\frac{\kappa'_{\gamma\uparrow,\gamma\downarrow}(q)}{\kappa_{\gamma\uparrow,\gamma\downarrow}(q)}} = \frac{m(q)-n_\xi(q)}{\ell(q)-n_\xi(q)}$  and using  $\ell(q) = m(q) + Dq(1-q)$ ,  $\ell'(q) - m'(q) = D(1-2q)$  and  $\ell(q)m'(q) - m(q)\ell'(q) = D(q(1-q)m'(q) - (1-2q)m)$  yields

$$\begin{aligned}
& \frac{\kappa''_{\gamma\uparrow,\gamma\downarrow}(q)}{\kappa_{\gamma\uparrow,\gamma\downarrow}(q)} - \left( \frac{\kappa'_{\gamma\uparrow,\gamma\downarrow}(q)}{\kappa_{\gamma\uparrow,\gamma\downarrow}(q)} \right)^2 \\
&= -\frac{(1-2q)}{q^2(1-q)^2} \frac{m(q)-n_\xi(q)}{\ell(q)-n_\xi(q)} \\
&\quad + \frac{(\ell(q)m'(q) - m(q)\ell'(q)) + n_\xi(q)(\ell'(q) - m'(q)) - n'(q)(\ell(q) - m(q))}{q(1-q)(\ell(q) - n_\xi(q))^2} \\
&= -\frac{(1-2q)}{q^2(1-q)^2} \frac{(m(q) - n_\xi(q))}{(\ell(q) - n_\xi(q))} \\
&\quad + \frac{D}{q(1-q)(\ell(q) - n_\xi(q))^2} \left( q(1-q)m'(q) - (1-2q)(m(q) - n_\xi(q)) - q \frac{1-S}{S} n_\xi(q) \frac{m(q) - n_\xi(q)}{\ell(q) - n_\xi(q)} \right)
\end{aligned}$$

It follows that, using  $m'(q) = \frac{R(1-S)}{S}\sigma^2 q - \frac{(1-S)}{S}(\mu - r)$ ,

$$\begin{aligned}
& \frac{b_X(x, y, q, \phi)}{x} \\
&= -\frac{Dq}{\ell(q) - n_\xi(q)} (n_\xi(q) - 2\theta S(\ell(q) - n_\xi(q))) + \mu - \frac{\sigma^2}{2} + \frac{\sigma^2(\ell(q) - n_\xi(q))^2}{2D^2} \frac{(1-q)^2 - q^2}{(1-q)^2 q^2} \\
&\quad - \frac{1}{2} \sigma^2 (\ell(q) - n_\xi(q))^2 \frac{(1-2q)(m(q) - n_\xi(q))}{(\ell(q) - m(q))^2 (\ell(q) - n_\xi(q))} \\
&\quad + \frac{\sigma^2}{2(\ell(q) - m(q))} \left( q(1-q) \left( \frac{R(1-S)}{S} \sigma^2 q - \frac{(1-S)}{S} (\mu - r) \right) - (1-2q)(m(q) - n_\xi(q)) \right. \\
&\quad \left. - q \frac{1-S}{S} \frac{n_\xi(q)(m(q) - n_\xi(q))}{\ell(q) - n_\xi(q)} \right) \\
&= \mu - \frac{\sigma^2}{2} + q\sigma^2(1-R) - \frac{Dqn_\xi(q)}{(\ell(q) - n_\xi(q))} + (R\sigma^2 q - (\mu - r)) + (1-2q) \frac{\sigma^2}{2} \frac{(\ell(q) - n_\xi(q))^2}{(\ell(q) - m(q))^2} \\
&\quad - (1-2q) \frac{\sigma^2}{2} \frac{(\ell(q) - n_\xi(q))(m(q) - n_\xi(q))}{(\ell(q) - m(q))^2} - (1-2q) \frac{\sigma^2}{2} \frac{(m(q) - n_\xi(q))}{(\ell(q) - m(q))} \\
&\quad - q \frac{\sigma^2}{2} \frac{(1-S)}{S} \frac{n_\xi(q)(m(q) - n_\xi(q))}{(\ell(q) - m(q))(\ell(q) - n_\xi(q))} \\
&= r - \frac{Dqn_\xi(q)}{\ell(q) - n_\xi(q)} \left( 1 + \frac{m(q) - n_\xi(q)}{\ell(q) - m(q)} \right) \\
&\quad + (1-2q) \frac{\sigma^2}{2} \left[ -1 + \frac{(\ell(q) - n_\xi(q))^2}{(\ell(q) - m(q))^2} - \frac{(\ell(q) - n_\xi(q))(m(q) - n_\xi(q))}{(\ell(q) - m(q))^2} - \frac{(m(q) - n_\xi(q))}{(\ell(q) - m(q))} \right] \\
&= r - \frac{n_\xi(q)}{1-q}
\end{aligned}$$

Then  $b_t = r\hat{X}_t - \frac{\hat{X}_t}{1-\hat{Q}_t}n(\hat{Q}_t) = r\hat{X}_t - \frac{Y_t \kappa_{\gamma\uparrow,\gamma\downarrow}(\hat{Q}_t)\hat{\Phi}_t}{\hat{Q}_t}n(\hat{Q}_t) = r\hat{X}_t - \hat{C}_t$ . Combining the expressions for  $a_t$ ,  $b_t$  and  $c_t$  we conclude that  $\hat{X}$  does indeed solve (A.17).

It only remains to show that  $(\hat{\Phi}, \hat{C}) \in \mathcal{A}$ . We consider the case where  $q_*, q^* > 0$ . Then  $\Phi_t > 0$  and it is sufficient to show that  $\hat{X}_t + \frac{\hat{\Phi}_t Y_t}{\gamma\downarrow} > 0$ . Substituting for  $\hat{X}_t$  using (5.15) we have  $\hat{X}_t + \frac{\hat{\Phi}_t Y_t}{\gamma\downarrow} = \hat{\Phi}_t Y_t \left( \frac{1-\hat{Q}_t}{\hat{Q}_t} \kappa_{\gamma\uparrow,\gamma\downarrow}(\hat{Q}_t) + \frac{1}{\gamma\downarrow} \right)$ . Set  $\chi(q) = \frac{(1-q)}{q} \kappa_{\gamma\uparrow,\gamma\downarrow}(q) + \frac{1}{\gamma\downarrow}$ : we want to show that  $\chi(q) > 0$  on  $(q_*, q^*)$ .

The result is immediate on  $(q_* \wedge 1, 1]$  so it is sufficient to consider  $\chi$  on  $[q_* \vee 1, q^*]$ . But on this interval  $\chi'(q) = -\frac{\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q)}{q^2} \left( \frac{\ell(q) - m(q)}{\ell(q) - n_\xi(q)} \right) < 0$  and the positivity of  $\chi$  on  $[q_* \vee 1, q^*]$  follows from the monotonicity of  $\chi$  and the fact that  $\chi(q^*) = \frac{1}{q^* \gamma^\downarrow}$ .  $\square$

**Proposition A.7.** *Let  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$  with  $\xi := \gamma^\uparrow \gamma^\downarrow > 1$  be such that  $\xi$  is in the well-posedness range of transaction costs. Let  $(x, y, \phi) \in \mathcal{S}_{(\gamma^\uparrow, \gamma^\downarrow)}^\circ$ . Let  $(\hat{\Phi}, \hat{C}) \in \mathcal{A}_{(\gamma^\uparrow, \gamma^\downarrow)}(x, y, \phi)$  be the candidate optimal investment-consumption pair,  $\hat{X} = (\hat{X}_t)_{t \geq 0} := (X_t^{x, y, \phi, \hat{\Phi}, \hat{C}})_{t \geq 0}$  the corresponding cash process, and  $\hat{Q} = (\hat{Q}_t)_{t \geq 0}$  the corresponding shadow fraction of wealth. Then*

$$\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, \hat{X}_t, Y_t, \hat{\Phi}_t) = \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(0, \hat{X}_0, Y_0, \hat{\Phi}_0) \mathcal{E}(\sigma(1 - R)\hat{Q} \bullet B)_t e^{-\theta \int_0^t n_\xi(\hat{Q}_s) ds}, \quad t \geq 0. \quad (\text{A.18})$$

Moreover, the process  $\hat{M} = (\hat{M}_t)_{t \geq 0}$  defined by

$$\hat{M}_t := \int_0^t g_{EZ}(s, \hat{C}_s, \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(s, \hat{X}_s, Y_s, \hat{\Phi}_s)) ds + \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, \hat{X}_t, Y_t, \hat{\Phi}_t).$$

is a uniformly integrable martingale with  $\hat{M}_\infty := \int_0^\infty g_{EZ}(s, \hat{C}_s, \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(s, \hat{X}_s, Y_s, \hat{\Phi}_s)) ds$

*Proof.* First, it follows from the definition of  $\hat{C}$  in Theorem 5.13(c), (A.8) and (A.11) that

$$g_{EZ}(t, \hat{C}_t, \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, \hat{X}_t, Y_t, \hat{\Phi}_t)) = \theta n_\xi(\hat{Q}_t) \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, \hat{X}_t, Y_t, \hat{\Phi}_t), \quad t \geq 0. \quad (\text{A.19})$$

Second, Itô's formula, Proposition A.5(a) and (b) and the properties of  $\hat{\Phi}, \hat{C}$  from Theorem 5.13 give

$$d\hat{M}_t = \sigma(1 - R) \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, \hat{X}_t, Y_t, \hat{\Phi}_t) \hat{Q}_t dB_t, \quad (\text{A.20})$$

Third, combining (A.20) with (A.19) yields

$$\begin{aligned} d\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, \hat{X}_t, Y_t, \hat{\Phi}_t) &= d\hat{M}_t - g_{EZ}(t, \hat{C}_t, \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, \hat{X}_t, Y_t, \hat{\Phi}_t)) dt \\ &= \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, \hat{X}_t, Y_t, \hat{\Phi}_t) \left( \sigma(1 - R) \hat{Q}_t dB_t - \theta n_\xi(\hat{Q}_t) dt \right). \end{aligned}$$

The formula for the stochastic exponential now gives (A.18).

Next, fix  $\beta > 0$  and let  $C > 0$  be a uniform bound for  $\hat{Q}$  and  $\nu > 0$  a lower bound for  $n_\xi$ . Since  $\hat{Q}$  is uniformly bounded, Novikov's condition implies that  $\mathcal{E}(\beta \sigma(1 - R)\hat{Q} \bullet B)$  is a true martingale. Hence, (A.18) yields

$$\begin{aligned} \mathbb{E} \left[ |\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, \hat{X}_t, Y_t, \hat{\Phi}_t)|^\beta \right] &= |\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(0, \hat{X}_0, Y_0, \hat{\Phi}_0)|^\beta \mathbb{E} \left[ e^{\beta(\beta-1) \frac{\sigma^2(1-R)^2}{2} \int_0^t \hat{Q}_s^2 ds - \theta \int_0^t n_\xi(\hat{Q}_s) ds} \right] \\ &\leq |\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(0, \hat{X}_0, Y_0, \hat{\Phi}_0)|^\beta e^{\beta(\beta-1) \frac{\sigma^2(1-R)^2}{2} C^2 - \theta \nu t}, \end{aligned} \quad (\text{A.21})$$

where  $C$  is a (upper or lower as required) bound on  $\hat{Q}$ , and  $\nu$  is a lower bound on  $n_\xi$ . For  $\beta = 2$ , (A.21) shows that  $\hat{M}$  is a square-integrable martingale. For  $\beta = 1$ , (A.21) shows that  $\lim_{t \rightarrow \infty} \mathbb{E} \left[ |\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, \hat{X}_t, Y_t, \hat{\Phi}_t)| \right] = 0$  and together with (A.19) and the fact that  $n_\xi$  is uniformly bounded, this implies that  $\hat{M}_t$  converges in  $L^1$  to  $\hat{M}_\infty := \int_0^\infty g_{EZ}(s, \hat{C}_s, \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(s, \hat{X}_s, Y_s, \hat{\Phi}_s)) ds$ . Hence,  $\hat{M}$  is a uniformly integrable martingale.  $\square$

**Definition A.8.** Suppose that  $X = (X_t)_{t \geq 0}$  and  $Y = (Y_t)_{t \geq 0}$  are nonnegative progressive processes. We say that  $X$  has the same order as  $Y$  (and write  $X \stackrel{\mathbb{O}}{=} Y$ , noting that  $\stackrel{\mathbb{O}}{=}$  is an equivalence relation) if there exist constants  $k, K \in (0, \infty)$  such that

$$0 \leq kY \leq X \leq KY.$$

Denote the set of progressive processes with the same order as  $X$  by  $\mathbb{O}(X) \subseteq \mathcal{P}_+$ .

**Definition A.9.** For  $X \in \mathcal{P}_+$ , define  $J^X = (J_t^X)_{t \geq 0}$  by  $J_t^X = \mathbb{E} [\int_t^\infty X_s \, ds \mid \mathcal{F}_t]$ . Then the set  $\mathbb{SO}$  of *self-order* consumption streams is given by

$$\mathbb{SO} := \left\{ X \in \mathcal{P}_{++} : \mathbb{E} \left[ \int_0^\infty X_t \, dt \right] < \infty \text{ and } X \stackrel{\mathbb{O}}{=} J^X \right\}.$$

Moreover, set

$$\mathbb{SO}_\nu := \left\{ X \in \mathbb{SO} : (e^{\nu t} X_t)_{t \geq 0} \in \mathbb{SO} \right\} \quad \text{for } \nu \geq 0, \quad \text{and} \quad \mathbb{SO}_+ = \cup_{\nu > 0} \mathbb{SO}_\nu.$$

**Lemma A.10.** Let  $X, Y \in \mathcal{P}_+$  with  $X \stackrel{\mathbb{O}}{=} Y$ . Then for  $\nu \geq 0$ ,  $X \in \mathbb{SO}_\nu$  if and only if  $Y \in \mathbb{SO}_\nu$ . In particular,  $X \in \mathbb{SO}_+$  if and only if  $Y \in \mathbb{SO}_+$ .

**Corollary A.11.** Let  $\gamma^\uparrow, \gamma^\downarrow \in [1, \infty)$  with  $\xi := \gamma^\uparrow \gamma^\downarrow > 1$  be such that  $\xi$  lies in the well-posedness range of transaction costs. Let  $(x, y, \phi) \in \mathcal{S}_{(\gamma^\uparrow, \gamma^\downarrow)}^\circ$ . Let  $(\hat{\Phi}, \hat{C}) \in \mathcal{A}_{(\gamma^\uparrow, \gamma^\downarrow)}(x, y, \phi)$  be the candidate optimal investment-consumption pair,  $\hat{X} = (\hat{X}_t)_{t \geq 0} := (X_t^{x, y, \phi, \hat{\Phi}, \hat{C}})_{t \geq 0}$  the corresponding cash process, and  $\hat{Q} = (\hat{Q}_t)_{t \geq 0}$  the corresponding shadow fraction of wealth. Then the process  $(e^{-\delta\theta t} \hat{C}_t^{1-R})_{t \geq 0}$  lies in  $\mathbb{SO}_+$ .

Further,  $\hat{C}$  is uniquely proper.

*Proof.* First, note that

$$e^{-\delta\theta t} \hat{C}_t^{1-R} = e^{-\delta\theta t} (\hat{X}_t + \hat{\Phi}_t \tilde{Y}_t)^{1-R} n_\xi(\hat{Q}_t)^{1-R} = (1-R) \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, \hat{X}_t, Y_t, \hat{\Phi}_t) n_\xi(\hat{Q}_t)^\theta$$

Since  $n_\xi$  is uniformly bounded from above and from below, by Lemma A.10 it remains to show that that  $((1-R) \hat{V}(t, \hat{X}_t, Y_t, \hat{\Phi}_t))_{t \geq 0} \in \mathbb{SO}_+$ .

To this end, it suffices to show that  $((1-R) \hat{V}(t, \hat{X}_t, Y_t, \hat{\Phi}_t))_{t \geq 0} \in \mathbb{SO}_\nu$  for some  $\nu > 0$ . Since  $n$  is uniformly bounded from above and below by a positive constants, there is  $D > \nu > 0$  such that  $D \geq \theta n + \nu \geq \nu > 0$ . Now using that  $\mathcal{E}(\sigma(1-R)\hat{Q} \bullet B)_t$  is a martingale by Novikov's condition and the fact that  $\hat{Q}$  is uniformly bounded, we obtain

$$\begin{aligned} J_t^{\exp(\nu \cdot) \hat{V}} &= \mathbb{E} \left[ \int_t^\infty \exp(\nu s) (1-R) \hat{V}(s, \hat{X}_s, Y_s, \hat{\Phi}_s) \, ds \mid \mathcal{F}_t \right] \\ &= (1-R) \hat{V}(t, \hat{X}_t, Y_t, \hat{\Phi}_t) \mathbb{E} \left[ \int_t^\infty \exp \left( - \int_t^s \theta n(\hat{Q}_u) - \nu \, du \right) \, ds \mid \mathcal{F}_t \right] \end{aligned} \quad (\text{A.22})$$

The claim that  $(e^{-\delta\theta t} \hat{C}_t^{1-R})_{t \geq 0}$  lies in  $\mathbb{SO}_+$  now follows from the fact that the last factor on the right hand side of (A.22) lies in  $[\frac{1}{D}, \frac{1}{\nu}]$ .

The fact that there is a unique utility process  $V^{\hat{C}}$  associated to  $\hat{C}$  now follows from Herdegen et al. [20, Theorem 4.7].  $\square$

*Proof of Theorem 5.15.* First, it follows from Proposition A.11 and Theorem 5.13 that  $\hat{C}$  is uniquely proper and has continuous paths.

Next, define the process  $\hat{M} = (\hat{M}_t)_{t \geq 0}$  by

$$\hat{M}_t := \int_0^t g_{EZ}(s, \hat{C}_s, \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(s, \hat{X}_s, Y_s, \hat{\Phi}_s)) \, ds + \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, \hat{X}_t, Y_t, \hat{\Phi}_t).$$

By Proposition A.7,  $\hat{M}$  is a uniformly integrable martingale with

$$\hat{M}_\infty = \int_0^\infty e^{-\delta s} \frac{(\hat{C}_s)^{1-S}}{1-S} ((1-R) \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(s, \hat{X}_s, Y_s, \hat{\Phi}_s))^\rho \, ds.$$

This implies that

$$\begin{aligned}
\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, \hat{X}_t, Y_t, \hat{\Phi}_t) &= \hat{M}_t - \int_0^t g_{EZ}(s, \hat{C}_s, \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(s, \hat{X}_s, Y_s, \hat{\Phi}_s)) \, ds \\
&= \mathbb{E} \left[ \hat{M}_\infty \mid \mathcal{F}_t \right] - \int_0^t g_{EZ}(s, \hat{C}_s, \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(s, \hat{X}_s, Y_s, \hat{\Phi}_s)) \, ds \\
&= \mathbb{E} \left[ \int_t^\infty g_{EZ}(s, \hat{C}_s, \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(s, \hat{X}_s, Y_s, \hat{\Phi}_s)) \mid \mathcal{F}_t \right].
\end{aligned}$$

Thus  $(\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(t, \hat{X}_t, Y_t, \hat{\Phi}_t))_{t \geq 0}$  is a utility process for  $\hat{C}$ . Note that it is never vanishes and hence is a proper solution. Thus, it coincides with  $V^{\hat{C}}$  by uniqueness of utility processes (for proper solutions).  $\square$

#### A.4 Verification argument

*Proof of Proposition 5.17.* First, note that the dynamics of  $X + \varepsilon \hat{X}$  are given by

$$d(X_t + \varepsilon \hat{X}_t) = r(X_t + \varepsilon \hat{X}_t) \, dt - (C_t + \varepsilon \hat{C}_t) \, dt - Y_t \gamma^\uparrow d(\Phi_t^\uparrow + \varepsilon \hat{\Phi}_t^\uparrow) + \frac{Y_t}{\gamma^\downarrow} d(\Phi_t^\downarrow + \varepsilon \hat{\Phi}_t^\downarrow).$$

Next, fix arbitrary bounded stopping times  $\tau_1 \leq \tau_2$ . Using Itô's formula, Proposition A.5, Corollary A.6 and the fact that  $\Phi_t^\uparrow + \varepsilon \hat{\Phi}_t^\uparrow - (\Phi_t + \varepsilon \hat{\Phi}_t)^\uparrow$  and  $\Phi_t^\downarrow + \varepsilon \hat{\Phi}_t^\downarrow - (\Phi_t + \varepsilon \hat{\Phi}_t)^\downarrow$  are nondecreasing and identical for  $t \geq 0$ , we obtain

$$\begin{aligned}
&\hat{V}(X_{\tau_2} + \varepsilon \hat{X}_{\tau_2}, Y_{\tau_2}, \Phi_{\tau_2} + \varepsilon \hat{\Phi}_{\tau_2}) - \hat{V}(X_{\tau_1} + \varepsilon \hat{X}_{\tau_1}, Y_{\tau_1}, \Phi_{\tau_1} + \varepsilon \hat{\Phi}_{\tau_1}) \\
&+ \int_{\tau_1}^{\tau_2} g_{EZ}(C_u + \varepsilon \hat{C}_u, \hat{V}(X_u + \varepsilon \hat{X}_u, Y_u, \Phi_u + \varepsilon \hat{\Phi}_u)) \, du \\
&= \int_{\tau_1}^{\tau_2} \left( \hat{V}_u \, du + \hat{V}_x \, d(X_u + \varepsilon \hat{X}_u) + \hat{V}_y \, dY_u + \hat{V}_\phi \, d(\Phi_u + \varepsilon \hat{\Phi}_u) + \frac{1}{2} \hat{V}_{yy} \, d\langle Y \rangle_u + g_{EZ}(C_u + \varepsilon \hat{C}_u, \hat{V}) \, du \right) \\
&= \int_{\tau_1}^{\tau_2} \left( \hat{V}_t + r(X_u + \varepsilon \hat{X}_u) \hat{V}_x - (C_u + \varepsilon \hat{C}_u) \hat{V}_x + \mu Y_u \hat{V}_y + \frac{1}{2} \sigma^2 Y_u^2 \hat{V}_{yy} + g_{EZ}(C_u + \varepsilon \hat{C}_u, \hat{V}) \right) \, du \\
&+ \int_{\tau_1}^{\tau_2} \left( \hat{V}_\phi - \gamma^\uparrow Y_u \hat{V}_x \right) \, d(\Phi_u + \varepsilon \hat{\Phi}_u)^\uparrow - \int_{\tau_1}^{\tau_2} \gamma^\uparrow Y_u \hat{V}_x \, d(\Phi_u^\uparrow + \varepsilon \hat{\Phi}_u^\uparrow - (\Phi_u + \varepsilon \hat{\Phi}_u)^\uparrow) \\
&+ \int_{\tau_1}^{\tau_2} \left( \frac{Y_u}{\gamma^\downarrow} \hat{V}_x - \hat{V}_\phi \right) \, d(\Phi_u + \varepsilon \hat{\Phi}_u)^\downarrow + \int_{\tau_1}^{\tau_2} \frac{Y_u}{\gamma^\downarrow} \hat{V}_x \, d(\Phi_u^\downarrow + \varepsilon \hat{\Phi}_u^\downarrow - (\Phi_u + \varepsilon \hat{\Phi}_u)^\downarrow) \\
&+ \int_{\tau_1}^{\tau_2} \sigma Y_u \hat{V}_s \, dW_u \\
&\leq - \int_{\tau_1}^{\tau_2} \hat{V}_x Y_u \left( \gamma^\uparrow - \frac{1}{\gamma^\downarrow} \right) \, d(\Phi_u^\uparrow + \varepsilon \hat{\Phi}_u^\uparrow - (\Phi_u + \varepsilon \hat{\Phi}_u)^\uparrow) + \int_{\tau_1}^{\tau_2} \sigma Y_u \hat{V}_y \, dW_u \leq \int_{\tau_1}^{\tau_2} \sigma Y_u \hat{V}_y \, dW_u.
\end{aligned}$$

Define the local martingale  $N = (N_t)_{t \geq 0}$  by

$$N_t = \int_0^t \sigma Y_u \hat{V}_y(u, X_u + \varepsilon \hat{X}_u, Y_u, \Phi_u + \varepsilon \hat{\Phi}_u) \, dW_u.$$

Moreover, for  $n \in \mathbb{N}$ , define the stopping time  $\zeta_n := \inf\{s \geq \tau_1 : \langle N \rangle_s - \langle N \rangle_{\tau_1} \geq n\}$ . Then

$$\begin{aligned}
&\hat{V}(X_{\tau_1} + \varepsilon \hat{X}_{\tau_1}, Y_{\tau_1}, \Phi_{\tau_1} + \varepsilon \hat{\Phi}_{\tau_1}) \geq \hat{V}(X_{\tau_2 \wedge \zeta_n} + \varepsilon \hat{X}_{\tau_2 \wedge \zeta_n}, Y_{\tau_2 \wedge \zeta_n}, \Phi_{\tau_2 \wedge \zeta_n} + \varepsilon \hat{\Phi}_{\tau_2 \wedge \zeta_n}) \\
&+ \int_{\tau_1}^{\tau_2 \wedge \zeta_n} e^{-\delta u} \frac{(C_u + \varepsilon \hat{C}_u)^{1-S}}{1-S} ((1-R) \hat{V}(u, X_u + \varepsilon \hat{X}_u, S_u, \Phi_u + \varepsilon \hat{\Phi}_u))^\rho \, du + N_{\tau_1} - N_{\tau_2 \wedge \zeta_n}.
\end{aligned}$$

Taking conditional expectations and using that  $(N_{t \wedge \zeta_n} - N_{t \wedge \tau_1})_{t \geq 0}$  is an  $L^2$ -bounded martingale, the Optional Sampling Theorem gives

$$\hat{V}_{\tau_1} \geq \mathbb{E} \left[ \hat{V}_{\tau_2 \wedge \zeta_n} + \int_{\tau_1}^{\tau_2 \wedge \zeta_n} e^{-\delta u} \frac{(C_u + \varepsilon \hat{C}_u)^{1-S}}{1-S} ((1-R)\hat{V}_u)^\rho \, du \middle| \mathcal{F}_{\tau_1} \right]. \quad (\text{A.23})$$

where in a slight abuse of notation we have written  $\hat{V}_u$  as shorthand for  $\hat{V}(u, X_u + \varepsilon \hat{X}_u, Y_u, \Phi_u + \varepsilon \hat{\Phi}_u)$ . Since  $\hat{V}_u \geq \hat{V}(u, \varepsilon \hat{X}_u, Y_u, \varepsilon \hat{\Phi}_u)$  by Corollary A.6 and using  $(\hat{V}(t, \varepsilon \hat{X}_t, Y_t, \varepsilon \hat{\Phi}_t))_{t \geq 0}$  is bounded below by zero if  $R < 1$  and by a uniformly integrable martingale if  $R > 1$  by Proposition A.7, taking the liminf as  $n \rightarrow \infty$  in (A.23), the conditional version of Fatou's Lemma (with a UI martingale lower bound if  $R > 1$ ) and the conditional Monotone Convergence Theorem yield

$$\hat{V}_{\tau_1} \geq \mathbb{E} \left[ \hat{V}_{\tau_2} + \int_{\tau_1}^{\tau_2} e^{-\delta u} \frac{(C_u + \varepsilon \hat{C}_u)^{1-S}}{1-S} ((1-R)\hat{V}_u)^\rho \, du \middle| \mathcal{F}_{\tau_1} \right]. \quad (\text{A.24})$$

Furthermore,  $\liminf_{t \rightarrow \infty} \mathbb{E}[\hat{V}_t] \geq \lim_{t \rightarrow \infty} \mathbb{E}[\hat{V}(t, \varepsilon \hat{X}_t, S_t, \varepsilon \hat{\Phi}_t)] = 0$ . Consequently,  $\hat{V}(I, X + \varepsilon \hat{X}, Y, \Phi + \varepsilon \hat{\Phi})$  is a supersolution for  $C + \varepsilon \hat{C}$ .  $\square$

*Proof of Theorem 5.18.* The argument is different for  $\theta \leq 1$  and  $\theta > 1$ .

**Case  $\theta \leq 1$ .** By Theorem 5.15, Proposition A.5(c) and continuity of  $\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}$ , it suffices to show that for each  $\varepsilon > 0$  and  $C \in \mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}(x, y, \phi)$ ,

$$V_0^C \leq \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(0, x(1 + \varepsilon), y, \phi(1 + \varepsilon)). \quad (\text{A.25})$$

So fix  $\varepsilon > 0$  and  $C \in \mathcal{C}^{(\gamma^\uparrow, \gamma^\downarrow)}(x, y, \phi)$  with corresponding investment process  $\Phi$  and cash process  $X = X^{x, y, \phi, \Phi, C}$ . Moreover, set  $\hat{X} = \hat{X}^{x, y, \phi, \hat{\Phi}, \hat{C}}$ . The cases  $R < 1$  and  $R > 1$  require slightly different arguments.

First, assume  $R < 1$ . Then  $\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(I, X + \varepsilon \hat{X}, Y, \Phi + \varepsilon \hat{\Phi})$  is a supersolution for the pair  $(g_{EZ}, C + \varepsilon \hat{C})$  by Proposition 5.17. Since  $C + \varepsilon \hat{C} > C$  and  $\frac{c^{1-S}}{1-S}$  is increasing in  $c$ ,  $\hat{V}_{\gamma^\uparrow, \gamma^\downarrow}(I, X + \varepsilon \hat{X}, Y, \Phi + \varepsilon \hat{\Phi})$  is also supersolution for  $C$  by (A.24). Moreover, the (generalised) utility process  $V^C$  for  $C$  is the minimal supersolution for  $C$  by [19, Theorem 6.5]. Consequently,  $\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(I, X + \varepsilon \hat{X}, Y, \Phi + \varepsilon \hat{\Phi}) \geq V^C$ , and (A.25) follows from considering  $t = 0$  and using Corollary A.6(b).

Next, assume  $R > 1$ , which implies  $S > 1$  by the fact that  $\theta > 0$ . Since  $(C + \varepsilon \hat{C})^{1-S} \leq (\varepsilon \hat{C})^{1-S}$ ,  $C + \varepsilon \hat{C}$  is uniquely evaluable by Corollary A.11 and [19, Corollary 6.3]. Hence, there exists a uniformly integrable utility process  $V^{C+\varepsilon \hat{C}}$ , which is in particular a subsolution for  $C$ . By the comparison theorem for sub- and supersolutions in the form of [19, Theorem 5.8], it follows that  $\hat{V}(I, X + \varepsilon \hat{X}, Y, \Phi + \varepsilon \hat{\Phi}) \geq V^{C+\varepsilon C}$ . Moreover, since  $V^{C+\varepsilon C} \geq V^C$  by [19, Proposition 6.8],  $\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(I, X + \varepsilon \hat{X}, Y, \Phi + \varepsilon \hat{\Phi}) \geq V^C$ , and (A.25) follows from considering  $t = 0$  and using Corollary A.6(b).

**Case  $\theta > 1$ .** By Theorem 5.15 and Proposition A.5(c), it suffices to show that for each  $C \in \mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}(x, y, \phi) \cap \mathbb{UP}^*$ ,

$$V_0^C \leq \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(0, x, y, \phi), \quad (\text{A.26})$$

So fix  $C \in \mathcal{C}^{(\gamma^\uparrow, \gamma^\downarrow)}(x, y, \phi) \cap \mathbb{UP}^*$  with corresponding investment process  $\Phi$  and cash process  $X = X^{x, y, \phi, \Phi, C}$ . Moreover, set  $\hat{X} = \hat{X}^{x, y, \phi, \hat{\Phi}, \hat{C}}$ . As in the case  $\theta \leq 1$ , the cases  $R < 1$  and  $R > 1$  require slightly different arguments.

First, assume  $R > 1$ . For each  $\varepsilon > 0$ ,  $\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(I, X + \varepsilon \hat{X}, Y, \Phi + \varepsilon \hat{\Phi})$  is a supersolution for the pair  $(g_{EZ}, C + \varepsilon \hat{C})$  by Proposition 5.17. Since  $e^{-\delta \theta t} (C_t + \varepsilon \hat{C}_t)^{1-R} \leq e^{-\delta \theta t} (\varepsilon \hat{C})^{1-R}$  for all  $t \geq 0$ , it follows from Corollary A.11 and [20, Proposition E.3] that there exists a unique extremal solution  $V^{C+\varepsilon \hat{C}}$  for  $(C + \varepsilon \hat{C})$  and  $V^{C+\varepsilon \hat{C}}$  is increasing in  $\varepsilon$ . It is the minimal supersolution by [20, Corollary E.4]. Hence, by minimality,

$V^{C+\varepsilon\hat{C}} \leq \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(I, X + \varepsilon\hat{X}, Y, \Phi + \varepsilon\hat{\Phi}) < 0$ . In particular,  $V_0^{C+\varepsilon\hat{C}} \leq \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(0, X_0 + \varepsilon\hat{X}_0, y, \Phi_0 + \varepsilon\hat{\Phi}_0)$ , and letting  $\varepsilon \downarrow 0$  and using Proposition A.5(c), we obtain

$$\lim_{\varepsilon \rightarrow 0} V_0^{C+\varepsilon\hat{C}} \leq V^{\gamma^\uparrow, \gamma^\downarrow}(0, x, y, \phi). \quad (\text{A.27})$$

Let  $V^* = \lim_{\varepsilon \rightarrow 0} V^{C+\varepsilon\hat{C}}$ . Then  $V_0^* \leq \hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(0, X_0 + \varepsilon\hat{X}_0, y, \Phi_0 + \varepsilon\hat{\Phi}_0)$ . Consequently, since  $g_{EZ}$  is increasing in its last two arguments, and  $C + \varepsilon\hat{C}$  and  $V^{C+\varepsilon\hat{C}}$  are increasing in  $\varepsilon$ ,  $g_{EZ}(\cdot, C + \varepsilon\hat{C}, V^{C+\varepsilon\hat{C}})$  is increasing in  $\varepsilon$ . Applying the Monotone Convergence Theorem for conditional expectations yields

$$V_t^* = \lim_{\varepsilon \rightarrow 0} V_t^{C+\varepsilon\hat{C}} = \lim_{\varepsilon \rightarrow 0} \mathbb{E} \left[ \int_t^\infty g_{EZ}(s, C_s + \varepsilon\hat{C}_s, V_s^{C+\varepsilon\hat{C}}) ds \middle| \mathcal{F}_t \right] = \mathbb{E} \left[ \int_t^\infty g_{EZ}(s, C_s, V_s^*) ds \middle| \mathcal{F}_t \right].$$

Therefore,  $V^*$  is a solution associated to  $C$ . Moreover, for each  $t \geq 0$ ,  $V_t^* < 0$   $\mathbb{P}$ -a.s. by the fact that  $V_t^{C+\varepsilon\hat{C}} < 0$   $\mathbb{P}$ -a.s. for each  $\varepsilon$  and  $V^{C+\varepsilon\hat{C}}$  is increasing in  $\varepsilon$ . Thus,  $V^*$  is proper. It therefore agrees with the unique proper solution  $V^C$  and via (A.27) we obtain (A.26).

Next, assume  $R < 1$ . For  $\nu > 0$  and let  $C_t^\nu := C_t \wedge \nu^{-1}e^{-t}\hat{C}_t$ . Then  $\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(I, X + \varepsilon\hat{X}, Y, \Phi + \varepsilon\hat{\Phi})$  is a supersolution for  $C + \varepsilon\hat{C}$  by Proposition 5.17, and a fortiori a supersolution for  $C^\nu$ .

Moreover, there exists an extremal solution  $V^{C^\nu}$  associated to  $C^\nu$  which is decreasing in  $\nu$  by [20, Proposition E.3]. Furthermore, since  $C$  is right-continuous,  $C^\nu$  is right-continuous. The extremal solution  $V^{C^\nu}$  associated to  $C^\nu$  is therefore proper by [20, Theorem 4.5 and Remark 5.7].

Furthermore, using that  $(C_t^\nu)^{1-R} \leq \nu^{R-1}e^{(R-1)t}\hat{C}_t^{1-R}$ , we obtain

$$\begin{aligned} (1-R)V^{\gamma^\uparrow, \gamma^\downarrow}(t, X_t + \varepsilon\hat{X}_t, Y, \Phi_t + \varepsilon\hat{\Phi}_t) &\geq e^{-\delta\theta t}(\nu\hat{C}_t)^{1-R}n_\xi(\hat{Q}_t)^{-\theta} \\ &\geq \nu^{2(1-R)}n_\xi(Q_t)^{-\theta}e^{(1-R-\delta\theta)t}(C_t^\nu)^{1-R} \\ &\geq \nu^{2(1-R)}D^{-\theta}e^{(1-R-\delta\theta)t}(C_t^\nu)^{1-R}, \end{aligned}$$

where  $D := m(q_*(\xi))$  is an upper bound for  $n_\xi$ .

Now using that  $V^{C^\nu}$  is uniformly integrable by [20, Remark 2.2], the comparison result [20, Corollary B.6], gives  $\hat{V}^{\gamma^\uparrow, \gamma^\downarrow}(I, X + \varepsilon\hat{X}, Y, \Phi + \varepsilon\hat{\Phi}) \geq V^{C^\nu}$ . In particular,  $V^{\gamma^\uparrow, \gamma^\downarrow}(0, X_0 + \varepsilon\hat{X}_0, y, \Phi_0 + \varepsilon\hat{\Phi}_0) \geq V_0^{C^\nu}$ , and letting  $\varepsilon, \nu \downarrow 0$  and using Proposition A.6(b), we obtain

$$\lim_{\varepsilon \rightarrow 0} V_0^{C^\nu} \leq V^{\gamma^\uparrow, \gamma^\downarrow}(0, x, y, \phi). \quad (\text{A.28})$$

Finally, set  $V^* = \lim_{\nu \rightarrow 0} V^{C^\nu}$ . Using that  $\frac{c^{1-S}}{1-S}((1-R)v)^\rho$  is increasing both in  $c$  and  $v$  (recall  $\rho > 0$ ), monotone convergence gives, for  $t \geq 0$

$$V_t^* = \lim_{\nu \rightarrow \infty} \mathbb{E} \left[ \int_t^\infty e^{-\delta s} \frac{(C_s^\nu)^{1-S}}{1-S} ((1-R)V_s^{C^\nu})^\rho ds \middle| \mathcal{F}_t \right] = \mathbb{E} \left[ \int_t^\infty e^{-\delta s} \frac{C_s^{1-S}}{1-S} ((1-R)V_s^*)^\rho ds \middle| \mathcal{F}_t \right].$$

Hence,  $V^*$  is a solution for  $C$ . It is a proper solution since for each  $t \geq 0$ ,  $V_t^* > 0$  if  $V_t^{C^\nu} > 0$  for some  $\nu$  and  $V_t^{C^\nu} > 0$  on  $\{J^{(C^\nu)^{1-R}} > 0\} = \{J^{C^{1-R}} > 0\}$  up to null sets by the fact that each  $V^{C^\nu}$  is proper and  $\hat{C}$  is strictly positive. Therefore,  $V^*$  must agree with the unique proper solution  $V^C$  associated to  $C$ , and via (A.28) we obtain (A.26).  $\square$

## A.5 The ill-posed case

*Proof of Proposition 5.19.* Define  $\bar{\xi}$  as in (5.1).

(a) We only provide an argument for the case  $S \leq R < 1$ ; the argument for the case  $R < S < 1$  is very similar and hence omitted. We distinguish three cases.

**Case 1:  $m(0) \leq 0$ .** Without loss of generality we may assume that  $x > 0$  and  $\phi = 0$ , since otherwise one can liquidate the entire risky position at  $t = 0$  and end up with a strictly positive pure-cash position, which lies in the interior of the solvency region. Consider a sequence of investment-consumption strategies  $(\Phi^n, C^n)_{n \in \mathbb{N}}$ , where  $\Phi^n \equiv 0$  and  $C^n = \eta_n Z^n$  where  $(\eta_n)_{n \in \mathbb{N}}$  is a decreasing sequence of positive real numbers converging to  $-\frac{S}{1-S}m(0) \geq 0$  and  $Z^n := Z^{x, \Phi^n, C^n}$ ; see (3.1). It is easy to verify that  $C^n \in \mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}(x, \phi, y)$  for each  $n$ . Then by the calculations in Section 3.1, the (unique) utility process  $V^{C^n} = (V_t^{C^n})_{t \geq 0}$  associated with  $C^n$  satisfies

$$V_0^{C^n} = \left( \frac{\eta_n^{1-S}}{H(0, \eta_n)} \right)^\theta \frac{x^{1-R}}{1-R} = \left( \frac{\eta_n^{1-S}}{S(m(0) + \frac{1-S}{S}\eta_n)} \right)^\theta \frac{x^{1-R}}{1-R},$$

where  $H(\cdot, \cdot)$  is defined in (3.5). Now letting  $n \rightarrow \infty$ , the claim follows.

**Case 2:  $m(0) > 0$  and  $m(1) \leq 0$ .** Without loss of generality we may assume that  $x = 0$  and  $\phi > 0$ , since otherwise at  $t = 0$  one can invest all the current wealth into the risky asset and end up with a pure-stock position, which lies in the interior of the solvency region. Consider a sequence of investment-consumption strategies  $(\Phi^n, C^n)_{n \in \mathbb{N}}$ , where  $\Phi^n = \frac{Z^n}{Y}$  and  $C^n = \frac{\eta_n}{\gamma^\downarrow} Z^n$  where  $(\eta_n)_{n \in \mathbb{N}}$  is a decreasing sequence of positive real numbers converging to  $-\frac{S}{1-S}m(1) \geq 0$  and  $Z^n := Z^{x, \Phi^n, C^n}$ ; see (3.1). It is easy to verify that  $C^n \in \mathcal{C}_{\gamma^\downarrow, \gamma^\downarrow}(x, \phi, y)$  for each  $n$ . Then by similar calculations as in Section 3.1, the (unique) utility process  $V^{C^n} = (V_t^{C^n})_{t \geq 0}$  associated with  $C^n$  satisfies

$$V_0^{C^n} = \left( \frac{\eta_n^{1-S}}{H(1, \eta_n)} \right)^\theta \frac{\left( \frac{x}{\gamma^\downarrow} \right)^{1-R}}{1-R} = \left( \frac{\eta_n^{1-S}}{S(m(1) + \frac{1-S}{S}\eta_n)} \right)^\theta \frac{\left( \frac{x}{\gamma^\downarrow} \right)^{1-R}}{1-R},$$

where  $H(\cdot, \cdot)$  is defined in (3.5). Now letting  $n \rightarrow \infty$ , the claim follows.

**Case 3:  $m(0) > 0$ ,  $m(1) > 0$ ,  $m(q_M) < 0$  and  $\xi \leq \bar{\xi}$ .** Consider a decreasing sequence  $(\xi_n)_{n \in \mathbb{N}}$  with  $\xi_n \downarrow \bar{\xi}$ . For each  $n \in \mathbb{N}$ , set  $\gamma_n^\uparrow := \xi_n/\gamma^\downarrow$  and denote by  $(\Phi^n, C^n)$  the optimal strategy from Theorem 5.13 corresponding to the (well-posed) problem with transaction cost  $\gamma_n^\uparrow$  on purchases and  $\gamma^\downarrow$  on sales. Note that each  $C^n$  is  $(\gamma^\uparrow, \gamma^\downarrow)$ -attainable as  $\mathcal{C}_{\gamma_n^\uparrow, \gamma^\downarrow}(x, \phi, y) \subseteq \mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}(x, \phi, y)$ . Since  $\underline{x} := x + \phi^+ y \frac{1}{\gamma^\downarrow} - \phi^- y \gamma_n^\uparrow > 0$  we have that for large enough  $n$ ,  $x + \phi^+ y \frac{1}{\gamma^\downarrow} - \phi^- y \gamma_n^\uparrow > \frac{x}{2} > 0$  and then using Theorem 5.15 together with the lower bound  $x + \phi y \bar{\kappa}_{\gamma_n^\uparrow, \gamma^\downarrow} \left( \bar{q}_{\gamma_n^\uparrow, \gamma^\downarrow} \left( \frac{\phi y}{x+\phi y} \right) \right) \geq x + \phi^+ y \frac{1}{\gamma^\downarrow} - \phi^- y \gamma_n^\uparrow \geq \frac{x}{2}$  gives

$$V_0^{C^n} = \frac{\left( x + \phi y \bar{\kappa}_{\gamma_n^\uparrow, \gamma^\downarrow} \left( \bar{q}_{\gamma_n^\uparrow, \gamma^\downarrow} \left( \frac{\phi y}{x+\phi y} \right) \right) \right)^{1-R}}{1-R} \bar{n}_{\xi_n} \left( \bar{q}_{\gamma_n^\uparrow, \gamma^\downarrow} \left( \frac{\phi y}{x+\phi y} \right) \right)^{-\theta S} \geq \frac{\underline{x}^{1-R}}{(1-R)2^{1-R}} (\|\bar{n}_{\xi_n}\|_\infty)^{-\theta S}$$

Using that  $\lim_{n \rightarrow \infty} \|\bar{n}_{\xi_n}\|_\infty = 0$  by Corollary A.3, the result follows.

(b) We only provide an argument for the case  $S \geq R < 1$ ; the argument for the case  $R > S > 1$  is very similar and hence omitted. We distinguish two cases.

**Case 1:  $m(q_M) \leq 0$ .** Since  $\mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}(x, \phi, y) \subseteq \mathcal{C}_0(x + \phi, y)$ , the claim follows from the fact that for each  $C \in \mathcal{C}_0(x + \phi, y)$ , we have  $V_0^C = -\infty$  by [19, Corollary 8.2].

**Case 2:  $m(q_M) > 0$  and  $\xi \geq \bar{\xi}$ .** Note that the finiteness of transaction costs implies  $\bar{\xi} < \infty$  while  $m(q_M) > 0$  implies  $\bar{\xi} > 1$ . Let  $(\xi_n)_{n \in \mathbb{N}}$  be an increasing sequence with  $\xi_n \uparrow \bar{\xi}$ . For each  $n \in \mathbb{N}$  set  $\gamma_n^\uparrow = \xi_n/\gamma^\downarrow$ . By Theorem 5.18, the problem with transaction costs  $(\gamma_n^\uparrow, \gamma^\downarrow)$  is well-posed for every  $n \in \mathbb{N}$  with value function given by (5.13). Fix  $C \in \mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}(x, \phi, y)$ . We have the upper bound  $x + \phi y \bar{\kappa}_{\gamma_n^\uparrow, \gamma^\downarrow} \left( \bar{q}_{\gamma_n^\uparrow, \gamma^\downarrow} \left( \frac{\phi y}{x+\phi y} \right) \right) \leq \xi x + \phi^+ y \gamma_n^\uparrow - \phi^- y \frac{1}{\gamma^\downarrow} \leq \xi x + \phi^+ y \gamma^\uparrow - \phi^- y \frac{1}{\gamma^\downarrow} = \xi \underline{x}$  and then, since

$\mathcal{C}_{\gamma^\uparrow, \gamma^\downarrow}(x, \phi, y) \subset \mathcal{C}_{\gamma_n^\uparrow, \gamma_n^\downarrow}(x, \phi, y)$  by Remark 4.3, Theorem 5.15 gives

$$\begin{aligned} V_0^C &\leq \sup_{C \in \mathcal{C}_{\gamma_n^\uparrow, \gamma_n^\downarrow}(x, \phi, y)} V_0^C = \frac{\left( x + \phi y \bar{\kappa}_{\gamma_n^\uparrow, \gamma^\downarrow} \left( \bar{q}_{\gamma_n^\uparrow, \gamma^\downarrow} \left( \frac{\phi y}{x + \phi y} \right) \right) \right)^{1-R}}{1-R} \bar{n}_{\xi_n} \left( \bar{q}_{\gamma_n^\uparrow, \gamma^\downarrow} \left( \frac{\phi y}{x + \phi y} \right) \right)^{-\theta S} \\ &\leq \frac{(\xi x)^{1-R}}{1-R} (\|\bar{n}_{\xi_n}\|_\infty)^{-\theta S}. \end{aligned}$$

By Corollary A.3 we have that  $\lim_{n \rightarrow \infty} \|\bar{n}_{\xi_n}\|_\infty = 0$  and the result follows.  $\square$

## B Proofs for small transaction costs

*Proof of Proposition 7.1.* We consider the case that  $q_M = \frac{\lambda}{\sigma R} \in (0, \infty) \setminus \{1\}$ ; the case  $q_M < 0$  can be treated in a similar way. The case  $q_M = 1$  is broadly similar, but the leading terms are of different order. See Hobson et al. [22] for a derivation of the expansion for the case  $q_M = 1$  in the additive setting.

By Definition 5.1, the assumptions that the problem (4.3) is well-posed for arbitrarily small transaction cost implies that  $m_M = m(q_M) \geq 0$  if  $R < 1$  and  $m(q_M) > 0$  if  $R > 1$ .

Recall  $n_{(z)}^+$  solves  $n' = O(q, n)$  subject to  $n_{(z)}^+(z) = m(z)$ . (Here  $z \in (0, q_M) \cap \{q : m(q) > 0\}$ .) Then  $\zeta^+ = \zeta^+(z) > q_M$  is such that  $n_{(z)}^+(\zeta^+) = m(\zeta^+)$  and moreover  $(1-R)n_{(z)}^+ > (1-R)m$  on  $(z, \zeta^+) \setminus \{1\}$ . Further,  $\Sigma^+(z) = \exp \left( \int_z^{\zeta^+(z)} \frac{1}{q(1-q)} \frac{n_{(z)}^+(q) - m(q)}{\ell(q) - n_{(z)}^+(q)} dq \right)$ .

For  $u \in (0, q_M) \cap \{q : m(q) > 0\}$  introduce  $N_u : [-u, \infty) \rightarrow \mathbb{R}^+$  defined by  $N_u(v) = n_{(q_M-u)}^+(q_M+v) - m_M$ . Introduce also  $M$  and  $L$  given by  $M(v) = m(q_M+v) - m_M$  and  $L(v) = \ell(q_M+v) - m_M$ . Note that  $M(v) = \frac{R(1-S)}{2S} \sigma^2 v^2 = RDv^2$  where  $D = \frac{(1-S)}{2S} \sigma^2$ .

Define  $G_u : [-1, \infty)$  by  $G_u(v) = \frac{1}{D}(N_u(uv) - M(uv))$ . (The normalisation  $\frac{1}{D}$  is largely cosmetic but will help highlight the point at which the elasticity of intertemporal consumption  $S$  enters the expansion.) Then  $G_u(-1) = 0$ . Further, for  $v \geq -1$ ,

$$\begin{aligned} G'_u(v) &= \frac{1}{D} u O(q_M + uv, DG_u(v) + M(uv) + m_M) - \frac{1}{D} u M'(uv) \\ &= -\frac{2}{\sigma^2} u \frac{D(G_u(v) + Ru^2 v^2) + m_M}{1 - q_M - vu} \frac{G_u(v)}{(q_M(1 - q_M) + (1 - 2q_M)uv - u^2 v^2 - G_u(v))} - 2Rvu^2 \end{aligned}$$

We look for a solution to (B.1) of the form  $G_u(v) = \sum_{k \geq 0} u^k g_k(v)$  defined for  $v \geq -1$ . We find that the zeroth order term  $g_0$  satisfies

$$g'_0(v) = 0 \quad g_0(-1) = 0$$

and conclude that  $g_0(v) = 0$ . Similarly, using that  $g_0 \equiv 0$  we find

$$g'_1(v) = 0 \quad g_1(-1) = 0$$

so that again  $g_1 \equiv 0$ . The next three terms (clearly, the expansion method can be continued to arbitrary order) solve

$$\begin{aligned} g'_2(v) &= -2Rv & g_2(-1) &= 0 \\ g'_3(v) &= -\frac{2m_M}{\sigma^2 q_M (1 - q_M)^2} g_2(v) & g_3(-1) &= 0 \\ g'_4(v) &= -\frac{2m_M}{\sigma^2 q_M (1 - q_M)^2} \left( g_3(v) + v g_2(v) \frac{(3q_M - 1)}{q_M(1 - q_M)} \right) & g_4(-1) &= 0 \end{aligned}$$

and we find

$$\begin{aligned} g_2(v) &= R(1 - v^2) \\ g_3(v) &= -A(1 + v)^2(2 - v) \\ g_4(v) &= B_1(1 + v)^3(3 - v) + B_2(1 - v^2)^2 \end{aligned}$$

where

$$\begin{aligned} A &= \frac{2R}{3\sigma^2} \frac{m_M}{q_M(1-q_M)^2} \\ B_1 &= \frac{R}{3\sigma^4} \frac{m_M^2}{q_M^2(1-q_M)^4} \\ B_2 &= \frac{R}{2\sigma^2} \frac{m_M(3q_M-1)}{q_M^2(1-q_M)^3} \end{aligned}$$

Note that if we truncate the expansion at the second order term (i.e. take only the non-zero first term of the expansion) we find  $G_u(v) \approx u^2(1-v^2)R$  so that  $N_u(uv) \approx u^2RD$  which is independent of  $v$ . It follows that for  $z$  close to  $q_M$ ,  $n_{(z)}^+(q) = m(z)$  and then  $\zeta^+(z) = 2q_M - z$ , and to leading order (in the shadow fraction of wealth co-ordinates) the no-transaction region is symmetric about the Merton ratio.

The expansion for  $G_u$  takes the form

$$G_u(v) = R(1-v^2)u^2 - A(1+v)^2(2-v)u^3 + [B_1(1+v)^3(3-v) + B_2(1-v^2)^2]u^4 + \mathcal{O}(u^5),$$

and if  $\zeta_{(u)} = \inf\{v > -1 : G_u(v) = 0\}$  then  $\zeta_{(u)}$  is approximately 1. We look for a root of the form  $\zeta_{(u)} = 1 - \zeta_1 u + \zeta_2 u^2 + \mathcal{O}(u^3)$ . Then from  $G_u(\zeta_{(u)}) = 0$  we find

$$0 = (2R\zeta_1 - 4A)u^3 + (-2R\zeta_2 - \zeta_1^2 R + 16B_1)u^4 + O(u^5)$$

It follows that

$$\zeta_1 = \frac{2A}{R} = \frac{4m_M}{3\sigma^2 q_M(1-q_M)^2}, \quad \zeta_2 = \zeta_1^2.$$

We can now fix  $u = q_M - q_*$  via the identity

$$\begin{aligned} \ln(\xi) &= u \int_{-1}^{1-\zeta_1 u + \zeta_1^2 u^2 + \mathcal{O}(u^3)} dv \frac{1}{(q_M + uv)(1 - q_M - uv)} \frac{G_u(v)}{(q_M + uv)(1 - q_M - vu) - G_u(v)} \\ &= \int_{-1}^{1-\zeta_1 u + \zeta_1^2 u^2 + \mathcal{O}(u^3)} \frac{dv}{q_M^2(1-q_M)^2} u \left(1 + \frac{uv}{q_M}\right)^{-1} \left(1 - \frac{uv}{1-q_M}\right)^{-1} G_u(v) \\ &\quad \times \left( \left(1 + \frac{uv}{q_M}\right) \left(1 - \frac{uv}{1-q_M}\right) - \frac{G_u(v)}{q_M(1-q_M)} \right)^{-1} \\ &= \int_{-1}^{1-\zeta_1 u + \zeta_1^2 u^2 + \mathcal{O}(u^3)} \frac{dv}{q_M^2(1-q_M)^2} \\ &\quad \times \left( g_2(v)u^3 + \left( g_3(v) + \frac{2(2q_M-1)}{q_M(1-q_M)} vg_2(v) \right) u^4 + \Psi(v)u^5 + \mathcal{O}(u^6) \right) \end{aligned}$$

where

$$\Psi(v) = g_4(v) + \frac{2(2q_M-1)}{q_M(1-q_M)} vg_3(v) + \frac{1}{q_M(1-q_M)} g_2(v)^2 + \frac{3-10q_M+10q_M^2}{q_M^2(1-q_M)^2} v^2 g_2(v)$$

Now  $\int_{-1}^z (1-v^2)dv = \frac{1}{3}(z+1)^2(2-z)$  so that

$$\int_{-1}^{1-\zeta_1 u + \zeta_1^2 u^2 + \mathcal{O}(u^3)} g_2(v)dv = \frac{R}{3}(4 - 3\zeta_1^2 u^2) + \mathcal{O}(u^3)$$

Similarly,  $\int_{-1}^z (1+v)^2(2-v)dv = \frac{1}{4}(z+1)^3(3-z)$  and  $\int_{-1}^z v(1-v^2)dv = -\frac{1}{4}(z^2-1)^2$  so that

$$\begin{aligned} &\int_{-1}^{1-\zeta_1 u + \mathcal{O}(u^2)} \left( g_3(v) + \frac{2(2q_M-1)}{q_M(1-q_M)} vg_2(v) \right) dv \\ &= -4A(1 - \zeta_1 u) + \mathcal{O}(u^2) \\ &= -\frac{8R}{3\sigma^2} \frac{m_M}{q_M(1-q_M)^2} + \left( \frac{8R}{3\sigma^2} \frac{m_M}{q_M(1-q_M)^2} \right) \zeta_1 u + \mathcal{O}(u^2) \end{aligned}$$

Finally,  $\int_{-1}^1 (1+v)^3(3-v)dv = \frac{48}{5}$ ,  $\int_{-1}^1 (1-v^2)^2dv = \frac{16}{15}$ ,  $\int_{-1}^1 v(1+v)^2(2-v)dv = \frac{8}{5}$ ,  $\int_{-1}^1 v^2(1-v^2)dv = \frac{4}{15}$  so that

$$\begin{aligned} \int_{-1}^{1+\mathcal{O}(u)} \Psi(v)dv &= \frac{16}{5} \frac{R}{\sigma^4} \frac{m_M^2}{q_M^2(1-q_M)^4} + \frac{8R}{15\sigma^2} \frac{m_M(3q_M-1)}{q_M^2(1-q_M)^3} - \frac{32R}{15\sigma^2} \frac{m_M(2q_M-1)}{q_M^2(1-q_M)^3} \\ &\quad + \frac{16}{15} \frac{1}{q_M(1-q_M)} R^2 + \frac{4}{15} \frac{(3-10q_M+10q_M^2)}{q_M^2(1-q_M)^2} R + \mathcal{O}(u) \\ &= \frac{16R}{5\sigma^4} \frac{m_M^2}{q_M^2(1-q_M)^4} - \frac{8R}{15\sigma^2} m_M \frac{(5q_M-3)}{q_M^2(1-q_M)^3} \\ &\quad + \frac{16}{15} \frac{1}{q_M(1-q_M)} R^2 + \frac{4}{15} \frac{(3-10q_M+10q_M^2)}{q_M^2(1-q_M)^2} R + \mathcal{O}(u) \end{aligned}$$

Putting this all together, writing  $\xi = 1 + \varepsilon$ , and then using the expression for  $\zeta_1$ ,

$$\begin{aligned} &\ln(1+\varepsilon) \left( = \varepsilon - \frac{\varepsilon^2}{2} + \mathcal{O}(\varepsilon^3) = \varepsilon^\uparrow + \varepsilon^\downarrow - \frac{(\varepsilon^\uparrow)^2 - (\varepsilon^\downarrow)^2}{2} + \mathcal{O}(\varepsilon^3) \right) \\ &= \frac{4R}{3q_M^2(1-q_M)^2} u^3 - \frac{8R}{3\sigma^2} \frac{m_M}{q_M^3(1-q_M)^4} u^4 \\ &\quad + \frac{R}{q_M^2(1-q_M)^2} \left( -\zeta_1^2 + \frac{8}{3\sigma^2} \frac{m_M \zeta_1}{q_M(1-q_M)^2} + \frac{4}{15} \frac{(3-10q_M+10q_M^2)}{q_M^2(1-q_M)^2} + \frac{16}{15} R \frac{1}{q_M(1-q_M)} \right. \\ &\quad \left. \frac{16}{5\sigma^4} \frac{m_M^2}{q_M^2(1-q_M)^4} - \frac{8}{15\sigma^2} m_M \frac{(5q_M-3)}{q_M^2(1-q_M)^3} \right) + \mathcal{O}(u^6) \\ &= \frac{4R}{3q_M^2(1-q_M)^2} u^3 - \frac{2R}{q_M^2(1-q_M)^2} \zeta_1 u^4 \\ &\quad + \frac{R}{q_M^2(1-q_M)^2} \left( \frac{14}{5} \zeta_1^2 - \frac{2}{5} \frac{(5q_M-3)}{q_M(1-q_M)} \zeta_1 + \frac{4}{15} \frac{(3-10q_M+10q_M^2)}{q_M^2(1-q_M)^2} + \frac{16}{15} R \frac{1}{q_M(1-q_M)} \right) u^5 + \mathcal{O}(u^6) \\ &=: C_3 u^3 + C_4 u^4 + C_5 u^5 + \mathcal{O}(u^6) \end{aligned}$$

Comparing the two sides of the equation we see that for the equation to balance we must have that  $u = q_M - q_*(1+\varepsilon) = \mathcal{O}(\varepsilon^{1/3})$ . We look for an expressing for  $q_*$  of the form

$$u = \Delta^{\frac{1}{3}} \varepsilon^{\frac{1}{3}} + \Sigma \Delta^{\frac{2}{3}} \varepsilon^{\frac{2}{3}} + \Psi \Delta \varepsilon + \mathcal{O}\left(\varepsilon^{\frac{4}{3}}\right)$$

or equivalently

$$q_*(1+\varepsilon) = q_M - \Delta^{\frac{1}{3}} \varepsilon^{\frac{1}{3}} - \Sigma \Delta^{\frac{2}{3}} \varepsilon^{\frac{2}{3}} - \Psi \Delta \varepsilon + \mathcal{O}\left(\varepsilon^{\frac{4}{3}}\right)$$

Then

$$\varepsilon = C_3 \left( \Delta \varepsilon + 3\Sigma(\Delta \varepsilon)^{4/3} + 3(\Sigma^2 + \Psi)(\Delta \varepsilon)^{5/3} \right) + C_4 \left( (\Delta \varepsilon)^{4/3} + 4\Sigma(\Delta \varepsilon)^{5/3} \right) + C_5 (\Delta \varepsilon)^{5/3} + \mathcal{O}(\varepsilon^2)$$

The constants  $\Sigma$ ,  $\Delta$  and  $\Psi$  can be fixed by the identities  $1 = C_3 \Delta$ ,  $0 = C_4 + 3\Sigma C_3$  and  $0 = 3(\Sigma^2 + \Psi)C_3 + 4\Sigma C_4 + C_5$ . We find

$$\begin{aligned} \Delta &= \frac{1}{C_3} = \frac{3q_M^2(1-q_M)^2}{4R} \\ \Sigma &= -\frac{C_4}{3C_3} = \frac{2m_M}{3\sigma^2 q_M(1-q_M)^2} = \frac{\zeta_1}{2} \\ \Psi &= -\frac{C_5 + 4\Sigma C_4}{3C_3} - \Sigma^2 = 3\Sigma^2 - \frac{C_5}{3C_3} \\ &= \frac{1}{5} \Sigma^2 + \frac{1}{5} \frac{(5q_M-3)}{q_M(1-q_M)} \Sigma - \frac{(3-10q_M+10q_M^2) + 4Rq_M(1-q_M)}{15q_M^2(1-q_M)^2} \end{aligned}$$

Further,

$$\begin{aligned} q^*(1+\varepsilon) - q_M &= u(\varepsilon)\zeta_{(u(\varepsilon))} = u(\varepsilon)(1 - \zeta_1 u(\varepsilon) + \zeta_1^2 u(\varepsilon)^2) + \mathcal{O}(u(\varepsilon)^4) \\ &= \Delta^{\frac{1}{3}}\varepsilon^{\frac{1}{3}} + \Sigma\Delta^{\frac{2}{3}}\varepsilon^{\frac{2}{3}} + \Psi\Delta\varepsilon - \zeta_1 \left( \Delta^{\frac{2}{3}}\varepsilon^{\frac{2}{3}} + 2\Sigma\Delta\varepsilon \right) + \zeta_1^2\Delta\varepsilon + \mathcal{O}(\varepsilon^{4/3}) \end{aligned}$$

It follows that

$$q^*(1+\varepsilon) = q_M + \Delta^{\frac{1}{3}}\varepsilon^{\frac{1}{3}} - \Sigma\Delta^{\frac{2}{3}}\varepsilon^{\frac{2}{3}} + \Psi\Delta\varepsilon + \mathcal{O}\left(\varepsilon^{\frac{4}{3}}\right) \quad \square$$

*Proof of Corollary 7.3.* Setting  $\varepsilon := (1 + \varepsilon^\uparrow)(1 + \varepsilon^\downarrow) - 1$  and using (5.10) and (5.11), we obtain

$$\begin{aligned} p_*(1 + \varepsilon^\uparrow, 1 + \varepsilon^\downarrow) &= \frac{q_*(1 + \varepsilon)}{1 + \varepsilon^\uparrow(1 - q_*(1 + \varepsilon))} = q_*(1 + \varepsilon) - \varepsilon^\uparrow q_*(1 + \varepsilon)(1 - q_*(1 + \varepsilon)) + \mathcal{O}(\varepsilon^2) \\ p^*(1 + \varepsilon^\uparrow, 1 + \varepsilon^\downarrow) &= \frac{(1 + \varepsilon^\downarrow)q^*(1 + \varepsilon)}{1 + \varepsilon^\downarrow q^*(1 + \varepsilon)} = q^*(1 + \varepsilon) + \varepsilon^\downarrow q^*(1 + \varepsilon)(1 - q^*(1 + \varepsilon)) + \mathcal{O}(\varepsilon^2) \end{aligned}$$

Now the result follows from Proposition 7.5.  $\square$

*Proof of Proposition 7.5.* Setting  $\gamma^\uparrow := 1 + \varepsilon^\uparrow$ ,  $\gamma^\downarrow := 1 + \varepsilon^\downarrow$ ,  $\xi := \gamma^\uparrow\gamma^\downarrow$  and  $\varepsilon := \xi - 1$ , and looking at the expression (4.15) for the optimal consumption, we need to consider how  $\kappa(q(p)) = \kappa_{\gamma^\uparrow, \gamma^\downarrow}(q(p))$  and  $n_\xi(q(p))$  vary with  $\varepsilon^\uparrow$  and  $\varepsilon^\downarrow$  as functions of  $p$ .

Set  $u = u(\varepsilon) = q_M - q_*(1 + \varepsilon)$ . Recalling that  $\kappa_{\gamma^\uparrow, \gamma^\downarrow} : [q_*(\xi), q^*(\xi)] \rightarrow [\frac{1}{\gamma^\downarrow}, \gamma^\uparrow]$  is monotonically decreasing and hence  $\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q(p)) = 1 + \mathcal{O}(\varepsilon)$ , set

$$\tilde{\kappa}_u(v) = \kappa_{\gamma^\uparrow, \gamma^\downarrow}(q_M + uv) = \kappa_{\gamma^\uparrow, \gamma^\downarrow}(q_M + (q_M - q_*(1 + \varepsilon))v).$$

Then  $\tilde{\kappa}_u(-1) = \kappa_{\gamma^\uparrow, \gamma^\downarrow}(q_*(\xi)) = \gamma^\uparrow$ . Moreover,

$$\begin{aligned} \frac{\tilde{\kappa}'_u(v)}{\tilde{\kappa}_u(v)} &= u \frac{\kappa'_{\gamma^\uparrow, \gamma^\downarrow}(q_M + uv)}{\kappa_{\gamma^\uparrow, \gamma^\downarrow}(q_M + uv)} \\ &= -\frac{u}{(q_M + uv)(1 - q_M - uv)} \frac{(n_{(q_*(1+\varepsilon))}^+(q_M + uv) - m(q_M + uv))}{(\ell(q_M + uv) - n_{(q_*(1+\varepsilon))}^+(q_M + uv))} \\ &= -\frac{u}{(q_M + uv)(1 - q_M - uv)} \frac{N_u(uv) - M(uv)}{L(uv) - N_u(uv)} \\ &= -\frac{u}{(q_M + uv)(1 - q_M - uv)} \frac{DG_u(v)}{D(q_M + uv)(1 - q_M - uv) - DG_u(v)} \\ &= -\frac{u^3 g_2(v)}{q_M^2(1 - q_M)^2} + \mathcal{O}(u^4) \end{aligned}$$

We find, with  $h_3(v) = (1 + v)^2(2 - v) = 3 \int_{-1}^v (1 - w^2)dw$ , for  $v \in [-1, \zeta(u)]$ ,

$$\begin{aligned} \tilde{\kappa}_u(v) &= \gamma^\uparrow \exp\left(-\frac{u^3}{q_M^2(1 - q_M)^2} \int_{-1}^v g_2(w)dw + \mathcal{O}(u^4)\right) \\ &= \gamma^\uparrow \exp\left(-\frac{R\varepsilon\Delta}{3q_M^2(1 - q_M)^2} h_3(v) + \mathcal{O}(\varepsilon^{4/3})\right) \\ &= \gamma^\uparrow \exp\left(-\frac{\varepsilon}{4} h_3(v) + \mathcal{O}(\varepsilon^{4/3})\right) \end{aligned}$$

Note that  $\kappa_u(1) = \gamma^\uparrow \exp(-\varepsilon) + \mathcal{O}(\varepsilon^{4/3}) = \gamma^\uparrow \frac{1}{1+\varepsilon} + \mathcal{O}(\varepsilon^{4/3}) = \frac{1}{\gamma^\downarrow} + \mathcal{O}(\varepsilon^{4/3})$ .

Since  $\kappa_{\gamma\uparrow,\gamma\downarrow}(q(p)) = 1 + \mathcal{O}(\varepsilon)$  it follows that  $p(q) = \frac{q}{1-q(1-\kappa_{\gamma\uparrow,\gamma\downarrow}(q))} = q + \mathcal{O}(\varepsilon)$  which can be inverted to give  $q(p) = p + \mathcal{O}(\varepsilon)$ . Further,  $q$  and  $p$  both lie in intervals of width of order  $\varepsilon^{1/3}$  around  $q_M$ , so  $q(p) = q_M + \mathcal{O}(\varepsilon^{1/3})$ . It follows that, with  $p = \frac{y\phi}{x+y\phi}$  and  $v$  as shorthand for  $\frac{p-q_M}{u}$  in later lines,

$$\begin{aligned}
x + y\phi\kappa_{\gamma\uparrow,\gamma\downarrow}(q(p)) &= (x + y\phi)(1 + p(\kappa_{\gamma\uparrow,\gamma\downarrow}(q(p)) - 1)) \\
&= (x + y\phi)\left(1 + q_M(\kappa_{\gamma\uparrow,\gamma\downarrow}(p) - 1) + \mathcal{O}(\varepsilon^{4/3})\right) \\
&= (x + y\phi)\left(1 + q_M(\tilde{\kappa}_u(v) - 1) + \mathcal{O}(\varepsilon^{4/3})\right) \\
&= (x + y\phi)\left(1 + q_M\left(\gamma^\uparrow \exp\left(-\frac{\varepsilon}{4}h_3(v)\right) - 1\right) + \mathcal{O}(\varepsilon^{4/3})\right) \\
&= (x + y\phi)\left(1 + q_M\left((1 + \varepsilon^\uparrow)\left(1 - \frac{\varepsilon^\uparrow + \varepsilon^\downarrow}{4}h_3(v)\right) - 1\right) + \mathcal{O}(\varepsilon^{4/3})\right) \\
&= (x + y\phi)\left(1 + q_M\left(1 - \frac{h_3(v)}{4}\right)\varepsilon^\uparrow - q_M\frac{h_3(v)}{4}\varepsilon^\downarrow\right) + \mathcal{O}(\varepsilon^{4/3})
\end{aligned}$$

Now we turn to the expansion of  $n_\xi(q(p))$ . We have  $n_\xi(q(p)) = n_{(q_M-u(\varepsilon))}^+(q(p))$ . Then, using  $g_3 = -Ah_3$  we have

$$\begin{aligned}
n_{(q_M-u(\varepsilon))}^+(q(p)) &= m_M + N_{u(\varepsilon)}(q(p) - q_M) \\
&= m_M + M(q(p) - q_M) + DG_{u(\varepsilon)}\left(\frac{(q(p) - q_M)}{u(\varepsilon)}\right) \\
&= m_M + RD(q(p) - q_M)^2 + Du(\varepsilon)^2 g_2\left(\frac{(q(p) - q_M)}{u(\varepsilon)}\right) + Du(\varepsilon)^3 g_3\left(\frac{(q(p) - q_M)}{u(\varepsilon)}\right) + \mathcal{O}(\varepsilon^{4/3}) \\
&= m_M + \frac{R(1-S)}{S} \frac{\sigma^2}{2} u(\varepsilon)^2 - DAu(\varepsilon)^3 h_3\left(\frac{(q(p) - q_M)}{u(\varepsilon)}\right) + \mathcal{O}(\varepsilon^{4/3}) \\
&= m_M \left(1 + \frac{R(1-S)}{S} \frac{\sigma^2}{2m_M} \left(\Delta^{2/3}\varepsilon^{2/3} + 2\Delta\Sigma\varepsilon\right) - \frac{DA\Delta}{m_M} \varepsilon h_3\left(\frac{p - q_M}{\Delta^{1/3}\varepsilon^{1/3}}\right)\right) + \mathcal{O}(\varepsilon^{4/3})
\end{aligned}$$

Note that  $\frac{DA\Delta}{m_M}$  can be written more simply as  $\frac{q_M}{4} \frac{1-S}{S}$  and  $\frac{R\Delta\Sigma\sigma^2}{m_M} = \frac{q_M}{2}$ . Hence this expression simplifies to

$$n_{(q_M-u(\varepsilon))}^+(q(p)) = m_M \left(1 + \frac{R(1-S)}{S} \frac{\sigma^2}{2m_M} \Delta^{2/3}\varepsilon^{2/3} + q_M \frac{(1-S)}{2S} \left(1 - \frac{1}{2}h_3\left(\frac{p - q_M}{\Delta^{1/3}\varepsilon^{1/3}}\right)\right)\varepsilon\right) + \mathcal{O}(\varepsilon^{4/3}).$$

Putting this all together and using (4.15),

$$\begin{aligned}
C_{1+\varepsilon^\uparrow, 1+\varepsilon^\downarrow}^*(t, x, \phi, y) &= (x + y\phi)m_M \left(1 + q_M \left(1 - \frac{1}{4}h_3\left(\frac{p - q_M}{\Delta^{1/3}\varepsilon^{1/3}}\right)\right)\varepsilon^\uparrow - q_M \frac{1}{4}h_3\left(\frac{p - q_M}{\Delta^{1/3}\varepsilon^{1/3}}\right)\varepsilon^\downarrow\right) \\
&\quad \times \left(1 + \frac{R(1-S)}{S} \frac{\sigma^2}{2m_M} \Delta^{2/3}\varepsilon^{2/3} + q_M \frac{(1-S)}{2S} \left(1 - \frac{1}{2}h_3\left(\frac{p - q_M}{\Delta^{1/3}\varepsilon^{1/3}}\right)\right)\varepsilon\right) + \mathcal{O}(\varepsilon^{4/3}) \\
&= (x + y\phi)m_M \left(1 + \frac{R(1-S)}{S} \frac{\sigma^2}{2m_M} \Delta^{2/3}(\varepsilon^\uparrow + \varepsilon^\downarrow)^{2/3} + \frac{q_M}{2S} \left(1 - \frac{1}{2}h_3\left(\frac{p - q_M}{\Delta^{1/3}\varepsilon^{1/3}}\right)\right)(\varepsilon^\uparrow + \varepsilon^\downarrow) + \frac{q_M}{2}(\varepsilon^\uparrow - \varepsilon^\downarrow)\right) \\
&\quad + \mathcal{O}(\varepsilon^{4/3})
\end{aligned}$$

□

## C Proofs for the results on comparative statics

### C.1 Comparative statics in $S$

*Proof of Proposition 8.1.* We will establish the ordering of the optimal purchase/sale boundaries in the  $q$ -coordinate, which can then be extended to that of  $\hat{p}_*^*$  and  $\tilde{p}_*^*$  by the monotonicity of the Möbius transformation over the relevant domain.

Let

$$\bar{m}(q) := \frac{S}{1-S} \left( m(q) - \frac{\delta}{S} \right) = -r - \lambda\sigma q + \frac{R\sigma^2}{2} q^2$$

which is independent of  $S$ . Similarly, let

$$\bar{\ell}(q) := \frac{S}{1-S} \left( \ell(q) - \frac{\delta}{S} \right) = -r - \left( \lambda\sigma - \frac{\sigma^2}{2} \right) q - (1-R) \frac{\sigma^2}{2} q^2$$

which is also independent of  $S$ .

Suppose  $(n(\cdot), q_*, q^*)$  is the solution to the free boundary problem in Proposition 5.3. Then  $n'(q) = O(q, n(q))$  on  $q \in [q_*, q^*]$ ,  $n(q_*) = m(q_*)$ ,  $n(q^*) = m(q^*)$ , and

$$\ln \gamma^\uparrow \gamma^\downarrow = \ln \xi = \int_{q_*}^{q^*} \frac{1}{q(1-q)} \frac{n(q) - m(q)}{\ell(q) - n(q)} dq.$$

Define  $\bar{n}(q) := \frac{S}{1-S} (n(q) - \frac{\delta}{S})$ . We have

$$\begin{aligned} \bar{n}'(q) &= \frac{S}{1-S} n'(q) = \frac{S}{1-S} O(q, n(q)) = \frac{S}{1-S} O(q, (1-S)\bar{n}(q)/S + \delta/S) \\ &= \frac{\bar{n}(q)}{1-q} \frac{\bar{n}(q) - \bar{m}(q)}{\bar{\ell}(q) - \bar{n}(q)} - \frac{1}{S} \frac{1}{1-q} \frac{\bar{n}(q) - \bar{m}(q)}{\bar{\ell}(q) - \bar{n}(q)} (\bar{n}(q) + \delta) \\ &= \bar{O}(q, \bar{n}(q)) \end{aligned}$$

over  $q \in [q_*, q^*]$ , where

$$\bar{O}(q, \bar{n}) := \frac{\bar{n}}{1-q} \frac{\bar{n} - \bar{m}(q)}{\bar{\ell}(q) - \bar{n}} - \frac{1}{S} \frac{1}{1-q} \frac{\bar{n} - \bar{m}(q)}{\bar{\ell}(q) - \bar{n}} (\bar{n} + \delta).$$

The boundary conditions at  $q = q_*$  can be restated as  $\bar{n}(q_*) = \bar{m}(q_*)$  and  $\bar{n}(q^*) = \bar{m}(q^*)$ . The integral constraint becomes

$$\ln \xi = \int_{q_*}^{q^*} \frac{1}{q(1-q)} \frac{\bar{n}(q) - \bar{m}(q)}{\bar{\ell}(q) - \bar{n}(q)} dq.$$

Finally, away from 1, the solution trajectory  $n = (n(q))_{q_* \leq q \leq q^*}$  always satisfies  $\frac{1}{1-q} \frac{n(q) - m(q)}{\ell(q) - n(q)} \operatorname{sgn}(q_M) \geq 0$  and in turn  $\frac{1}{1-q} \frac{\bar{n}(q) - \bar{m}(q)}{\bar{\ell}(q) - \bar{n}(q)} \operatorname{sgn}(q_M) \geq 0$ .

(a) If (8.1) holds, then

$$\bar{n}(q) + \delta = \frac{S(n(q) - \delta)}{1-S} > \frac{S(m_M - \delta)}{1-S} = \delta - r - \frac{\lambda^2}{2R} \geq 0$$

as  $\operatorname{sgn}(1-S)n(q) > \operatorname{sgn}(1-S)m_M$ . If we view  $\bar{O}(q, \bar{n}) = \bar{O}(q, \bar{n}; S)$  as a function of  $S$ , then clearly  $\bar{O}$  is strictly increasing (resp. decreasing) in  $S$  over  $(q, \bar{n})$  on which  $\frac{1}{1-q} \frac{\bar{n} - \bar{m}(q)}{\bar{\ell}(q) - \bar{n}} (\bar{n} + \delta)$  is positive (resp. negative). Hence given (8.1), we must have  $\operatorname{sgn}(q_M)\bar{O}(q, \bar{n}(q); \hat{S}) > \operatorname{sgn}(q_M)\bar{O}(q, \bar{n}(q); \tilde{S})$  over  $q \in (q_*, q^*) \setminus \{1\}$  for  $\hat{S} > \tilde{S}$ .

Recall the definitions of  $(n_{(z)}^+(q))_{q \in [z, \zeta^+(z)]}$ ,  $\zeta^+(z)$  and  $\Sigma^+(z)$  as well as the relevant domain of  $z$  defined in Proposition A.1 or A.2. Define  $(a_{(z)}^+(q))_{q \in [z, \zeta^+(z)]}$  via  $a_{(z)}^+(q) := \frac{S}{1-S}(n_{(z)}^+(q) - \frac{\delta}{S})$ . By construction,

$$\frac{d}{dq} a_{(z)}^+(q) = \bar{O}(q, a_{(z)}^+(q); S), \quad q \in [z, \zeta^+(z)],$$

$a_{(z)}^+(z) = \bar{m}(z)$ ,  $a_{(z)}^+(\zeta^+(z)) = \bar{m}(\zeta^+(z))$ , and

$$\Sigma^+(z) = \exp \left( \int_z^{\zeta^+(z)} \frac{1}{q(1-q)} \frac{a_{(z)}^+(q) - \bar{m}(q)}{\bar{\ell}(q) - a_{(z)}^+(q)} dq \right).$$

Let  $\hat{n}_{(z)}^+$  and  $\tilde{n}_{(z)}^+$  be the two versions of  $n_{(z)}^+$  under  $S = \hat{S}$  and  $S = \tilde{S} < \hat{S}$  respectively while all other model parameters are fixed. We will apply the same overhead notations to  $\zeta^+(z)$ ,  $a_{(z)}^+$  and  $\Sigma^+$  which are defined relative to the underlying parameters  $\hat{S}$  and  $\tilde{S}$ . Since  $\bar{m}(\cdot)$  does not depend on  $S$ ,  $\hat{a}_{(z)}^+(z) = \bar{m}(z) = \tilde{a}_{(z)}^+(z)$ . Also  $(\hat{a}_{(z)}^+)'(z) = 0 = (\tilde{a}_{(z)}^+)'(z)$  for  $z \neq 1$ .

Suppose that  $q_M \in (0, 1)$ . We first show that  $\hat{a}_{(z)}^+(q) > \tilde{a}_{(z)}^+(q)$  near  $q = z$  for any  $z \in (0, q_M)$ . Using the ODE satisfied by  $a_{(z)}^+$  and L'Hospital rule, we obtain

$$\lim_{q \downarrow z} \frac{a_{(z)}^+(q) - \bar{m}(z)}{(q - z)^2} = \frac{\bar{m}'(z)}{\sigma^2 z(1-z)^2} \left[ \frac{1}{S}(\bar{m}(z) + \delta) - \bar{m}(z) \right],$$

where the leading factor is negative and in turn the whole expression is strictly increasing in  $S$  as  $\bar{m}(z) + \delta = \frac{S(\bar{m}(z) + \delta)}{1-S} > \frac{S(\bar{m}_M + \delta)}{1-S} = \delta - r - \frac{\lambda^2}{2R} \geq 0$  for all  $z \neq q_M$  under condition (8.1). Hence  $\hat{a}_{(z)}^+(q)$  must be initially strictly larger than  $\tilde{a}_{(z)}^+(q)$ . Together with the facts that  $\bar{O}(q, a; \hat{S}) > \bar{O}(q, a; \tilde{S})$  along any solution trajectory (excluding the end points), we conclude  $\hat{\zeta}^+(z) \geq \tilde{\zeta}^+(z)$  and  $\hat{a}_{(z)}^+(q) \geq \tilde{a}_{(z)}^+(q)$  at least up to  $q = \tilde{\zeta}^+(z) < 1$ . Then

$$\begin{aligned} \hat{\Sigma}^+(z) &= \exp \left( \int_z^{\hat{\zeta}^+(z)} \frac{1}{q(1-q)} \frac{\hat{a}_{(z)}^+(q) - \bar{m}(q)}{\bar{\ell}(q) - \hat{a}_{(z)}^+(q)} dq \right) \geq \exp \left( \int_z^{\zeta^+(z)} \frac{1}{q(1-q)} \frac{\hat{a}_{(z)}^+(q) - \bar{m}(q)}{\bar{\ell}(q) - \hat{a}_{(z)}^+(q)} dq \right) \\ &\geq \exp \left( \int_z^{\tilde{\zeta}^+(z)} \frac{1}{q(1-q)} \frac{\tilde{a}_{(z)}^+(q) - \bar{m}(q)}{\bar{\ell}(q) - \tilde{a}_{(z)}^+(q)} dq \right) \\ &= \tilde{\Sigma}^+(z), \end{aligned}$$

and in turn  $\hat{q}_* = (\hat{\Sigma}^+)^{-1}(\xi) \geq (\tilde{\Sigma}^+)^{-1}(\xi) = \tilde{q}_*$  since  $\Sigma^+(\cdot)$  is decreasing.

The arguments are similar in the case of  $q_M > 1$  except some extra care is needed around  $q = 1$ . For any  $z \in (0, 1)$ ,  $a_{(z)}^+(1) = \bar{m}(1)$  and  $(a_{(z)}^+)'(1) = \bar{m}'(1)$  such that the ordering of  $\hat{a}_{(z)}^+(q)$  and  $\tilde{a}_{(z)}^+(q)$  is not immediately clear on  $q > 1$ . But again we can deduce via L'Hospital rule that

$$\lim_{q \rightarrow 1} \frac{a_{(z)}^+(q) - \bar{m}(q)}{(1-q)^2} = -\frac{\sigma^2}{2} \frac{\bar{m}'(1)}{\frac{1}{S}(\bar{m}(1) + \delta) - \bar{m}(1)},$$

which is positive and strictly increasing in  $S$  under condition (8.1). Hence we conclude  $\hat{a}_{(z)}^+(q) > \tilde{a}_{(z)}^+(q)$  in the neighbourhood of  $q = 1$ .

The results for  $q^*$  can be established similarly by parameterising the solutions to the family of ODEs via their right boundary points. Results for the case of  $q_M < 0$  can be derived using similar ideas.

(b) If instead the parameters are such that (8.2) holds, then the argument in part (a) does not work as  $\bar{n}(q) + \delta$  may change sign and hence the monotonicity of  $\bar{O}(q, \bar{n}; S)$  in  $S$  is ambiguous. But suppose  $q_M > 0$  and the problems under  $S \in \{\hat{S}, \tilde{S}\}$  are well-posed for all sufficiently small transaction costs. Then there exists  $\delta > 0$  such that for all  $z \in (q_M - \delta, q_M)$ ,  $n_{(z)}^+(q)$  is well-defined and is strictly positive over  $q \in [z, \zeta^+(z)]$ . Moreover, since the family of solutions  $(n_{(z)}^+(\cdot))_z$  is monotonic in  $z$  with  $n_{(z)}^+(q) \rightarrow m_M$  as  $z \rightarrow q_M$ , one can choose  $\Delta$  such that  $|n_{(z)}^+(q) - m_M| < \omega$  and  $|m(q) - m_M| < \omega$  on  $q \in [z, \zeta^+(z)]$  for all  $z \in (q_M - \Delta, q_M)$  where  $\omega > 0$  is arbitrary. Then

$$\begin{aligned} a_{(z)}^+(q) + \delta &= \frac{S(n_{(z)}^+(q) - \delta)}{1 - S} \leq \frac{S(m_M + \operatorname{sgn}(1 - S)\omega - \delta)}{1 - S} \\ &= \delta - r - \frac{\lambda^2}{2R} + \frac{S}{|1 - S|}\omega < 0 \end{aligned}$$

and similarly  $\bar{m}(q) + \delta < 0$  if  $\omega$  is chosen to be sufficiently small. In such case,  $\bar{O}(q, a_{(z)}^+(q); \hat{S}) < \bar{O}(q, a_{(z)}^+(q); \tilde{S})$  along the solution trajectory  $(a_{(z)}^+(q))_{z < q < \zeta^+(z)}$ , and that  $\hat{a}_{(z)}^+(q) < \tilde{a}_{(z)}^+(q)$  near  $q = z$  (and near  $q = 1$  as well if  $q_M > 1$ ). Following similar arguments in part (a), we can conclude  $\hat{\zeta}^+(z) \leq \zeta^+(z)$  and  $\hat{a}_{(z)}^+(q) \leq \tilde{a}_{(z)}^+(q)$  up to  $q = \hat{\zeta}^+(z)$ , and in turn  $\hat{\Sigma}^+(z) \leq \tilde{\Sigma}^+(z)$  for  $z \in (q_M - \delta, q_M)$ . Thus there exists  $\varepsilon := \hat{\Sigma}^+(q_M - \Delta) - 1 > 0$  such that  $\hat{q}_* = (\hat{\Sigma}^+)^{-1}(\xi) \leq (\tilde{\Sigma}^+)^{-1}(\xi) = \tilde{q}_*$  for all  $\xi \leq 1 + \varepsilon$ .

The proof for  $q^*$  is similar upon parameterising the solutions to the family of ODEs via the right boundary points. The analysis can be analogously extended to  $q_M < 0$ .

□

## C.2 Comparative statics in $\mathbf{R}$

Now we turn to comparative statics in  $R$ . For the case  $R, S > 1$  it is useful to introduce an auxiliary function. For  $q \in (0, \infty) \setminus \{1, 2S/(S+1)\}$ , define  $c = c(q)$  as the solution to the equation  $O(q, c)/c = 2/q$ . Straightforward algebra leads to

$$c(q) = \ell(q) + \frac{(S-1)q}{2S-(S+1)q}(m(q) - \ell(q)) = m(q) - \frac{2S(1-q)}{2S-(S+1)}(m(q) - \ell(q)). \quad (\text{C.1})$$

The definition of  $c$  is extended to  $q = 0$  and  $q = 1$  via  $c(0) := \lim_{q \downarrow 0} c(q) = \ell(0) = m(0)$  and  $c(1) := \lim_{q \rightarrow 1} c(q) = m(1) = \ell(1)$ . Note that  $c(q)$  is continuous on  $q \in [0, 2S/(S+1)] \supset (0, 1)$  and we have  $c'(0) = \ell'(0)$  and  $c'(1) = m'(1)$ . We also have

$$\begin{cases} \ell(q) < c(q) < m(q), & q \in (0, 1); \\ c(q) < m(q) < \ell(q), & q \in \left(1, \frac{2S}{S+1}\right); \\ m(q) < \ell(q) < c(q), & q \in \left(\frac{2S}{S+1}, \infty\right). \end{cases}$$

Also, on  $(0, 1) \times (0, \infty)$ ,  $\frac{O(q, z)}{z} > \frac{2}{q}$  for  $z \in (\ell(q), c(q))$  and  $\frac{O(q, z)}{z} < \frac{2}{q}$  for  $z \in (c(q), m(q))$ . Further, on  $(1, \frac{2S}{S+1}) \times (0, m(q))$ ,  $\frac{O(q, z)}{z} < \frac{2}{q}$  on  $(0, c(q)^+)$  and  $\frac{O(q, z)}{z} > \frac{2}{q}$  on  $(c(q)^+, m(q))$ . Finally, on  $(\frac{2S}{S+1}, \infty)$ ,  $\frac{O(q, z)}{z} < \frac{2}{q}$  on  $(0, m(q))$ .

**Lemma C.1.** *Suppose  $\lambda > 0$  and the problem is well-posed such that  $(n(\cdot), q_*, q^*)$  is the solution to the free boundary problem in Proposition 5.3. Then any of the following is a sufficient condition for  $2n(q) - qn'(q) > 0$  on  $q \in [q_*, q^*]$ .*

- (a)  $S < 1$ .
- (b)  $S > 1$ ,  $m(0) > 0$ , and either (recall the definition of  $c(q)$  in (C.1)):
  - (i)  $2c(q) - qc'(q) > 0$  on  $(0, 1)$  if  $q_M < 1$  or on  $(0, \frac{2S}{S+1})$  if  $q_M > 1$ ,
  - (ii)  $R \geq 2$ .

A further sufficient condition for (b)(i) is that  $\ell'(0) \geq 0$ , or equivalently  $(1 - S)(\sigma - 2\lambda) \geq 0$ .

*Proof.* (a) The result under  $R, S < 1$  is trivial since the solution  $n(q)$  is positive and decreasing (under  $\lambda > 0$ ).

- (b) We first show that  $\ell'(0) \geq 0$  is sufficient for condition (b)(i) to hold. By direct calculation,

$$\begin{aligned}
 2c(q) - qc'(q) &= 2 \left[ \frac{m(q) - \ell(q)}{\frac{2S}{S-1} \frac{1-q}{q} + 1} + \ell(q) \right] - q \left[ \frac{\left( \frac{2S}{S-1} \frac{1-q}{q} + 1 \right) (m'(q) - \ell'(q)) + (m(q) - \ell(q)) \frac{2S}{S-1} q^{-2}}{\left( \frac{2S}{S-1} \frac{1-q}{q} + 1 \right)^2} + \ell'(q) \right] \\
 &= \frac{\sigma^2}{2} \frac{(S-1)q}{S \left( \frac{2S}{S-1} \frac{1-q}{q} + 1 \right)^2} + 2\ell(q) - q\ell'(q) \\
 &= \frac{\sigma^2}{2} \frac{(S-1)q}{S \left( \frac{2S}{S-1} \frac{1-q}{q} + 1 \right)^2} + \frac{2(\delta + r(S-1))}{S} + \frac{S-1}{S} \left( \lambda\sigma - \frac{\sigma^2}{2} \right) q \\
 &= \frac{\sigma^2}{2} \frac{(S-1)q}{S \left( \frac{2S}{S-1} \frac{1-q}{q} + 1 \right)^2} + 2m(0) + \ell'(0)q > 0
 \end{aligned}$$

on  $q > 0$  away from  $q = 2S/(S+1)$ .

- (i) Suppose  $q_M \in (0, 1)$  such that  $[q_*, q^*] \subset (0, 1)$ . By the given condition,  $O(q, c(q)) = 2c(q)/q > c'(q)$  for  $q \in [q_*, q^*]$ . If  $n_{(z)}^+ = (n_{(z)}^+(q))_{q \geq z}$  is the solution to the initial value problem  $n' = O(q, n)$  with  $n_{(z)}^+(z) = m(z)$ ,  $n_{(z)}^+(q)$  can then only upcross  $c(q)$  from below on  $q \in (0, 1)$ . Since  $c(0) = \ell(0) = m(0)$  and the family of solutions  $(n_{(z)}^+(\cdot))_z$  is increasing in  $z$ , we must have  $n(q) = n_{(q_*)}^+(q) > n_{(0)}^+(q) \geq c(q) > \ell(q)$  for all  $q \in [q_*, q^*] \subset (0, 1)$ . Finally,  $\Xi(q, n) := \frac{O(q, n)}{n} = \frac{S-1}{S} \frac{1}{1-q} \frac{n-m(q)}{\ell(q)-n}$  is continuously decreasing in  $n \in (\ell(q), \infty)$  for any fixed  $q \in (0, 1)$  and therefore

$$\frac{n'(q)}{n(q)} = \Xi(q, n(q)) < \Xi(q, c(q)) = \frac{2}{q}.$$

Suppose  $q_M > 1$  and the transaction costs are sufficiently large such that  $q_* < 1 < q_M < q^*$ . Using the same argument as in the case of  $q_M \in (0, 1)$ , we can still conclude  $n(q) \geq c(q)$  and in turn  $n'(q)/n(q) \leq 2/q$  up to  $q < 1$ . But since  $n$  now passes through the singular point  $(1, m(1))$  at which we have  $c(1) = m(1) = n(1)$  and  $c'(1) = m'(1) = n'(1)$ , the ordering of  $c(q)$  and  $n(q)$  is ambiguous to the right of  $q = 1$ . We thus need to first argue that  $n(q) > c(q)$  in the neighbourhood of  $q = 1$ .

Write  $f(q) := n(q) - c(q)$ . Then

$$f'(q) = n'(q) - c'(q) = \frac{S-1}{S} \frac{f(q) + c(q)}{1-q} \frac{f(q) + c(q) - m(q)}{\ell(q) - f(q) - c(q)} - c'(q)$$

and a rearrangement of terms yields

$$\frac{f(q)}{(q-1)^2} = \frac{S(f'(q) + c'(q))}{(S-1)(f(q) + c(q))} \left( \frac{\ell(q) - c(q)}{1-q} - \frac{f(q)}{1-q} \right) - \frac{c(q) - m(q)}{(1-q)^2}.$$

Since it is known that  $n(1) = m(1) = c(1)$  and  $n'(1) = m'(1) = c'(1)$ , we have  $f(1) = f'(1) = 0$  and  $\lim_{q \rightarrow 1} f(q)/(1-q) = 0$ . Upon taking limit, we get

$$\begin{aligned} \lim_{q \rightarrow 1} \frac{f(q)}{(q-1)^2} &= \frac{S(m'(1) - \ell'(1))m'(1)}{(S-1)m(1)} - \frac{c''(1) - m''(1)}{2} \\ &= \frac{S(m'(1) - \ell'(1))m'(1)}{(S-1)m(1)} - \frac{2S(m'(1) - \ell'(1))}{S-1} \\ &= \frac{S(\ell'(1) - m'(1))}{S-1} \left[ 2 - \frac{m'(1)}{m(1)} \right] > 0 \end{aligned}$$

because  $\ell'(1) > m'(1)$  and  $2m(1) > m'(1)$  under the assumption of  $m(0) > 0$ ,  $q_M > 1$  and  $S > 1$ . Hence we conclude  $n(q) > c(q)$  near  $q = 1$ .

Using similar arguments in the case of  $q_M \in (0, 1)$ , we must have  $\ell(q) > m(q) > n(q) > c(q)$  on  $1 < q < q^* \wedge \frac{2S}{S+1}$ . As  $\Xi(q, n) = \frac{O(q, n)}{n} = \frac{S-1}{S} \frac{1}{1-q} \frac{n-m(q)}{\ell(q)-n}$  is continuously decreasing in  $n \in (0, \ell(q))$  for any fixed  $q > 1$ , we have  $\frac{n'(q)}{n(q)} = \frac{O(q, n(q))}{n(q)} < \frac{O(q, c(q))}{c(q)} = \frac{2}{q}$  for  $1 < q < q^* \wedge \frac{2S}{S+1}$ . If  $q^* > \frac{2S}{S+1}$ , then  $c(q) > \ell(q) > m(q) > n(q) > 0$  on  $q \in (\frac{2S}{S+1}, q^*)$ . Note that  $\Xi(q, n)$  is discontinuous at  $n = \ell(q)$ . But since  $H(q, n)$  is still locally continuous and decreasing on  $n \in [0, \ell(q))$  and  $n \in (\ell(q), \infty)$  respectively under fixed  $q > 1$ , we have

$$\frac{2}{q} = \Xi(q, c(q)) > \lim_{n \rightarrow \infty} \Xi(q, n) = \frac{S-1}{S} \frac{1}{q-1} \geq \frac{S-1}{S} \frac{1}{q-1} \frac{m(q)}{\ell(q)} = \Xi(q, 0) > \Xi(q, n(q)) = \frac{n'(q)}{n(q)}$$

on  $\frac{2S}{S+1} < q < q^*$ . By continuity of  $n(q)$  and  $n'(q)$ , the above inequality holds at  $q = \frac{2S}{S+1}$  and  $q = 1$  as well.

Finally, the case of  $1 \leq q_* < q_M < q^*$  under small transaction costs can be handled similarly using the fact that the solution  $n(q)$  in such case is bounded below by  $n_{(1)}^+(q)$ .

(ii) If  $m(0) > 0$  and  $\ell'(0) \geq 0$  then we are in Case b(i), so we only need to consider  $m(0) > 0$  and  $\ell'(0) < 0$ . Since  $\ell$  is concave we have  $m(1) = \ell(1) < \ell(0) = m(0)$  so necessarily  $q_M \in (0, 1)$  and  $[q_*, q^*] \subset (0, 1)$ . In what follows we only consider functions over  $(0, 1)$ .

Define  $A(q) := qm(q) + (1-q)\ell(q)$ .  $A$  is chosen so that  $A(z) = c(z)$  and  $A'(z) = c'(z)$  for  $z \in \{0, 1\}$ , and to be a simple function. It is easily seen that  $A(q) > c(q)$ . So, if we can show that  $n$  is always above  $A$  it will follow that  $2n(q) - qn'(q) > 0$  on  $(q_*, q^*)$  as required.

Note that  $O(q, A(q)) = \frac{(S-1)}{S} \frac{A(q)}{q}$  so that if  $n = A$  then  $n' > A'$  is equivalent to  $\frac{(S-1)}{S} \frac{A(q)}{q} > A'$ . So, a sufficient condition for  $n > A$  on  $(0, 1)$  is (here we use  $m(q) - \ell(q) = Dq(1-q)$  where  $D = \frac{(S-1)}{S} \frac{\sigma^2}{2}$ )

$$\begin{aligned} \frac{(S-1)}{S} \frac{A(q)}{q} &> \ell'(q) + (m(q) - \ell(q)) + q(m'(q) - \ell'(q)) \\ \Leftrightarrow (S-1)(\ell(q) + q(m(q) - \ell(q))) &> Sq\ell'(q) + Sq(m(q) - \ell(q)) + Sq^2(m'(q) - \ell'(q)) \\ \Leftrightarrow (S-1)\ell(q) &> Sq\ell'(q) + q(m(q) - \ell(q)) + Sq^2(m'(q) - \ell'(q)) \\ \Leftrightarrow (S-1)\ell(0) + (S-1)q\ell'(0) + (S-1)\frac{q^2}{2}\ell''(0) &> Sq(\ell'(0) + q\ell''(0)) + Dq^2(1-q) + SDq^2(1-2q) \\ \Leftrightarrow (S-1)\ell(0) &> q\ell'(0) + \frac{q^2}{2}\ell''(0)(2S - (S-1)) + Dq^2((1-q) + S(1-2q)) \\ \Leftrightarrow (S-1)\ell(0) &> q\ell'(0) + \frac{q^2}{2}\ell''(0)(S+1) + Dq^2((S+1) - q(1+2S)). \end{aligned}$$

Thus we only need to establish  $(S - 1)\ell(0) > \sup_{q \in (0,1)} \varphi(q)$  where  $\varphi(q)$  is a cubic function defined via

$$\varphi(q) := q\ell'(0) + \frac{\sigma^2}{2S}(S+1)(S-1)(2-R)q^2 - \frac{\sigma^2}{2S}(S-1)(1+2S)q^3,$$

where we use that  $\frac{1}{2}\ell''(0) + D = \frac{\sigma^2}{2S}(S-1)(2-R)$ . Note that  $\varphi(0) = 0 < (S-1)\ell(0)$  and  $\varphi'(0) = \ell'(0) < 0$ . Hence we are done if the coefficient of  $q^2$  is negative, which is equivalent to  $R \geq 2$ .

□

*Proof of Proposition 8.3.* We will show that  $\hat{p}_* \leq \tilde{p}_*$  and  $\hat{p}^* \leq \tilde{p}^*$  under the stated conditions in Lemma C.1. The last part of the proposition will then follow trivially.

In what follows, we write

$$\begin{aligned} m(q) &= m(q; R) = \frac{\delta - r(1-S)}{S} - \frac{1-S}{S}\lambda\sigma q + R\frac{1-S}{S}\frac{\sigma^2}{2}q^2 = m(q; 0) + R\frac{1-S}{S}\frac{\sigma^2}{2}q^2, \\ \ell(q) &= \ell(q; R) = \frac{\delta - r(1-S)}{S} - \frac{1-S}{S}\left(\lambda\sigma - \frac{\sigma^2}{2}\right)q - (1-R)\frac{1-S}{S}\frac{\sigma^2}{2}q^2 = \ell(q; 0) + R\frac{1-S}{S}\frac{\sigma^2}{2}q^2, \end{aligned}$$

where  $m(q; 0)$  and  $\ell(q; 0)$  do not depend on  $R$ . Note that  $\ell(q) - m(q) = \ell(q; 0) - m(q; 0)$ .

Let  $f(q) := \frac{n(q) - m(q)}{\ell(q) - n(q)}$  where  $n$  is a solution to  $n' = O(q, n)$ . Then

$$\begin{aligned} f'(q) &= \frac{(\ell(q) - n(q))(n'(q) - m'(q)) - (n(q) - m(q))(\ell'(q) - n'(q))}{(\ell(q) - n(q))^2} \\ &= \frac{1 + f(q)}{\ell(q) - m(q)} \left[ \frac{S-1}{S(1-q)} f(q)(m(q) + \ell(q)f(q)) - m'(q) - \ell'(q)f(q) \right] \\ &= \frac{1 + f(q)}{\ell(q; 0) - m(q; 0)} \left[ \frac{S-1}{S(1-q)} f(q) \left( m(q; 0) + R\frac{1-S}{S}\frac{\sigma^2}{2}q^2 + \left( \ell(q; 0) + R\frac{1-S}{S}\frac{\sigma^2}{2}q^2 \right) f(q) \right) \right. \\ &\quad \left. - m'(q; 0) - R\frac{1-S}{S}\sigma^2q - \left( \ell'(q; 0) + R\frac{1-S}{S}\sigma^2q \right) f(q) \right] \\ &= \Psi(q, f(q); 0) - \frac{R(1+f(q))^2}{S(1-q)^2} ((1-S)qf(q) + 2S(1-q)) =: \Psi(q, f(q); R) \end{aligned}$$

for some function  $\Psi(q, f; R)$  with  $\Psi(q, f; 0)$  being independent of  $R$ . In particular,  $\Psi$  depends on  $R$  linearly and we have

$$\frac{\partial}{\partial R} \Psi(q, f; R) = -\frac{(1+f)^2}{S(1-q)^2} ((1-S)qf + 2S(1-q)).$$

Along a solution trajectory of the free boundary value problem,

$$\begin{aligned} (1-S)qf(q) + 2S(1-q) &= (1-S)q \frac{n(q) - m(q)}{\ell(q) - n(q)} + 2S(1-q) \\ &= -Sq(1-q) \frac{O(q, n(q))}{n(q)} + 2S(1-q) \\ &= \frac{S(1-q)}{n(q)} (2n(q) - qn'(q)). \end{aligned}$$

Then if any condition in Lemma C.1 holds such that  $2n - qn' > 0$ , we have  $\operatorname{sgn}(\frac{\partial}{\partial R} \Psi(q, f(q); R)) = \operatorname{sgn}(q-1)$ .

Recall the definitions of  $(n_{(z)}^+(q))_{q \in [z, \zeta^+(q)]}$ ,  $\zeta^+(z)$  and  $\Sigma^+(z)$  defined in Proposition A.1 or A.2. Set  $f_{(z)}^+(q) := \frac{n_{(z)}^+(q) - m(q)}{\ell(q) - n_{(z)}^+(q)}$ . Note that  $\operatorname{sgn}(f_{(z)}^+(q)) = \operatorname{sgn}(1 - q)$ . By construction,

$$\frac{d}{dq} f_{(z)}^+(q) = \Psi(q, f_{(z)}^+(q); R), \quad q \in [z, \zeta^+(z)],$$

$f_{(z)}^+(z) = 0 = f_{(z)}^+(\zeta(z))$  and

$$\Sigma^+(z) = \exp \left( \int_z^{\zeta^+(z)} \frac{f_{(z)}^+(q)}{q(1-q)} dq \right).$$

Fix  $\tilde{R} < \hat{R}$ , and we will use the corresponding overhead notations to denote the particular quantities evaluated under  $R = \tilde{R}$  and  $R = \hat{R}$  respectively similar to what we have done in the proof of Proposition 8.1.

If  $\tilde{R} < \hat{R}$  are such that  $\tilde{q}_M \geq 1 \geq \hat{q}_M$ , then we obviously have  $\hat{q}^* \leq 1 \leq \tilde{q}^*$ . It is therefore sufficient to only consider values of  $\tilde{R}, \hat{R}$  such that  $\tilde{q}_M, \hat{q}_M < 1$  or  $\tilde{q}_M, \hat{q}_M > 1$ .

If  $\tilde{q}_M, \hat{q}_M < 1$ , then the domain of a solution to the free boundary problem must be contained within  $(0, 1)$ . Since  $\Psi(q, f_{(z)}^+(q); R)$  is strictly decreasing in  $R$  on  $q \in (0, 1)$  and  $\hat{f}_{(z)}^+(z) = 0 = \tilde{f}_{(z)}^+(z)$  with

$$(f_{(z)}^+)'(z) = -\frac{m'(z; 0) + R^{\frac{1-S}{S}}\sigma^2 z}{\frac{1-S}{S}\frac{\sigma^2}{2}z(1-z)} \quad (\text{C.2})$$

which is decreasing in  $R$  such that  $(\hat{f}_{(z)}^+)'(z) < (\tilde{f}_{(z)}^+)'(z)$  for  $z \in (0, 1)$ , we must have  $\hat{\zeta}(z) \leq \tilde{\zeta}(z) < 1$  and  $0 \leq \hat{f}_{(z)}(q) \leq \tilde{f}_{(z)}(q)$  for  $q \in [z, \hat{\zeta}^+(z)]$ . Then

$$\tilde{\Sigma}^+(z) = \exp \left( \int_z^{\tilde{\zeta}^+(z)} \frac{\tilde{f}_{(z)}^+(q)}{q(1-q)} dq \right) \geq \exp \left( \int_z^{\hat{\zeta}^+(z)} \frac{\tilde{f}_{(z)}^+(q)}{q(1-q)} dq \right) \geq \exp \left( \int_z^{\hat{\zeta}^+(z)} \frac{\hat{f}_{(z)}^+(q)}{q(1-q)} dq \right) = \hat{\Sigma}^+(z)$$

and then  $\hat{q}_* = (\hat{\Sigma}^+)^{-1}(\xi) \leq (\tilde{\Sigma}^+)^{-1}(\xi) = \tilde{q}_*$ .

Suppose  $\tilde{q}_M, \hat{q}_M > 1$ . In case of  $z > 1$  such that the domain of the solution lives on a subset of  $(1, \infty)$ ,  $\Psi(q, f_{(z)}^+(q); R)$  is now increasing in  $R$  and  $(\hat{f}_{(z)}^+)'(z) > (\tilde{f}_{(z)}^+)'(z)$  for  $z > 1$  using (C.2). Hence  $\tilde{\zeta}(z) \geq \hat{\zeta}(z) > q_M$  and  $0 \geq \hat{f}_{(z)}^+(q) \geq \tilde{f}_{(z)}^+(q)$  for  $1 < z \leq q \leq \hat{\zeta}^+(z)$ . Then we can deduce  $\tilde{\Sigma}^+(z) \geq \hat{\Sigma}^+(z)$  on  $z > 1$ . The same conclusion can be extended to  $z = 1$  upon showing that  $(\hat{f}_{(1)}^+)'(1) > (\tilde{f}_{(1)}^+)'(1)$ . It is sufficient to verify that

$$\lim_{q \rightarrow 1} \frac{f(q)}{q-1} = \frac{S}{1-S} \frac{m'(1)}{m(1)} = \frac{R\sigma^2 - \lambda\sigma}{R^{\frac{1-S}{S}}\frac{\sigma^2}{2} + \frac{\delta-r(1-S)}{S} - \frac{1-S}{S}\lambda\sigma}$$

is increasing in  $R$ . One can check that the first order derivative of this expression with respect to  $R$  is proportional to

$$\frac{\delta - r(1-S)}{S} - \frac{\lambda\sigma(1-S)}{2S} = m(0) + \frac{m'(0)}{2} = \ell(1) + \frac{1-S}{2S}R\sigma^2(q_M - 1),$$

which is strictly positive in both cases of  $S < 1$  (as  $\ell(1) > 0$  under well-posedness and we work with  $q_M > 1$ ) and  $S > 1$  (since  $m'(0) > 0$ , and  $m(0) > 0$  as a required condition).

Finally, for  $\tilde{q}_M, \hat{q}_M > 1$  and  $z < 1$  such that  $\tilde{\zeta}^+(z) = \tilde{\zeta}^+(1) \geq \hat{\zeta}^+(1) = \hat{\zeta}^+(z)$ , we have  $\tilde{f}_{(z)}^+(q) \geq \hat{f}_{(z)}^+(q) \geq 0$  on  $q \leq 1$ ,  $\tilde{f}_{(z)}^+(1) = \hat{f}_{(z)}^+(1) = 0$  and  $0 \geq \hat{f}_{(z)}^+(q) = \hat{f}_{(1)}^+(q) \geq \tilde{f}_{(1)}^+(q) = \tilde{f}_{(z)}^+(q)$  on  $1 \leq q \leq \hat{\zeta}(1)$ .

Then

$$\begin{aligned}\tilde{\Sigma}^+(z) &= \exp \left( \int_z^{\tilde{\zeta}^+(z)} \frac{\tilde{f}_{(z)}^+(q)}{q(1-q)} dq \right) = \exp \left( \int_z^1 \frac{\tilde{f}_{(z)}^+(q)}{q(1-q)} dq + \int_1^{\tilde{\zeta}^+(1)} \frac{\tilde{f}_{(z)}^+(q)}{q(1-q)} dq \right) \\ &\geq \exp \left( \int_z^1 \frac{\hat{f}_{(z)}^+(q)}{q(1-q)} dq + \int_1^{\tilde{\zeta}^+(1)} \frac{\hat{f}_{(z)}^+(q)}{q(1-q)} dq \right) = \hat{\Sigma}^+(z).\end{aligned}$$

Hence  $\tilde{\Sigma}^+(z) \geq \hat{\Sigma}^+(z)$  for all  $z$  which establishes  $\hat{q}_* \leq \tilde{q}_*$ .

The result for the sale boundary can again be established by parametrising the solutions via their right boundary points.  $\square$

## References

- [1] M. Akian, J. L. Menaldi, and A. Sulem, *On an investment-consumption model with transaction costs*, SIAM J. Control Optim. **34** (1996), no. 1, 329–364.
- [2] C. Atkinson and S. Mokkhavesa, *Multi-asset portfolio optimization with transaction cost*, Applied Mathematical Finance **11** (2004), no. 2, 95–123.
- [3] Maxim Bichuch, *Asymptotic analysis for optimal investment in finite time with transaction costs*, SIAM J. Financial Math. **3** (2012), no. 1, 433–458.
- [4] G. Chacko and L.M. Viceira, *Dynamic consumption and portfolio choice with stochastic volatility in incomplete markets*, The Review of Financial Studies **18** (2005), no. 4, 1369–1402.
- [5] X.F. Chen and M. Dai, *Characterisation of optimal strategy for multi-asset investment and consumption with transaction costs*, SIAM Journal on Financial Mathematics **4** (2013), no. 1, 857–883.
- [6] J. Choi, *Asymptotic analysis for Merton’s problem with transaction costs in power utility case*, Stochastics **86** (2014), no. 5, 803–816.
- [7] ———, *Optimal consumption and investment with liquid and illiquid assets*, Mathematical Finance **30** (2020), no. 2, 621–663.
- [8] J. Choi, M. Sirbu, and G. Zitkovic, *Shadow prices and well-posedness in the problem of optimal investment and consumption with transaction costs*, SIAM Journal on Control and Optimization **51** (2013), no. 6, 4414–4449.
- [9] G. M. Constantinides, *Capital market equilibrium with transaction costs*, Journal of political Economy **94** (1986), no. 4, 842–862.
- [10] M.H.A. Davis and A.R. Norman, *Portfolio selection with transaction costs*, Mathematics of Operations Research **15** (1990), no. 4, 676–713.
- [11] B. Dumas and E. Luciano, *An exact solution to a dynamic portfolio choice problem under transactions costs*, The Journal of Finance **46** (1991), no. 2, 577–595.
- [12] L.G. Epstein and S.E. Zin, *Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework*, Econometrica **57** (1989), no. 4, 937–969.
- [13] W. H. Fleming and H. M. Soner, *Controlled Markov processes and viscosity solutions*, second ed., Springer, New York, 2006.
- [14] S. Gerhold, P. Guasoni, J. Muhle-Karbe, and W. Schachermayer, *Transaction costs, trading volume, and the liquidity premium*, Finance Stoch. **18** (2014), no. 1, 1–37.
- [15] S. Gerhold, J. Muhle-Karbe, and W. Schachermayer, *Asymptotics and duality for the Davis and Norman problem*, Stochastics **84** (2012), no. 5-6, 625–641.
- [16] P. Guasoni and J. Muhle-Karbe, *Portfolio choice with transaction costs: a user’s guide*, Paris-Princeton Lectures on Mathematical Finance 2013, Lecture Notes in Math., vol. 2081, Springer, 2013, pp. 169–201.
- [17] M. Herdegen, D. Hobson, and J. Jerome, *An elementary approach to the Merton problem*, Mathematical Finance **31** (2021), no. 4, 1218–1239.

[18] ———, *The infinite-horizon investment-consumption problem for Epstein-Zin stochastic differential utility. I: Foundations*, Finance and Stochastics **27** (2023), no. 1, 127–158.

[19] ———, *The infinite-horizon investment-consumption problem for Epstein-Zin stochastic differential utility. II: Existence, uniqueness and verification for  $\vartheta \in (0, 1)$* , Finance and Stochastics **27** (2023), no. 1, 159–188.

[20] ———, *Proper solutions for Epstein-Zin stochastic differential utility*, ArXiV:2112.06708, 2023.

[21] D. Hobson, A. S. L. Tse, and Y. Zhu, *A multi-asset investment and consumption problem with transaction costs*, Finance and Stochastics **23** (2019), no. 3, 641–676.

[22] ———, *Optimal consumption and investment under transaction costs*, Mathematical Finance **29** (2019), no. 2, 483–506.

[23] K. Janeček and S.E. Shreve, *Asymptotic analysis for optimal investment and consumption with transaction costs*, Finance and Stochastics **18** (2004), no. 2, 181–206.

[24] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991.

[25] H. Kraft, T. Seiferling, and F.T. Seifried, *Optimal consumption and investment with Epstein-Zin recursive utility*, Finance and Stochastics **21** (2017), no. 1, 187–226.

[26] M.J.P. Magill and G.M. Constantinides, *Portfolio selection with transaction costs*, Journal of Economic Theory **13** (1976), no. 2, 245–263.

[27] R. Mehra and E.C. Prescott, *The equity premium: A puzzle*, Journal of Monetary Economics **15** (1985), no. 2, 145–161.

[28] Y. Melnyk, J. Muhle-Karbe, and F.T. Seifried, *Lifetime investment and consumption with recursive preferences and small transaction costs*, Mathematical Finance **30** (2020), no. 3, 1135–1167.

[29] R.C. Merton, *Lifetime portfolio selection under uncertainty: the continuous-time case*, The Review of Economics and Statistics **51** (1969), no. 3, 247–257.

[30] ———, *Optimum consumption and portfolio rules in a continuous-time model*, Journal of Economic Theory **3** (1971), no. 4, 373–413.

[31] H. Pham, *Continuous-time stochastic control and optimization with financial applications*, Springer-Verlag, Berlin, 2009.

[32] Dylan Possamaï, H. Mete Soner, and Nizar Touzi, *Homogenization and asymptotics for small transaction costs: the multidimensional case*, Comm. Partial Differential Equations **40** (2015), no. 11, 2005–2046.

[33] L.C.G Rogers, *Why is the effect of proportional transaction costs  $o(\delta^{2/3})$ ?*, Mathematics of Finance (G. Yin and Q. Zhang, eds.), American Mathematical Society, Providence, 2004, pp. 303–308.

[34] S. E. Shreve and H. M. Soner, *Optimal investment and consumption with transaction costs*, Annals of Applied Probability **4** (1994), no. 3, 609–692.

[35] H. Mete Soner and Nizar Touzi, *Homogenization and asymptotics for small transaction costs*, SIAM J. Control Optim. **51** (2013), no. 4, 2893–2921.

[36] M. Taksar, M. J. Klass, and D. Assaf, *A diffusion model for optimal portfolio selection in the presence of brokerage fees*, Math. Oper. Res. **13** (1988), no. 2, 277–294.

[37] H. Tanaka, *Stochastic differential equations with reflecting boundary condition in convex regions*, Hiroshima Mathematical Journal **9** (1979), no. 1, 163–177.

[38] A. S. L. Tse, *Dynamic economic decision problems under behavioural preferences and market imperfections*, Ph.D. thesis, University of Warwick, 2016.

[39] H. Xing, *Consumption-investment optimization with Epstein-Zin utility in incomplete markets*, Finance and Stochastics **21** (2017), no. 1, 227–262.