OPTIMAL CONSUMPTION AND INVESTMENT UNDER TRANSACTION COSTS

DAVID HOBSON, ALEX S.L. TSE, AND YEQI ZHU

Abstract. In this article we consider the Merton problem in a market with a single risky asset and proportional transaction costs. We give a complete solution of the problem up to the solution of a first-crossing problem for a first-order differential equation. We find that the characteristics of the solution (for example well-posedness) can be related to some simple properties of a univariate quadratic whose coefficients are functions of the parameters of the problem.

Our solution to the problem via the value function includes expressions for the boundaries of the no-transaction wedge. Using these expressions we prove a precise condition for when leverage occurs. One new and unexpected result is that when the solution to the Merton problem (without transaction costs) involves a leveraged position, and when transaction costs are large, the location of the boundary at which sales of the risky asset occur is independent of the transaction cost on purchases.

1. Introduction

In this article we consider the problem of maximizing expected utility of consumption over the infinite horizon in a financial market consisting of a riskless bond and a single risky asset. Merton (1969, 1971) considered this problem in a perfect, frictionless market and showed that the optimal strategy is to keep a constant fraction of wealth in the risky asset. Of the possible market frictions, arguably the most significant is transaction costs, and this paper adds to the growing literature on optimal consumption/investment problems with proportional transaction costs.

In the Merton setting the market is complete. Constantinides and Magill (1976) generalized the problem to the incomplete case by introducing transaction costs. They argued that in the case of a single risky asset following exponential Brownian motion and power utility the scalings of the problem should mean that there is a no-transaction wedge, and the optimal strategy should be to trade in a minimal

Date: April 20, 2018.

The authors would like to thank the referees for their careful reading of the paper and their insightful comments.

David Hobson: Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK. D.Hobson@warwick.ac.uk;

Alex S.L. Tse: Cambridge Endowment for Research in Finance, Judge Business School, University of Cambridge, Cambridge, CB2 1AG, UK. S.Tse@jbs.cam.ac.uk;

Yeqi Zhu: Credit Suisse, London, UK. (The opinions expressed in the paper are those of the author and not of Credit Suisse.) Yeqi.Zhu@credit-suisse.com.

fashion so as to keep the fraction of wealth in the risky asset within an interval I. If the initial portfolio is such that the initial fraction of wealth in the risky asset is outside this interval then the agent makes an instantaneous transaction to bring the fraction of wealth in the risky asset to the closest boundary of I. Thereafter, the agent only trades when this fraction is on the boundary of I. In (Cash wealth, Wealth in risky asset) space, the interval I becomes a no-transaction wedge.

The model and accompanying intuition was formulated precisely by Davis and Norman (1990). They used the language of stochastic control and martingales to give a rigorous description of the problem for power utility. For a certain subset of parameter combinations they gave a full description of the solution, specifying both the optimal consumption, and the optimal investment strategy. The optimal investment strategy involves a process which receives a local-time push at both boundaries of an interval, and these pushes are just sufficient to keep the process within the interval. Davis and Norman (1990) reduce the problem to solving a pair of first-order ordinary differential equations (ODEs) subject to value matching conditions at unknown free-boundaries. Their analysis was extended by Shreve and Soner (1994) to a larger set of parameter combinations using methodologies from viscosity solutions.

Both Davis and Norman (1990) and Shreve and Soner (1994) consider the value function and the primal problem. Recently, there have been a trio of papers considering the dual problem and shadow prices. Kallsen and Muhle-Karbe (2010) were the first to use the shadow-price approach in this context, but only consider logarithmic utility. Herczegh and Prokaj (2015) extend their results to power utility. The most complete treatment of the problem via the shadow-price approach is the paper of Choi et al. (2013). Choi et al. give a full analysis of the problem, covering all parameter combinations (which involve an appreciating risky asset). They reduce the problem to the solution of a free-boundary problem for a single first-order ODE. There are multiple solutions to this free-boundary problem, and the one that is wanted is the one for which the solution to the free-boundary problem satisfies an integral condition.

In this paper we revisit the consumption-investment problem with transaction costs considered by Davis and Norman (1990), Shreve and Soner (1994), Choi et al. (2013) and Herczegh and Prokaj (2015), again taking the primal approach. We show that the problem can be transformed into finding the solution of a first-order ODE which starts and ends on a given function, subject to an integral condition, as in Choi et al. (2013). The dual approach via shadow prices is very attractive conceptually, but the advance in this paper relative to Choi et al. (2013) is that it is much easier to understand the character of the solutions to our differential equation. For our solution the possible behaviors correspond to the possible shapes of a quadratic function of a single variable, whereas in Choi et al. (2013) it is necessary to consider a phase-diagram in which the behaviors depend on a pair of ellipses and/or hyperbolas. Although the two problems must be transformations of each other, our approach leads to a simpler representation of the solution. Nonetheless, many features of our characterizing ODE are to be found in the characterizing ODE of Choi et al. (2013); in particular in both cases the ODE has a singular point, and for some parameter combinations, though not all, the solution we want passes through this singular point. Our main contribution relates to the fact that our characterization of the problem involves the solution of a first-order ODE crossing the aforementioned quadratic and the fact that there is a direct interpretation of these crossing points as the sale and purchase boundaries of the no-transaction wedge. This allows us re-derive more simply many results from the literature, to give new interpretations and explanations, and to give extensions.

The remainder of this paper is structured as follows. In the next section we formulate the problem as a problem in stochastic optimal control. In Section 3 we introduce an auxiliary problem which involves finding the solution of a first order ODE which starts and ends on a quadratic, and which is subject to an integral condition. Using simple properties of the ODE, and more especially of the quadratic boundary, we describe when solutions to the auxiliary problem exist.

Section 4 contains the main financial results. The first result, Theorem 5 states a one-to-one correspondence between existence or otherwise of a solution to the auxiliary problem and well-posedness of the Merton problem with transaction costs. This result mirrors the main result of Choi et al. (2013), but our formulation is a small improvement in that we cover an extra case and we give an algebraic expression for a quantity that Choi et al. can only express as an integral. The second result, Theorem 6, connects the endpoints of the solution of the auxiliary problem to the boundaries of the no-transaction wedge.

Theorem 6 forms the cornerstone of the analysis in Section 5 in which we consider how the boundaries of the no-transaction wedge depend on the parameters. Analysis of this type seems to be new, and would be difficult under previous approaches. We show that if the expected return on the risky asset is small, then the no-transaction wedge includes the Merton line, and the no-transaction wedge gets larger as transaction costs increase. However, if the expected return increases further, then we may lose both the monotonicity property of the no-transaction wedge, and the property that the Merton line (corresponding to zero transaction costs) lies within the no-transaction region. Remarkably, although in general the locations of both the sale and purchase boundaries depend on the transaction costs on both sale and purchases, in some circumstances the sale boundary is independent of the transaction cost on purchases.

Both Choi et al. (2013) and this paper make great strides towards a complete solution of the Merton problem for general levels of the transaction cost parameters. We can also connect to the literature on the Merton problem in the small transaction cost regime (Janeček and Shreve (2004); Choi (2014)). Janeček and Shreve (2004) show that, typically, for small transaction costs, the no transaction wedge is centered around the Merton line and has width of the order of the one-third power of the size of the transaction cost. Choi (2014) applies results from Choi et al. (2013) to calculate the next order term. Based on Theorem 6 we give a short and simple argument which yields the results of Janeček and Shreve (2004) and Choi

(2014), and extend to a boundary case in which the width of the no-transaction wedge is exceptionally of square-root order.

Section 6 concludes. Some proofs are given in an Appendix rather than in the main text.

2. Problem specification

Let $Y = (Y_t)_{t\geq 0}$ denote the price of a risky asset and suppose Y is an exponential Brownian motion with drift μ and volatility σ; then $Y_t = Y_0 e^{\sigma B_t + (\mu - \sigma^2/2)t}$ where $B = (B_t)_{t\geq 0}$ is a Brownian motion. Let $C = (C_t)_{t\geq 0}$ denote the consumption rate of the individual and let Θ_t denote the number of units of the risky asset held by the investor. We assume that C is non-negative and progressively measurable and that Θ is progressively measurable with finite variation; in particular $\Theta_t = \Theta_0 + \Phi_t - \Psi_t$ where Φ and Ψ are increasing, adapted, càdlàg processes with $\Phi_{0-} = \Psi_{0-} = 0$ representing purchases and sales of the risky asset respectively.

Suppose cash wealth is right-continuous and evolves according to

$$dX_t = -C_t dt - Y_t(1+\lambda)d\Phi_t + Y_t(1-\gamma)d\Psi_t. \tag{2.1}$$

Here $\lambda \in [0,\infty)$ represents the transaction cost paid on purchases and $\gamma \in [0,1)$ represents the transaction cost paid on sales. We assume $\lambda + \gamma > 0$, else we are in the case of no transaction costs.

Define $\xi = \frac{1+\lambda}{1-\gamma} - 1 = \frac{\lambda+\gamma}{1-\gamma}$ so that $1+\xi$ is the ask to bid ratio. We call ξ the round trip transaction cost: an investor who starts with $\frac{1+\lambda}{1-\gamma}Y_t$ and who buys one unit of the risky asset, only to sell it again immediately, is left with Y_t, and may be considered to have paid a proportional transaction cost of size ξ. It will turn out that it is ξ which governs the nature of the solution to the problem, and not the individual transaction costs λ and γ. Indeed, if we define $\hat{Y}$ via $\hat{Y}_t = Y_t(1-\gamma)$ then (2.1) becomes

$$dX_t = -C_t dt - \hat{Y}_t(1+\xi)d\Phi_t + \hat{Y}_t d\Psi_t, \tag{2.2}$$

and the problem with proportional transaction costs on both purchases and sales reduces to a problem with transaction cost ξ on purchases only. Conversely, if we set $\tilde{Y}_t = Y_t(1+\lambda)$, then we have a problem in which the wealth process satisfies $dX_t = -C_t dt - \tilde{Y}_t d\Phi_t + \frac{1}{1+\xi}\tilde{Y}_t d\Psi_t$ corresponding to a problem with transaction cost $\frac{\xi}{(1+\xi)}$ on sales only. We prefer to specify the problem with distinct transaction costs on buying and selling, because in some circumstances (see especially Section 5.2) it is crucial to separate the two components.

We say that a wealth portfolio (X_t, Θ_t) is *solvent at time t* if

$$X_t + (1-\gamma)\Theta_t^+ Y_t - (1+\lambda)\Theta_t^- Y_t \geq 0,$$

or equivalently if instantaneous liquidation of the risky position yields a cash wealth which is non-negative. A consumption/investment strategy (C,Θ) is *solvent from time t_0* if the resulting wealth portfolio process $(X_t,\Theta_t)_{t\geq t_0}$ is solvent for each $t \geq t_0$. Write $\mathcal{A} = \mathcal{A}(x,y,\theta,t)$ for the set of strategies which are solvent from time t when $(X_{t-} = x, Y_t = y, \Theta_{t-} = \theta)$.

The objective of the agent is to maximize the discounted expected utility of consumption over the infinite horizon, where the discount factor is β and the utility function of the agent is assumed to have constant relative risk aversion with risk aversion co-efficient $R \in (0,\infty) \setminus 1$. The maximization takes place over the set of consumption/investment strategies which are solvent from time zero. In particular, the goal is to find

$$\sup_{(C,\Theta)\in\mathcal{A}(x_0,y_0,\theta_0,0)} \mathbb{E}\left[\int_0^\infty e^{-\beta t}\frac{C_t^{1-R}}{1-R}dt\right]. \tag{2.3}$$

Due to the Markovian structure of the set-up, we expect the value function, optimal consumption and optimal portfolio strategy to be functions of the current wealth portfolio of the agent and of the price of the risky asset. Let $V = V(x,y,\theta,t)$ be the forward starting value function for the problem so that

$$V(x,y,\theta,t) = \sup_{(C,\Theta)\in\mathcal{A}(x,y,\theta,t)} \mathbb{E}\left[\int_t^\infty e^{-\beta s}\frac{C_s^{1-R}}{1-R}ds \middle| X_{t-}=x, Y_t=y, \Theta_{t-}=\theta\right].$$

The goal is to solve for the value function $V = V(x,y,\theta,t)$ and the key quantities of economic interest. Note that it is the value $y\theta$ of the holdings of the risky asset which is important rather than the price level and quantity individually. Further, from the scalings of the problem we expect that we can write $V(x,y,\theta,t) = e^{-\beta t}\frac{x^{1-R}}{1-R}g\left(\frac{y\theta}{x}\right)$, where the key variable is the ratio $z = y\theta/x$ of wealth held in the risky asset to cash wealth. However, there are circumstances in which the agent seeks to leverage her position by allowing cash wealth to go negative (always respecting the solvency condition that after transaction costs her position in the risky asset covers any short cash position). Then we write instead

$$V(x,y,\theta,t) = e^{-\beta t}\frac{(x+y\theta)^{1-R}}{1-R}G\left(\frac{y\theta}{x+y\theta}\right), \tag{2.4}$$

where we call $(x+y\theta)$ the paper wealth, and set $p = \frac{y\theta}{x+y\theta}$ to be the ratio of wealth in the risky asset to paper wealth. Then intuitive arguments of Constantinides and Magill (1976) and the concrete results of Davis and Norman (1990) lead us to expect that the no-transaction region will be a wedge, see Figure 2.1.

3. An auxiliary problem

Let m and ℓ be the quadratic functions

$$m(q) = 1-\epsilon(1-R)q+\frac{\delta^2}{2}R(1-R)q^2 \tag{3.1}$$

$$\ell(q) = 1+\left(\frac{\delta^2}{2}-\epsilon\right)(1-R)q-\frac{\delta^2}{2}(1-R)^2q^2, \tag{3.2}$$

and note that $\ell(q) = m(q)+q(1-q)\frac{\delta^2}{2}(1-R)$. Let $q_M = \frac{\epsilon}{\delta^2 R}$ be the location of the turning point of m and let $m_M = m(q_M) = 1-\frac{\epsilon^2(1-R)}{2\delta^2 R}$ be the value at this point. In the case where $m_M < 0$ let q_- be the root of m in $(0,q_M)$, and let q_+ be the other root.

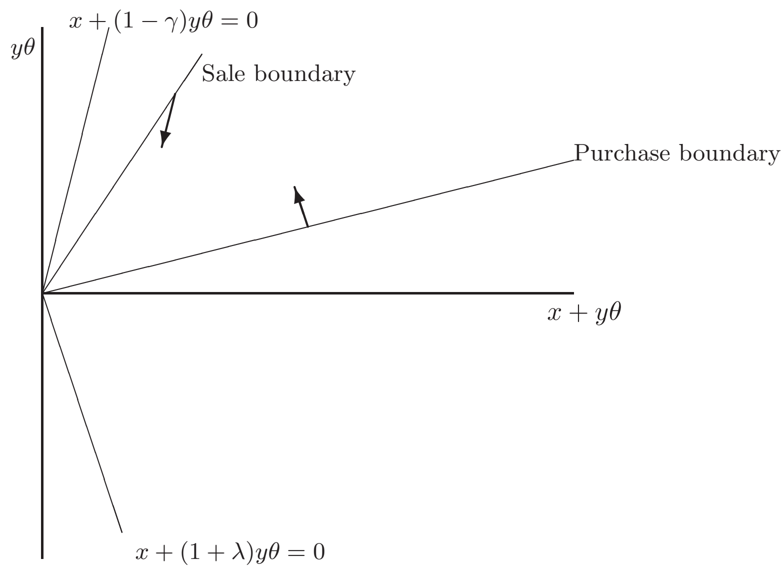


FIGURE 2.1. *The solvency and no-transaction regions.* The solvency region has boundaries given by the lines $x + (1+\lambda)y\theta = 0$ (for $y\theta < 0$) and $x+(1-\gamma)y\theta = 0$ (for $y\theta > 0$). The no-transaction wedge is bounded by the sale and purchase boundaries and lies within the solvency region. On and outside these boundaries, transactions are made to keep the process $(X_t + Y_t\Theta_t, Y_t\Theta_t)$ inside the wedge. The arrows represent the impact of transactions on the boundaries of the no-transaction wedge. As all transactions incur costs, they decrease paper wealth.

In the subsequent analysis the dimensionless quantities ϵ and δ will be related to the parameters of the financial problem via the identities $\epsilon = \frac{\mu}{\beta}$ and $\delta^2 = \frac{\sigma^2}{\beta}$. Then $q_M = \frac{\epsilon}{\delta^2 R} = \frac{\mu}{\sigma^2 R}$ is the Merton proportion and $m_M > 0$ (equivalently $(1-R)\mu^2 < 2\sigma^2 R\beta$) is the condition for the Merton problem without transaction costs to be well posed. q has an interpretation as the shadow portfolio weight on the sale and purchase boundaries.

Suppose $\epsilon > 0$. Let $n = n_r(\cdot)$ be a non-negative solution of the ODE

$$n'(q) = O(q, n(q)) \tag{3.3}$$

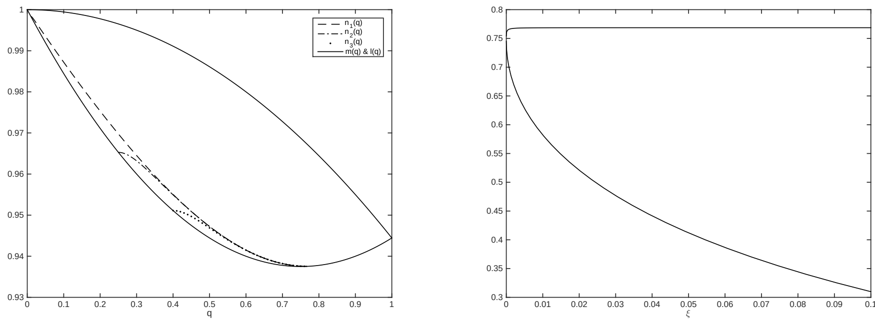
where

$$O(q,n) = \frac{(1-R)}{R}\frac{n}{(1-q)}\frac{m(q)-n}{\ell(q)-n}, \tag{3.4}$$

subject to the initial condition $n_r(r) = m(r)$. We are interested in $n = n_r(\cdot)$ on the interval $[r, \zeta(r)]$ where $\zeta(r) = \inf\{u : u > r, (1-R)n_r(u) < (1-R)m(u)\}$. (When $\epsilon < 0$ we consider the solution $n_r(s)$ for $s < r$.)

Figure 3.1 shows m, ℓ and some typical solutions n in the case where $R < 1$, $q_M \in (0,1)$ and $m_M > 0$. It follows that $\ell > m$ on $(0,1)$. For $r \in (0, q_M)$, as $n_r'(r) = 0$ we see that initially $m(q) < n_r(q) < \ell(q)$ and this property holds true on

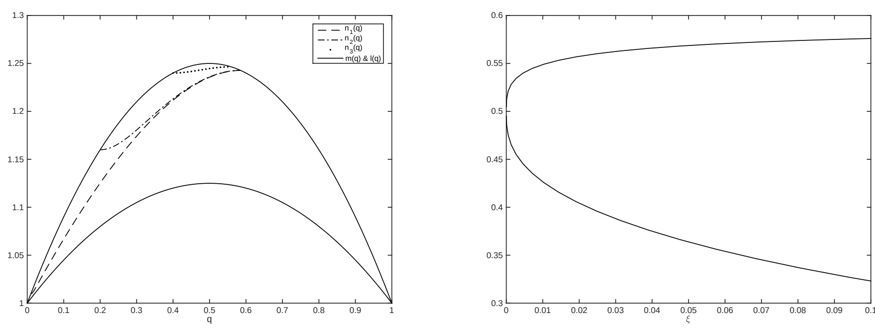
$(r, \zeta(r))$. Hence n_r is decreasing over this range and $\zeta(r) \leq 1$ (and a close look at the solution near $r = 1$ yields $\zeta(r) < 1$).



(A) Typical solutions n_r together with m and ℓ.

(B) $q_* = \Sigma^{-1}(\xi)$ and $q^* = \zeta(q_*)$

FIGURE 3.1. *Parameter values are $\epsilon = \frac{1}{2}$ and $\delta = 1$ and $R = 2/3$. (The equivalent financial parameters are $\mu = \beta/2$, $\sigma^2 = \beta$ and $R = 2/3$.) Panel 3.1a shows some solutions $n_r(q)$. Panel 3.1b shows $q_*(\xi)$ and $q^*(\xi)$ (equivalently the boundaries of the no-transaction region as a function of the level of round-trip transaction cost). Note that the boundary q^* (corresponding to asset sales) is insensitive to the level of transaction costs.*



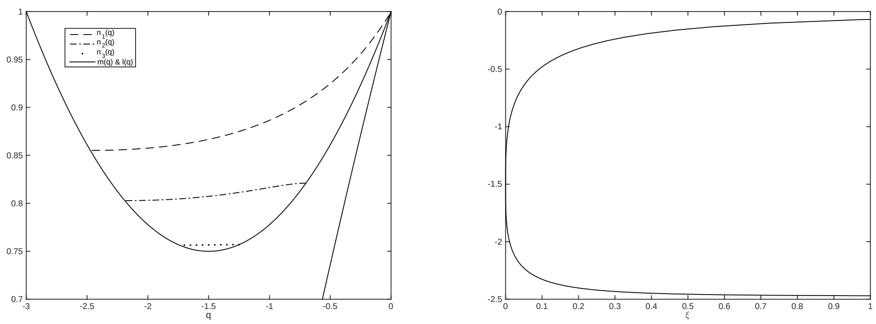
(A) Typical solutions n_r together with m and ℓ.

(B) $q_* = \Sigma^{-1}(\xi)$ and $q^* = \zeta(q_*)$

FIGURE 3.2. *Risk aversion $R > 1$. Parameter values are $\epsilon = \delta = 1$ (equivalently $\mu = \beta = \sigma^2$) and $R = 2$. As $R > 1$ we now find $m > \ell$ over $(0, 1)$ and the solutions we want satisfy $\ell < n < m$ and are increasing.*

The analysis extends easily to the case $R > 1$ (when m becomes concave rather than convex and $\ell < m$ on $(0, 1)$ so that solutions n to (3.3) are increasing, see

Figure 3.2) and to $\epsilon < 0$ when we need to work in the domain $q < 0$ and look for solutions defined to the left of a point $(r, n_r(r) = m(r))$, see Figure 3.3.



(A) Typical solutions n_r together with m and ℓ.

(B) $q^* = \Sigma^{-1}(\xi)$ and $q_* = \zeta(q^*)$.

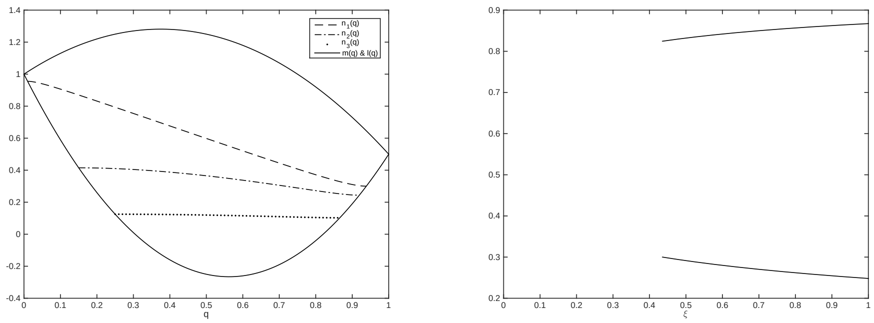
FIGURE 3.3. *Negative drift. Parameter values are $\epsilon = -1$, $\delta = 1$ (equivalently $-\mu = \beta = \sigma^2 > 0$), and $R = 2/3$. Now $\mu < 0$ and we are interested in m, ℓ and n on $q < 0$. For $\mu < 0$ we define solutions $n_r(q)$ for $q \le r < 0$.*

New features of the solution arise when $m_M < 0$ or when $q_M > 1$ (or equivalently $(1-R)m'(1) < 0$). When $m_M < 0 < m(1)$ we can only define solutions n_r for $0 < r < q_-$ (recall q_- is the root of m in $(0, q_M)$) and then $\zeta(r) > q_+$, see Figure 3.4.

When $m(1) > 0$ and $q_M > 1$, all solutions n_r for $0 < r < 1$ pass through the singular point $(1, m(1))$. The next lemma says that these solutions can be extended in a unique way to the right of 1.

Lemma 1. *Suppose $q_M > 1$ and $m(1) > 0$.*
(i) $n_1(\cdot)$ is well defined. Further $\zeta(1) > q_M > 1$ and $(1-R)n_1'(1) = (1-R)m'(1) < 0$.
(ii) For $0 < r < 1$, $n_r(1) = m(1)$ and $n_r'(1) = m'(1)$.
(iii) For $0 < r < 1 < q < \zeta(1)$, $n_r(q) = n_1(q)$. In particular, for $0 < r < 1$, $\zeta(r) = \zeta(1)$.

The intuition behind this lemma is as follows. (To simplify the exposition we assume $R < 1$ and $\epsilon > 0$, but the general case simply requires more care about, for example, whether n lies above or below m and is increasing or decreasing.) As $q_M > 1$ and m is decreasing on $(0, 1)$, for $r \in (0, 1)$ we have that n_r cannot cross m before $q = 1$. As we also have $n_r(q) \le \ell(q)$ on $(0, 1)$ we must have that n passes through the singular point $(1, m(1))$. It turns out that at this point $n'(1) = m'(1)$. A solution for n can be constructed beyond $q = 1$, but because n solves a first order equation, the solution does not depend in any way on the behavior of n to the left of 1. Thus, if $r < 1$, $\zeta(r)$ does not depend on r.



(A) Typical solutions n_r together with m and ℓ.

(B) $q_* = \Sigma^{-1}(\xi)$ and $q^* = \zeta(q_*)$

FIGURE 3.4. $m_M < 0$. *Parameter values are* $\epsilon = \frac{27}{2}$, $\delta = 6$ *(equivalently* $\mu = \frac{27}{2}\beta$, $\sigma^2 = 6\beta$*) and* $R = 2/3$. *Note that if* ξ *is too small then the problem is ill-posed. At the critical* ξ, $\xi = e^{\underline{\Lambda}} - 1$ *we have* $n_{q_*} = 0$ *on* (q_*, q^*).

For $r \in (0, q_M \wedge q_-)$ define

$$\Sigma(r) = \exp\left\{ \int_r^{\zeta(r)} dq \frac{1}{q(1-q)} \frac{n_r(q) - m(q)}{\ell(q) - n_r(q)} \right\} - 1 \tag{3.5}$$

The auxiliary problem is to find a pair (n_{q_*}, q_*) such that n_{q_*} solves (3.3) subject to $n_{q_*}(q_*) = m(q_*)$ and such that $\Sigma(q_*) = \xi$. Recall $1 + \xi = \frac{1+\lambda}{1-\gamma}$ is the ask to bid ratio. Then we set $q^* = \zeta(q_*)$.

Proposition 2. *(1) If* $m_M \geq 0$ *then for every* $\xi \in (0, \infty)$ *there is a unique pair* (n_{q_*}, q_*) *solving the auxiliary problem.*

(2) If $m(1) < 0$ *then there is no pair* (n_{q_*}, q_*) *solving the auxiliary problem for any* $\xi \in (0, \infty)$.

(3) If $m_M < 0$ *and* $m(1) > 0$ *then there is a unique pair* (n_{q_*}, q_*) *solving the auxiliary problem if and only if* $\xi > \underline{\xi}$ *where*

$$\underline{\xi} = \Sigma(q_-) = \exp\left\{ - \int_{q_-}^{q_+} dq \frac{1}{q(1-q)} \frac{m(q)}{\ell(q)} \right\} - 1.$$

Remark 3. The integral in the definition of $\underline{\xi}$ can be evaluated explicitly and written in terms of the roots $q_\pm$ of m and $p_\pm$ of ℓ. We have

$$\begin{aligned} \ln(1+\underline{\xi}) &= -\ln\frac{q_+}{q_-} - \ln\frac{1-q_-}{1-q_+} \\ &\quad + \frac{R}{1-R}\frac{(p_+ - q_+)(p_+ - q_-)}{p_+(p_+ - 1)(p_+ - p_-)} \ln\frac{p_+ - q_-}{p_+ - q_+} \\ &\quad - \frac{R}{1-R}\frac{(q_+ - p_-)(q_- - p_-)}{p_-(1-p_-)(p_+ - p_-)} \ln\frac{q_+ - p_-}{q_- - p_-}. \end{aligned}$$

Here $q_\pm$ and $p_\pm$ can be expressed in terms of the parameters of the financial problem as

$$q_\pm = \frac{\mu \pm \sqrt{\mu^2 - \sigma^2 \frac{2R\beta}{1-R}}}{\sigma^2 R}; \qquad p_\pm = \frac{\frac{1}{2}\sigma^2 - \mu \pm \sqrt{2\sigma^2\beta + (\frac{1}{2}\sigma^2 - \mu)^2}}{\sigma^2(1-R)}.$$

The intuition behind Proposition 2 is based on the fact that Σ is continuous and strictly decreasing on $(0, q_M \wedge q_-)$ and hence has a well-defined inverse. It can also be shown that $\lim_{r\downarrow 0} \Sigma(r) = \infty$ and that if $m_M > 0$ then $\lim_{r\uparrow q_M} \Sigma(r) = 0$. In the case where $m_M < 0$, Σ is bounded below by $\Sigma(q_-)$, where $\Sigma(q_-)$ is calculated by putting the solution $n_{q_-}(q) = 0$ on $[q_-, q_+ = \zeta(q-)]$ into the definition of Σ. It follows that there is no solution to $q_* = \Sigma^{-1}(\xi)$ for small ξ.

For $\xi > 0$ (or for $\xi > \Sigma(q_-)$ when $m_M < 0$) we set $q_* = q_*(\xi) = \Sigma^{-1}(\xi)$ and set $q^* = \zeta(q^*)$. Then (n_{q_*}, q_*) solves the auxiliary problem.

4. Main results

4.1. **Well-posedness.** The quantities $q_M = \frac{\mu}{\sigma^2 R}$ and $m_M = 1 - \frac{\mu^2(1-R)}{2R\beta\sigma^2}$ have been defined as the location of the turning point of m and the value of m at that point. However, they have a direct interpretation in terms of the solution of the Merton problem with zero transaction costs. In the Merton problem with zero transaction costs the optimal strategy is to invest so that the ratio of wealth in the risky asset to total wealth (ie $\frac{\Theta_t Y_t}{X_t + \Theta_t Y_t}$) is kept equal to the constant q_M. Provided $m_M > 0$, the value function for the Merton problem under zero transaction costs is given by $V = V(x, y, \theta, 0) = \frac{(x+y\theta)^{1-R}}{1-R}(\frac{R}{\beta})^R m_M^{-R}$. In particular, positivity of m_M is precisely the condition for well-posed of the Merton problem without transaction costs.

Our first main result relates well-posedness of the problem with transaction costs to the behavior of the quadratic m.

Proposition 4. *(1) If $m_M \geq 0$ then for every $\xi \in (0, \infty)$ the optimal consumption-investment problem is well-posed.*

(2) If $m(1) < 0$ then for every $\xi \in (0, \infty)$ the optimal consumption-investment problem is ill-posed.

(3) If $m_M < 0$ and $m(1) > 0$ then the optimal consumption-investment problem is well-posed if and only if $\xi > \underline{\xi}$.

This proposition can be rewritten in terms of the fundamental parameters of the problem, rather than the quadratic m. When we do so we recover the main result of Choi et al. (2013), Theorem 2.6, except that Choi et al. (2013) do not identify the constant $\underline{\xi}$ (and, less significantly, Choi et al. (2013) do not include the case $\mu < 0$).

Theorem 5 (See also Choi et al. (2013))**.** *(1) If $R > 1$ or if $R < 1$ and $|\mu| \leq \sigma\sqrt{\frac{2R\beta}{1-R}}$ then for every $\xi \in (0, \infty)$ the problem is well-posed.*

(2) If $R < 1$ and $\mu > \frac{\beta}{1-R} + \frac{R\sigma^2}{2}$ then for every $\xi \in [0, \infty)$ the problem is ill-posed.

(3) *If $R < 1$ and $\mu < -\sigma\sqrt{\frac{2R\beta}{1-R}}$ or $\sigma\sqrt{\frac{2R\beta}{1-R}} < \mu < \frac{\beta}{1-R} + \frac{R\sigma^2}{2}$ then the problem is well-posed if and only if $\xi > \underline{\xi}$.*

The idea behind our proof of Theorem 5 is to consider the value function which solves a (second-order, non-linear) Hamilton-Jacobi-Bellman (HJB) equation and to exploit various scalings and symmetries of the problem. Crucially, we use a change of independent variable to reduce the order of the problem. Following this order reduction second-order smooth fit at an unknown free boundary becomes starting and ending on a fixed boundary. (For a similar use of this change of independent variable, see Evans et al. (2008). Choi et al. (2013), see also Gerhold et al. (2014), use a different transformation to reduce the order of their second-order equation.) After several transformations and simplifications (details are provided in Appendix A below) we are left to consider the auxiliary problem of Section 3. Conversely, by inverting the same transformations, the solution to the auxiliary problem can be used to define a candidate value function, which can be confirmed to be the true value function by a classical verification argument. The formal analysis is helped by the fact that the candidate value function is C^2 on the solvency region.

4.2. **The location of the no-transaction wedge.** Part of the solution of the auxiliary problem is the pair of free-boundaries q_* and $q^* = \zeta(q_*)$. One of the advantages of our approach is that there is a direct interpretation of these quantities in terms of the no-transaction wedge.

Theorem 6. *Let (n_{q_*}, q_*) solve the auxiliary problem and let $q^* = \zeta(q_*)$. Then the no-transaction wedge is given by $p_* \leq \frac{y\theta}{x+y\theta} \leq p^*$ where*

$$p_*(\lambda,\gamma) = \frac{q_*(\xi)}{1+\lambda-\lambda q_*(\xi)}; \qquad\qquad p^*(\lambda,\gamma) = \frac{q^*(\xi)}{1-\gamma+\gamma q^*(\xi)}, \tag{4.1}$$

and $\xi = \frac{\lambda+\gamma}{1-\gamma}$.

Remark 7. The first equation in (4.1) can be rewritten as $q_* = \frac{p_*(1+\lambda)}{1+p_*\lambda}$. Recall that $p = \frac{y\theta}{x+y\theta}$ so that at the purchase boundary we find $q_* = \frac{y\theta(1+\lambda)}{x+y\theta(1+\lambda)}$ which is the proportion of wealth in the risky asset when wealth in the risky asset is valued using the bid price. Similar considerations apply on the sale boundary.

5. Discussion and implications of the main results

5.1. **Comparative statics for the dependence of the no-transaction wedge on the transaction cost parameters.** For most of this paper we have argued that the transaction costs λ on purchases and γ on sales only enter the problem through the bid to ask ratio $1 + \xi = \frac{1+\lambda}{1-\gamma}$. Whilst that is true for the construction of n (and the locations of the free-boundaries q_* and q^* from which the solution is built), the boundaries of the no-transaction wedge do depend on the individual transaction costs and we have $p_* = p_*(\lambda,\gamma)$ and $p^* = p^*(\lambda,\gamma)$ given by (4.1).

Lemma 8. *Suppose all parameters are fixed, except for the transaction costs λ and γ. Suppose also that the problem is well-posed. Then q^* is non-decreasing in ξ, and q_* is decreasing in ξ.*

Theorem 9. *Suppose parameters are such that the problem is well-posed.*

(i) *If $0 < \mu < \sigma^2 R$ then the sale boundary p^* is increasing in λ and increasing in γ and the purchase boundary p_* is decreasing in λ and decreasing in γ. Further, the Merton line lies within the no-transaction wedge. We have $0 < p_* < q_M = \frac{\mu}{\sigma^2 R} < p^* < 1$.*

(ii) *Suppose $\mu > \sigma^2 R$ or $\mu < 0$. If $\mu > \sigma^2 R$ then $p^* > 1$ and the agent will (at least sometimes) take a leveraged position. If $\mu < 0$ then $p_* < p^* < 0$ and the agent will take a short position. There are parameter combinations for which p_* and p^* are not monotonic in the individual transaction costs. Further, the Merton line need not lie within the no-transaction wedge.*

Proof. We have

$$\frac{dp^*}{d\lambda} = \frac{\partial \xi}{\partial \lambda}\frac{\partial q^*}{\partial \xi}\frac{dp^*}{dq^*} = \frac{1}{(1-\gamma)}\frac{\partial q^*}{\partial \xi}\frac{1-\gamma}{(1-\gamma+\gamma q^*)^2} > 0$$

and

$$\frac{dp^*}{d\gamma} = \frac{q^*(1-q^*)}{(1-\gamma(1-q^*))^2} + \frac{\partial \xi}{\partial \gamma}\frac{\partial q^*}{\partial \xi}\frac{dp^*}{dq^*} = \frac{q^*(1-q^*)}{(1-\gamma(1-q^*))^2} + \frac{1+\lambda}{(1-\gamma)^2}\frac{\partial q^*}{\partial \xi}\frac{1-\gamma}{(1-\gamma+\gamma q^*)^2}$$

If $0 < \mu < \sigma^2 R$ then $0 < q^* = \frac{\mu}{\sigma^2 R} < 1$ and the sign of both terms is positive, but if $q^* \notin [0,1]$ then either term may dominate.

Similarly,

$$\frac{dp_*}{d\gamma} = \frac{\partial \xi}{\partial \gamma}\frac{\partial q_*}{\partial \xi}\frac{dp_*}{dq_*} = \frac{1+\lambda}{(1-\gamma)^2}\frac{\partial q_*}{\partial \xi}\frac{1+\lambda}{(1+\lambda-\lambda q_*)^2} < 0$$

and

$$\frac{dp_*}{d\lambda} = \frac{-q_*(1-q_*)}{(1+\lambda(1-q_*))^2} + \frac{\partial \xi}{\partial \lambda}\frac{\partial q_*}{\partial \xi}\frac{dp_*}{dq_*} = \frac{-q_*(1-q_*)}{(1+\lambda-\lambda q_*))^2} + \frac{1}{(1-\gamma)}\frac{\partial q_*}{\partial \xi}\frac{1+\lambda}{(1+\lambda-\lambda q_*)^2}$$

If $0 < \mu < \sigma^2 R$ then $0 < q^* < 1$ and the sign of both terms is negative, but if $q_* \notin [0,1]$ then either term may dominate.

Note that solvency requires that $p^* < \frac{1}{\gamma}$. So, when $\mu > \frac{\sigma^2 R}{\gamma}$ we have $1 < p^* < \frac{1}{\gamma} < q_M$ and the Merton line lies outside the no transaction wedge. □

Of interest is the location of the no-transaction wedge and the relationship between the Merton line and the no-transaction wedge. The key advantage we have over the previous literature (Davis and Norman (1990); Shreve and Soner (1994)) is that we have decoupled the expressions for the locations of the boundaries of the no-transaction wedge into two parts: we have p_* and p^* given by (4.1) where $q_* < q_M < q^*$.

Davis and Norman (1990) argue that if $0 < \mu < \sigma^2 R \wedge \sigma\sqrt{\frac{2R\beta}{1-R}}$ (and a further technical condition, Condition B holds) then the no-transaction wedge lies in the subspace $x > 0$ and contains the Merton line. They also conjecture (Davis and Norman (1990), p704) that if the problem is well-posed and $\mu > \sigma^2 R$ then the no-transaction wedge lies in the subspace $x < 0$. As we have seen, if transaction costs are sufficiently large, we may have $p_* < q_* < 1$ and then this is not the case.

Shreve and Soner (1994) give bounds on p_* and p^*. They state in (11.4), (11.5) and (11.6) of Shreve and Soner (1994) that

$$p^* < \frac{\mu}{\frac{1}{2}(1-\gamma)\sigma^2 R + \gamma\mu}; \tag{5.1}$$

if $R > 1$ or $R < 1$ and $0 < \mu < \sigma\sqrt{\frac{2R\beta}{1-R}}$

$$p^* > \frac{\mu}{(1-\gamma)\sigma^2 R + \gamma\mu}; \tag{5.2}$$

and if $R > 1$ or $R < 1$ and $0 < \mu < \sigma\sqrt{\frac{2R\beta}{1-R}}$, and if $\mu < \sigma^2 R\frac{1+\lambda}{\lambda}$

$$p_* < \frac{\mu}{(1+\lambda)\sigma^2 R - \lambda\mu}. \tag{5.3}$$

The bounds (5.1), (5.2) and (5.3) can be seen to follow from our results, sometimes under weaker assumptions.

If $0 < q_M < 1$ (equivalently $0 < \mu < \sigma^2 R$) and the problem is well-posed then as m is a quadratic and n is monotone we must have $(1-R)m(q^*) = (1-R)n(q^*) < (1-R)n(q_*) = (1-R)m(q_*)$ and so $q^* - q_M < q_M - q_* < q_M$. We conclude that $q^* < \min\{2q_M, 1\}$. Then, as $q_* < q_M < q^*$,

$$p^* = \frac{q^*}{(1-\gamma)+\gamma q^*} < \frac{2q_M}{(1-\gamma)+\gamma 2q_M} = \frac{\mu}{\frac{1}{2}(1-\gamma)\sigma^2 R + \gamma\mu}; \tag{5.4}$$

$$p^* = \frac{q^*}{(1-\gamma)+\gamma q^*} > \frac{q_M}{(1-\gamma)+\gamma q_M} = \frac{\mu}{(1-\gamma)\sigma^2 R + \gamma\mu}; \tag{5.5}$$

$$p_* = \frac{q_*}{(1+\lambda)-\lambda q_*} < \frac{q_M}{(1+\lambda)-\lambda q_M} = \frac{\mu}{(1+\lambda)\sigma^2 R - \lambda\mu}. \tag{5.6}$$

Note that from $q^* < 1$ we also have the bound $p^* < 1$, and the no-transaction wedge lies in the region where $x > 0$ and the agent never leverages her position.

If $q_M > 1$ (equivalently $\mu > \sigma^2 R$) and the problem is well-posed then due to $(1-R)m(q^*) = (1-R)n(q^*) < (1-R)n(1) = (1-R)m(1)$ we have $q^* - q_M < q_M - \max\{q_*, 1\} \le q_M - 1$. Then $1 < q_M < q^* < 2q_M - 1$ and (5.4) can be refined to

$$p^* < \frac{2q_M - 1}{(1-\gamma)+\gamma(2q_M - 1)} = \frac{2\mu - \sigma^2 R}{(1-2\gamma)\sigma^2 R + 2\gamma\mu}.$$

(5.5) and (5.6) hold as before, (5.6) provided $\mu < \sigma^2 R\frac{1+\lambda}{\lambda}$.

Shreve and Soner (1994) also conjecture that (p675) if $q_M > 1$ then $p^* < q_M$ and the Merton line lies outside the no-transaction wedge. If $q_M > 1$ then we have $q^* < 2q_M - 1$ and $p^* < \frac{2q_M-1}{(1-\gamma)+\gamma(2q_M-1)}$. Then if $\frac{1}{2q_M} < \gamma < 1$ so that transaction costs on sales are large we have $p^* < \frac{2q_M-1}{(1-\gamma)+\gamma(2q_M-1)} < q_M$ and the Shreve-Soner conjecture is true. However, if transaction costs on sales are small we may find $p_* < q_M < p^*$, and the Merton line lies inside the no-transaction wedge. In particular, if $\gamma = 0$ then $p^* = q^*$ and $p^* > q_M$.

5.2. **Leverage and the independence of the sale boundary on the transaction cost parameter.** Suppose $\mu > \sigma^2 R$, or equivalently $q_M > 1$. Then the investor in the problem without transaction costs seeks to leverage her position by borrowing to finance a large position in the risky asset. We want to discuss further the implications for the behavior of the investor in the presence of transaction costs.

Recall the definition of Σ in (3.5). When $1 < r < \zeta(r)$ it is clear that Σ is well-defined. When $r < 1 < \zeta(r)$ it can be shown that the potential singularity at $q = 1$ can be removed (note $m(1) = n(1)$ and $n'(1) = m'(1)$) and hence that $\Sigma(r)$ is still continuous and strictly decreasing. Set

$$\overline{\xi} = \Sigma(1) = \exp\left\{ \int_1^{\zeta(1)} dq \frac{1}{q(q-1)} \frac{n_1(q) - m(q)}{n_1(q) - \ell(q)} \right\} - 1.$$

Note that if $R < 1$ and $m_M < 0$ then for $r > 1$, $n_r(q) < n_1(q)$ over the domain where both are defined, and hence $\Sigma(q_-) < \Sigma(1)$. Thus $\underline{\xi} < \overline{\xi}$.

Lemma 1 leads to the following result and corollary:

Lemma 10. *If $\mu > \sigma^2 R$ and $\xi \geq \overline{\xi}$ then $q_* \leq 1$ and q^* does not depend on ξ.*

Corollary 11. *If $\mu > \sigma^2 R$ and $\xi \geq \overline{\xi}$ then the ray $x = 0$ is contained in the no-transaction wedge and p^* does not depend on λ.*

At first sight the final conclusion of Corollary 11 may appear surprising. At a mathematical level, the result is a consequence of the fact that the relevant solution n_r passes through the singular point $(1, m(1))$ and on doing so 'forgets' its starting point $(r, m(r))$. Hence q^* does not depend on ξ for $\xi \geq \overline{\xi}$. As $p^* = \frac{q^*}{(1-\gamma)+\gamma q^*}$ we find that p^* does not depend on λ^*. The financial explanation of this result is fairly simple also. If the no-transaction wedge includes the ray corresponding to $x = 0$ and if ever $x = 0$, then the agent finances consumption first by spending cash reserves, then by borrowing, and then when borrowing levels become too great, from sales of the risky asset. But, once cash-wealth is non-positive, the agent will trade in such a way that cash wealth is never positive at any future moment. Thus, once $x = 0$ the agent will never again purchase units of risky asset, and the transaction cost on purchases becomes irrelevant. Hence the location of the sell threshold does not depend on λ. The same arguments show that the value function in the region $(x \leq 0)$ does not depend on λ (for $\xi \geq \overline{\xi}$), although it continues to depend on λ in the region $x > 0$.

Choi et al. (2013) show a related result and conclude that the shadow price and value function are independent of the value of transaction costs in certain domains (provided the level of the transaction cost is above a certain critical value). However, they do not give a financial explanation of this result.

5.3. **The small transaction cost limit.** Both this paper and Choi et al. (2013) give solutions to the Merton problem with general levels of proportional transaction costs. Given these results for general transaction costs it is interesting to consider the implications in the (often financially relevant) small transaction cost regime. The aim is to understand how the location and width of the no-transaction region depend on the transaction cost parameters.

Rogers (2004) gives a general argument to show that for proportional transaction costs we expect the width of the no-transaction wedge to be of the order of the size of transaction costs to the one-third power. Janeček and Shreve (2004) formalize this result and show that (provided the problem without transaction costs is well-posed and $\mu \notin \{0, \sigma^2 R\}$)

$$p_* = q_M - \Delta_1^{1/3}\xi^{1/3} + O(\xi^{2/3}) \qquad p^* = q_M + \Delta_1^{1/3}\xi^{1/3} + O(\xi^{2/3}). \tag{5.7}$$

where $q_M = \frac{\mu}{\sigma^2 R}$ is the fraction of wealth invested in the risky asset (the Merton proportion) and

$$\Delta_1 = \left(\frac{3q_M^2(1-q_M)^2}{4R} \right) = \left(\frac{3\mu^2(\sigma^2 R - \mu)^2}{4\sigma^8 R^5} \right)$$

Choi (2014) uses the results of Choi et al. (2013) to show that (when $m_M > 0$ and $q_m \notin \{0,1\}$)

$$p_* = q_M - \Delta_1^{1/3}\xi^{1/3} - \Delta_2\Delta_1^{2/3}\xi^{2/3} + O(\xi); \quad p^* = q_M + \Delta_1^{1/3}\xi^{1/3} - \Delta_2\Delta_1^{2/3}\xi^{2/3} + O(\xi). \tag{5.8}$$

where

$$\Delta_2 = \frac{2m_M}{3q_M(1-q_M)^2\delta^2 R} = \frac{\sigma^2 R(2\sigma^2 R\beta - \mu^2(1-R))}{3\mu(\sigma^2 R - \mu)^2}.$$

In principle, Choi's results can be extended to give an expasion to any order. (In fact Janeček and Shreve (2004) assume $\lambda = \gamma$ and Choi (2014) assumes $\gamma = 0$ but it is straightforward to translate their results to a more general setting.)

From Theorem 6 we have

$$p_*(\lambda,\gamma) = \frac{q_*}{1+\lambda-\lambda q_*} = q_*(\xi) - \lambda q_*(\xi)(1-q_*(\xi)) + O(\lambda^2,\gamma^2)$$

and $p^*(\lambda,\gamma) = q^*(\xi) + \gamma q^*(\xi)(1-q^*(\xi)) + O(\lambda^2,\gamma^2)$. Hence in calculating an expansion for p_* or p^* up to order $\frac{2}{3}$ as in (5.7) or (5.8) it is sufficient to consider expansions of $q_*(\xi)$ in powers of $\xi^{1/3}$. It is only when we consider expansions to order one that the individual transaction costs become important.

In this section we explain how our results yield (5.7) almost immediately. In Appendix C we show how to derive (5.8), and give a method to extend the results to higher order. We also consider what happens when $q_M = 1$. Exceptionally, in this case the leading order term is of different order, see Appendix C.

Proposition 12. *Suppose $\mu = \sigma^2 R$ and $2\beta > \sigma^2 R(1-R)$. Then $p^* = 1$ and $p_* = 1 - \Upsilon^{1/2}\xi^{1/2} + O(\xi)$ where $\Upsilon = \frac{2\beta m_M}{\sigma^2 R^2} = \frac{2\beta - \sigma^2 R(1-R)}{\sigma^2 R^2}$.*

For ease of exposition we discuss a derivation of (5.7) in the case $R < 1$ and $\mu > 0$. Suppose $\mu \neq \sigma^2 R$ and that the problem is well-posed. Consider $n_r(q)$ for r just a little bit smaller than q_M. On $[r, \zeta(r)]$ we have $0 \leq n_r(q) - m(q) \leq n_r(r) - m(q_M) = m(r) - m(q_M)$ which is of order $(r - q_M)^2$. It follows that solutions to $n' = O(q,n)$ are approximately horizontal lines. If we write $\hat{n}$ for this approximate solution, and $\hat{\zeta}$ and $\hat{\Sigma}$ for the corresponding first re-crossing of m and approximate value of Σ

then for small $u > 0$, $\hat{n}_{q_M-u}(q) = m(q_M - u)$, $\hat{\zeta}(q_M - u) = q_M + u$ and (we write $a \sim b$ if the terms agree to leading order)

$$\begin{aligned}\ln(1 + \hat{\Sigma}(q_M - u)) &= \int_{q_M-u}^{\hat{\zeta}(q_M-u)} \frac{dq}{q(1-q)} \frac{\hat{n}_{q_M-u}(q) - m(q)}{\ell(q) - \hat{n}_{q_M-u}(q)} \\ &\sim \frac{1}{q_M(1-q_M)} \frac{1}{\ell(q_M) - m(q_M)} \int_{-u}^{u} dv[m(q_M - u) - m(q_M + v)] \\ &= \frac{R}{q_M^2(1-q_M)^2} \int_{-u}^{u} [u^2 - v^2] dv = \Delta_1^{-1} u^3\end{aligned}$$

Taking inverses we find (ignoring terms of order $\xi^{2/3}$ or higher) $q_*(\xi) = q_M - \Delta_1^{1/3}\xi^{1/3}$ and hence also $q^*(\xi) = q_M + \Delta_1^{1/3}\xi^{1/3}$ as in (5.7).

5.4. **Dependence of the no-transaction wedge on the drift.**

Theorem 13. *Suppose all parameters except the drift are constant and that the problem is well posed. Then both the purchase and sale boundaries of the no-transaction wedge are increasing in the drift in the underlying asset.*

Proof. We want to show that both p_* and p^* are increasing in μ, which is equivalent to q_* and q^* increasing in ϵ. We consider the case $R < 1$ and $\epsilon > 0$; similar arguments work for $R > 1$ and/or $\epsilon < 0$.

Fix $\hat{\epsilon} > \tilde{\epsilon}$ and let $\hat{n}_r$ and $\tilde{n}_r$ denote the solutions of $n' = O(q, \hat{m}, n)$ and $n' = O(q, \tilde{m}, n)$ subject to $n_r(r) = 0$ where

$$O(q, m, n) = -\frac{1-R}{R} \frac{n}{1-q} \frac{n - m(q)}{\left(m(q) + \frac{\delta^2}{2}(1-R)q(1-q) - n\right)}.$$

Here $\hat{m}(q)$ (respectively $\tilde{m}$) is the quadratic $\hat{m}(q) = 1 - \hat{\epsilon}(1-R)q + \frac{\delta^2}{2}R(1-R)q^2$ (respectively $\tilde{m}(q) = 1 - \tilde{\epsilon}(1-R)q + \frac{\delta^2}{2}R(1-R)q^2$). In general let the $\hat{\cdot}$ and $\tilde{\cdot}$ symbols denote solutions defined relative to $\hat{\epsilon}$ and $\tilde{\epsilon}$. Let $m_0(q) = 1 + \frac{1}{2}\delta^2 R(1-R)q^2$.

Let $\hat{a}(q) = \hat{a}_r(q) = \hat{n}_r(q) - \hat{m}(q)$. Then

$$\begin{aligned}\hat{a}'(q) &= O(q, m_0(q) - \hat{\epsilon}(1-R)q, m_0(q) - \hat{\epsilon}(1-R)q + \hat{a}(q)) + \hat{\epsilon}(1-R) - \delta^2 R(1-R)q \\ &= O(q, m_0(q), m_0(q) + \hat{a}) - \delta^2 R(1-R)q \\ &\quad + \hat{\epsilon}(1-R)\left[\frac{1-R}{R} \frac{q}{(1-q)} \frac{\hat{a}}{(\frac{1}{2}\delta^2(1-R)q(1-q) - \hat{a})} + 1\right] \\ &=: \hat{O}(q, \hat{a}).\end{aligned}$$

For $q < 1$ we have $0 < a < \ell(q) - m(q)$ and $\hat{O}(q, a) > \tilde{O}(q, a)$, and we conclude that away from $q = r$, $\hat{a}_r$ and $\tilde{a}_r$ cannot cross. Consideration of the case $q > 1$ leads to the a similar conclusion.

Suppose first that $\hat{\epsilon} < \delta^2 R$ so that we may restrict attention to $r < q < 1$. Fix r. Then $\hat{a}_r(q) > \tilde{a}_r(q)$ at least until $\hat{\zeta}(r) \wedge \tilde{\zeta}(r)$ and then it follows both that $\hat{\zeta}(r) > \tilde{\zeta}(r)$ and $\hat{\Sigma}(r) > \tilde{\Sigma}(r)$, where we make use of $a = n - m$ and the representation

$$\Sigma(r) = \exp\left(\int_r^{\zeta(r)} \frac{dq}{q(1-q)} \frac{a(q)}{\frac{\delta^2}{2}R(1-R)q(1-q) - a(q)}\right) - 1,$$

which we note only depends on ϵ through a. As Σ is decreasing in r we conclude that $\hat{q}_* = \hat{\Sigma}^{-1}(\xi) > \tilde{\Sigma}^{-1}(\xi) = \tilde{q}_*$ and that q_* is increasing in ϵ.

In order to consider the sale boundary p^* it is convenient to parameterize solutions of the free boundary problem by the boundary point q^* rather than q_*. Let n solve $n' = O(q,n)$ in $q \leq s$ subject to $n_s(s) = m(s)$ and let $a_s(q) = n_s(q) - m(q)$. Then, we have $\zeta^{-1}(s) = \sup\{u \leq s : n_s(u) < m(s)\}$ and we get that solutions $(a_s(q))_{\{\zeta^{-1}(s) \leq q \leq s\}}$ are decreasing in ϵ; hence $\zeta^{-1}(s)$ is increasing in ϵ and Σ is decreasing in ϵ. It follows that q^* is also increasing in ϵ.

Now we relax the assumption that $\hat{\epsilon} < \delta^2 R$. If $\tilde{\epsilon} \leq \delta^2 R < \hat{\epsilon}$, then $\tilde{q}^* \leq 1 < \hat{q}^*$. For q_* the same proof as given above can be used.

Finally, if $\tilde{\epsilon} > \delta^2 R$ then for sufficiently small transaction costs we have $\hat{q}_* > 1$, and then by the arguments as above we can conclude that q_* and q^* are monotonic. The only point of delicacy is when transaction costs are larger, when we must consider the case where both $\hat{q}_*$ and $\tilde{q}_*$ lie below the singular point. Then, for $r < 1$, $\hat{a}_r(1) = \tilde{a}_r(1) = 0$. Nonetheless, for $r < 1$ we have the inequality $\hat{a}_r \geq \tilde{a}_r$ with strict inequality on $(r, 1)$, and hence $\hat{\Sigma}(r) > \tilde{\Sigma}(r)$. Note that $\hat{q}^* = \tilde{q}^*$ for large transaction costs.

□

6. Conclusion and further remarks

Our goal in this paper was to analyze the Merton problem with transaction costs. Following Davis and Norman (1990) and Shreve and Soner (1994), our approach is via the primal problem rather than the shadow-price approach of Kallsen and Muhle-Karbe (2010),Choi et al. (2013) and Herczegh and Prokaj (2015). Thus our approach brings different insights to the shadow-price literature. We were able to show via judicious transformations that the problem could be reduced to solving a first-order ordinary differential equation. There is a family of solutions to this ODE, and the one we want satisfies an additional integral equation. The value function for the consumption-investment problem can be constructed via integrating the solution to this auxiliary problem; however, the locations of the boundaries to the no-transaction wedge can be read immediately from the pair (q_*, q^*) which forms part of the solution to the auxiliary problem. Thus, especially when the questions of interest concern the location of the no-transaction wedge, our approach is very powerful.

At one level our results are a re-parametrization of the results of Choi et al. (2013) although the derivation is completely different. Choi et al. (2013) also reduce the problem to solving a first order ODE, subject to smooth fit conditions on a free-boundary, and subject to an integral condition. But in their case the points on the free-boundary lie on an ellipse (rather than a quadratic) and the phase-diagram is considerably more complicated. Our approach to the problem brings simplifications to the analysis and allows us to prove further results. First, we can relate the different cases to the different possible behaviors of a quadratic function of one variable. Second, in the case where the problem is ill-posed for zero transaction costs, we can give an algebraic expression for the value of the transaction costs at which the problem becomes ill-posed. (Choi et al. (2013) are only able to give this as an integral involving the roots of a quadratic, see their Lemma 6.11) Third, it is immediate from our approach that the integral equation which determines which of the family of candidate solutions of the ODE we want has a monotonicity property. In particular, it is immediate from our approach that Σ is strictly decreasing and has an inverse: Choi et al. (2013) are not able to give a corresponding monotonicity argument for their equivalent function (see their Remark 6.15). Fourth, a crucial component of our solution is the pair (q_*, q^*) which correspond to the boundaries of the no-transaction wedge. The fact that these quantities are an explicit element of the solution allows us to give several results on the comparative statics associated with these quantities. Some of these results have appeared in the literature, (in which case the advance is that our analysis typically gives shorter derivations of the key results). In other cases our results (for example, on the width of the no-transaction wedge when $\mu = \sigma^2 R$, and the comparative statics with respect to drift) are new. Finally, our results bring an important insight into the phenomenon that for some parameter values the location of the sell boundary does not depend in any way on the transaction cost for purchases. Mathematically, this relates to the fact that candidate solutions of the auxiliary problem all pass through the same singular point. Financially, it relates to the fact that the ray corresponding to zero cash wealth is in the no-transaction wedge, and if cash wealth ever hits zero, then under optimal behavior it remains non-positive thereafter, and the purchase boundary is never reached again.

In this paper we restrict attention to the case of a single risky asset. The situation with multiple risky assets and transaction costs is much more complicated. Nonetheless, the methods of this paper can be generalized to the multi-asset case, albeit only to a special case in which transaction costs are payable on one asset only. See Hobson et al. (2016), and also Choi (2018).

References

Choi, J. (2014). Asymptotic analysis for merton's problem with transaction costs in power utility case. *Stochastics*, 86(5):803–816.

Choi, J. (2018). Optimal investment and consumption with liquid and illiquid assets. *arXiv:1602.06998v3*.

Choi, J., Sirbu, M., and Zitkovic, G. (2013). Shadow prices and well posedness in the problem of optimal investment and consumption with transaction costs. *SIAM Journal on Control and Optimization*, 51(6):4414–4449.

Constantinides, G. M. and Magill, M. J. P. (1976). Portfolio selection with transaction costs. *Journal of Economic Theory*, 13:245–263.

Davis, M. H. A. and Norman, A. R. (1990). Portfolio selection with transaction costs. *Mathematics of Operations Research*, 15(4):676–713.

Evans, J. D., Henderson, V., and Hobson, D. (2008). Optimal timing for an indivisible asset sale. *Mathematical Finance*, 18(4):545–567.

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

Herczegh, A. and Prokaj, V. (2015). Shadow price in the power utility case. *The Annals of Applied Probability*, 25(5):2671–2707.

Hobson, D., Tse, A., and Zhu, Y. (2016). A multi-asset investment and consumption problem with transaction costs. *arXiv:1612.01327*.

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

Kallsen, J. and Muhle-Karbe, J. (2010). On using shadow prices in portflio optimization with transaction costs. *The Annals of Applied Probability*, 20(4):1341–1358.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: the continuous-time case. *The Review of Economics and Statistics*, 51(3):247–257.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413.

Rogers, L. (2004). Why is the effect of transaction costs $o(\delta^{2/3})$? *Mathematics of Finance, AMS Contemporary Mathematics Series 351. Editors G. Yin, Q. Zhang*, pages 303–308.

Shreve, S. E. and Soner, H. M. (1994). Optimal investment and consumption with transaction costs. *The Annals of Applied Probability*, 4(3):609–692.

Zhu, Y. (2015). Investment-consumption model with infinite transaction costs. *PhD Thesis, University of Warwick*.

Appendix A. Derivation of the candidate value function

Define $P = (P_t)_{t \geq 0}$ by $P_t = \frac{Y_t \Theta_t}{X_t + Y_t \Theta_t}$. The solvency requirement can be expressed as $-\frac{1}{\lambda} \leq P_t \leq \frac{1}{\gamma}$. Write

$$V(x, y, \theta, t) = e^{-\beta t} \frac{(x + y\theta)^{1-R}}{1-R} \left(\frac{R}{\beta} \right)^R G\left(\frac{y\theta}{x + y\theta} \right), \tag{A.1}$$

and consider

$$M_t := \int_0^t e^{-\beta s} \frac{C_s^{1-R}}{1-R} ds + V(X_t, Y_t, \Theta_t, t). \tag{A.2}$$

Applying Itô's formula we find (subscripts x, y, θ denote space derivatives and $\dot{V}$ denotes a time derivative)

$$\begin{aligned} dM_t &= \frac{C_t^{1-R}}{1-R} e^{-\beta t} dt + \dot{V} dt + V_x dX_t + V_y dY_t + V_\theta d\Theta_t + \frac{1}{2} V_{yy} d[Y]_t \\ &= \left\{ \frac{C_t^{1-R}}{1-R} e^{-\beta t} - V_x C_t \right\} dt + [V_\theta - V_x(1+\lambda)Y_t] d\Phi_t + [V_x(1-\gamma)Y_t - V_\theta] d\Psi_t \\ &\quad + e^{-\beta t} \frac{(X_t + Y_t \theta_t)^{1-R}}{1-R} \left(\frac{R}{\beta} \right)^R LG(P_t) dt + \sigma Y_t V_y dB_t \end{aligned}$$

where for $H = H(p)$

$$LH = -\beta H + \mu\left[(1-R)pH + p(1-p)H'\right] + \frac{\sigma^2}{2}\left[-R(1-R)p^2H - 2Rp^2(1-p)H' + p^2(1-p)^2H''\right].$$

As M is a martingale under the optimal strategy and a super-martingale otherwise, maximizing over C_t we find $C_t = e^{-\frac{\beta}{R}t}V_x^{-1/R} = \frac{\beta}{R}(X_t + Y_t\theta_t)\left[G(P_t) - \frac{P_tG'(P_t)}{1-R}\right]^{-1/R}$. As consumption is non-negative we must have that $G(p) > \frac{pG'(p)}{1-R}$. Further, if $d\Phi_t > 0$ then $V_\theta = (1+\lambda)yV_x$. Hence, if $d\Phi_t > 0$ then $(1-R)G(P_t) + (1-P_t)G'(P_t) = (1+\lambda)[(1-R)G(P_t) - P_tG'(P_t)]$ or equivalently

$$-\lambda(1-R)G(P_t) + (1+\lambda P_t)G'(P_t) = 0.$$

Similarly, if $d\Psi_t > 0$ then

$$\gamma(1-R)G(P_t) + (1-\gamma P_t)G'(P_t) = 0. \tag{A.3}$$

It follows that for $-\frac{1}{\lambda} \le p \le p_*$ we have $G(p) = \left(\frac{1+\lambda p}{1+\lambda p_*}\right)^{1-R} G(p_*)$ and for $p^* \le p \le \frac{1}{\gamma}$ we have $G(p) = \left(\frac{1-\gamma p}{1-\gamma p^*}\right)^{1-R} G(p^*)$.

Substituting for the optimal consumption we find that in the continuation region, $p_* < p < p^*$,

$$\begin{aligned} 0 &= \beta\left[G(p) - \frac{pG'(p)}{1-R}\right]^{1-1/R} + LG(p) \\ &= \beta\left[G(p) - \frac{pG'(p)}{1-R}\right]^{1-1/R} - \beta G(p) + \mu\left[(1-R)pG(p) + p(1-p)G'(p)\right] \\ &\quad + \frac{\sigma^2}{2}\left[-R(1-R)p^2G(p) - 2Rp^2(1-p)G'(p) + p^2(1-p)^2G''(p)\right]. \end{aligned} \tag{A.4}$$

Now we can see the merit of the factor $\left(\frac{R}{\beta}\right)^R$ in the definition of V: we can divide through by β to reduce the problem to one expressed in the dimensionless quantities $\epsilon = \frac{\mu}{\beta}$ and $\delta^2 = \frac{\sigma^2}{\beta}$. (This completes the parameter reduction; the original parameters μ, σ, β, R, λ, γ have been replaced by ϵ, δ, R and ξ.)

Set $h(p) = \text{sgn}(p(1-p))|1-p|^{R-1}G(p)$ and define $w(h) = p(1-p)\frac{dh}{dp}$. Then away from 0 and 1,

$$\frac{dh}{dp} = \text{sgn}(p(1-p))|1-p|^{R-1}\left[G'(p) + (1-R)\frac{G(p)}{1-p}\right]$$

and

$$w(h) = \text{sgn}(p(1-p))|1-p|^{R-1}\left[p(1-p)G'(p) + (1-R)pG(p)\right]. \tag{A.5}$$

Moreover,

$$w(h)\frac{d}{dh}w(h) = p(1-p)\frac{dh}{dp}\frac{d}{dh}w(h) = p(1-p)\frac{d}{dp}w(h)$$

and on differentiating (A.5) we find for $p \notin \{0,1\}$

$$\begin{aligned} &w(h)w'(h) \\ &= |1-p|^{R-1}\text{sgn}(p(1-p))\left[p^2(1-p)^2G''(p) + p(1-p)(1-2Rp)G'(p) + (1-R)p(1-Rp)G(p)\right]. \end{aligned}$$

Then

$$\left[-R(1-R)p^2G(p) - 2Rp^2(1-p)G'(p) + p^2(1-p)^2G''(p)\right] = w(h)[w'(h)-1]|1-p|^{1-R}\mathrm{sgn}(p(1-p)).$$

Also using (A.5) we have

$$G'(p) = \frac{w(h)}{|p||1-p|^R} - (1-R)\frac{G(p)}{1-p}$$

and it follows that

$$G(p) - \frac{pG'(p)}{1-R} = |1-p|^{-R}\mathrm{sgn}(p)h(p)\left(1 - \frac{w(h)}{(1-R)h}\right).$$

As consumption must be non-negative this expression must be positive so we can write it as $G(p) - \frac{pG'(p)}{1-R} = |1-p|^{-R}|h||1 - \frac{w(h)}{(1-R)h}|$ and then

$$\left(G(p) - \frac{pG'(p)}{1-R}\right)^{1-1/R} = |1-p|^{1-R}|h|^{1-1/R}\left|1 - \frac{w(h)}{(1-R)h}\right|^{1-1/R}.$$

Cancelling factors of $|1-p|^{1-R}$ and dividing by $\mathrm{sgn}(p(1-p)) = \mathrm{sgn}(h)$, (A.4) becomes

$$0 = h|h|^{-1/R}\left|1 - \frac{w(h)}{(1-R)h}\right|^{1-1/R} - h + \epsilon w(h) + \frac{\delta^2}{2}w(h)\left[w'(h) - 1\right],$$

and with $w(h) = (1-R)hW(h)$,

$$\frac{\delta^2}{2}(1-R)^2hW'(h)W(h) = -|h|^{-1/R}\left|1 - W(h)\right|^{1-1/R} + \ell(W(h)).$$

Then setting $N = W^{-1}$ we find

$$\frac{1}{N(q)}\frac{dN(q)}{dq} = \frac{\delta^2}{2}(1-R)^2\frac{q}{\ell(q) - |N(q)|^{-1/R}|1-q|^{1-1/R}}.$$

Finally set $n(q) = |N(q)|^{-1/R}|1-q|^{1-1/R}$. Then $n > 0$ and

$$\frac{n'(q)}{n(q)} = \frac{1-R}{R(1-q)} - \frac{1}{R}\frac{N'(q)}{N(q)}.$$

In particular, n solves $n' = O(q,n)$ where O is as given by (3.4) for all values of $q \in [q_*, q^*]$ (except perhaps at the singular points $q = 0$ and $q = 1$).

Consider now the boundary conditions. Under the candidate optimal strategy, the impact of a change in portfolio from a point outside the no-transaction wedge to a point on the boundary does not affect the value function. Hence for $-\frac{1}{\lambda} \le p \le p_*$ we have $G(p) = A_*(1+\lambda p)^{1-R}$ for some constant A_*. (We calculated A_* in the discussion after (A.3).) Then $h(p) = \mathrm{sgn}(p(1-p))|1-p|^{R-1}A_*(1+\lambda p)^{1-R}$ and

$$h'(p) = (1-R)h(p)\left[\frac{1}{1-p} + \frac{\lambda}{1+\lambda p}\right] = (1-R)h(p)\frac{1+\lambda}{(1-p)(1+\lambda p)}.$$

It follows that $W(h) = \frac{(1+\lambda)p}{(1+\lambda p)}$; then $|1 - W(h)| = \frac{|1-p|}{1+\lambda p} = (\frac{A_*}{|h|})^{1/(1-R)}$. Writing $q = W(h)$ and $h = N(q)$ for $N = W^{-1}$ we have

$$n(q) = |N(q)|^{-1/R}|1-q|^{1-1/R} = A_*^{-1/R}.$$

Note that $q = W(h) = \frac{(1+\lambda)p}{(1+\lambda p)}$ can be rewritten as

$$\frac{q}{1-q} = (1+\lambda)\frac{p}{1-p} \tag{A.6}$$

which is valid for $-\frac{1}{\lambda} < p \le p_*$ or equivalently $-\infty < q < q_* = \frac{(1+\lambda)p_*}{(1+\lambda p_*)}$. A similar analysis gives $n(q) = (A^*)^{-1/R}$ for $q \in [q^*, \infty)$ where $q^* = \frac{(1-\gamma)p^*}{(1-\gamma p^*)}$. Thus, the condition of continuity of n' at the free boundaries is equivalent to $n' = 0$, which in turn means that candidate locations of the boundary can be identified with $O(q, n(q)) = 0$ or equivalently $n(q) = m(q)$.

Note that $q > 1$ is equivalent to $p > 1$ and that each of these conditions corresponds to the case of leverage (where the agent borrows to finance the position in the risky asset). Similarly $q < 0$ is equivalent to $p < 0$. These conditions corresponds to a short position in the risky asset.

From (A.6) at (q_*, p_*) and the similar condition $\frac{q}{1-q} = (1-\gamma)\frac{p}{1-p}$ at (q^*, p^*) we have

$$1+\xi = \frac{1+\lambda}{1-\gamma} = \frac{p^*}{1-p^*}\frac{1-p_*}{p_*}\frac{q_*}{1-q_*}\frac{1-q^*}{q^*}$$

and hence using $w(h) = p(1-p)\frac{dh}{dp}$ and a change of variable

$$\ln(1+\xi) = \int_{p_*}^{p^*} \frac{dp}{p(1-p)} - \int_{q_*}^{q^*} \frac{dq}{q(1-q)} = \int_{h_*}^{h^*} \frac{dh}{w(h)} - \int_{q_*}^{q^*} \frac{dq}{q(1-q)}. \tag{A.7}$$

But

$$\int_{h_*}^{h^*} \frac{dh}{w(h)} = \int_{q_*}^{q^*} dq \frac{N'(q)}{(1-R)N(q)q} = \frac{\delta^2(1-R)}{2} \int_{q_*}^{q^*} dq \frac{1}{(\ell(q) - n(q))}. \tag{A.8}$$

Further,

$$\int_{q_*}^{q^*} \frac{dq}{q(1-q)} = \frac{\delta^2(1-R)}{2} \int_{q_*}^{q^*} dq \left[\frac{1}{\ell(q) - m(q)}\right], \tag{A.9}$$

where we use $\ell(q) - m(q) = \frac{\delta^2}{2}(1-R)q(1-q)$. Hence

$$\ln\left(\frac{p^*}{1-p^*}\frac{1-p_*}{p_*}\frac{(1-q^*)}{q^*}\frac{q_*}{(1-q_*)}\right) = \int_{q_*}^{q^*} dq \frac{1}{q(1-q)}\left[\frac{n(q)-m(q)}{\ell(q)-n(q)}\right]. \tag{A.10}$$

Then the solution we want must have $\ln(1+\xi) = \int_{q_*}^{q^*} dq \frac{1}{q(1-q)}\left[\frac{n(q)-m(q)}{\ell(q)-n(q)}\right]$.

Appendix B. Proofs

Proof of Lemma 1. The point at issue is to understand solutions of $n' = O(q, n)$ which pass through $(1, m(1))$. Similar issues arises in the analysis in Choi et al. (2013), and more discussion can be found there.

We assume that $\delta^2 R < \epsilon$ and also if $R < 1$ that $\epsilon < \frac{1}{1-R} + \frac{\delta^2 R}{2}$. Then $(1-R)m'(1) < 0$ and $m(1) > 0$. We are interested in the behavior of n as it passes through the singular point $(1, m(1))$.

Let $\eta(x) = \frac{n(1+x) - m(1+x)}{\frac{1}{2}\delta^2(1-R)}$. Then the singular point is now at the origin. Then

$$\eta'(x) = -\frac{a(x,\eta)}{x^2}\eta + b(x) \tag{B.1}$$

where

$$a(x,\eta) = \frac{2}{\delta^2 R}\frac{m(1)+(1-R)(\delta^2 R-\epsilon)x+\frac{1}{2}\delta^2 R(1-R)x^2+\frac{1}{2}\delta^2(1-R)\eta}{1+x+\frac{\eta}{x}}$$

$$b(x) = -\frac{2}{\delta^2(1-R)}m'(1+x) = \frac{2}{\delta^2}\left[\epsilon-\delta^2 R-\delta^2 Rx\right].$$

We have $m(1) > 0$ and $m'(1) < 0$ whence $\lim_{x\downarrow 0} a(x,0) = \frac{2}{\delta^2 R}m(1) > 0$ and $b(0) > 0$. It can be shown that a typical solution to (B.1) passing through the origin has $\eta(x) = O(x^2)$. Hence a and b are bounded near the origin and in order to understand the behavior of n near 1 it is sufficient to understand the behavior of solutions to

$$f' = -A\frac{f}{x^2}+B, \qquad f(0)=0. \tag{B.2}$$

for positive constants A, B. It can be shown (see (Choi et al., 2013, Lemma 6.8)) that there are multiple solutions for $x \le 0$, but a unique solution for $x \ge 0$. All these solutions have $f'(0-) = f'(0+) = 0$.

This result extends to our more general case, and if $m'(1) < 0$ there is a family of solutions n_r which pass through $(1, m(1))$ but to the right of this point these solutions are identical. □

Proof of Proposition 2. They key step is to show that Σ is well-defined and monotonic. We prove the results for $R < 1$ and $\mu > 0$, the other cases being similar.

If $0 < r < q_- \wedge q_M < 1$ then $\zeta(r) < 1$ and Σ is well-defined. If $q_M > 1$ then there is a potential singularity at $q = 1$ for solutions n_r with $r < 1$. But, at $q = 1$ then necessarily $n_r(1) = m(1)$ and $n_r'(1) = m'(1)$. Hence any singularity at 1 can be removed.

As the solutions n_r cannot cross and $n_r(r) = m(r)$ and $n_r(\zeta(r)) = m(\zeta(r))$ we have

$$\frac{\partial}{\partial r}\ln(1+\Sigma(r)) = \int_r^{\zeta(r)} dq\frac{1}{q(1-q)}\frac{\ell(q)-m(q)}{[\ell(q)-n_r(q)]^2}\frac{\partial}{\partial r}n_r(q) < 0$$

Then Σ is continuous and decreasing in r. By the monotonicity in r of n_r we can deduce that $n_0(\cdot) = \lim_{r\downarrow 0} n_r(q)$ exists. Some further arguments yield that $\lim_{r\downarrow 0}\Sigma(r) = \infty$.

□

Remark 14. It is immediate from our approach that the integral equation which determines which of the family of candidate solutions of the auxiliary we want has a monotonicity property. In particular, it is immediate from our approach that Σ is strictly decreasing and has an inverse. In contrast, in the shadow-price approach (Choi et al., 2013, Remark 6.15) are not able to give a corresponding monotonicity argument for their equivalent function.

Proof of Theorem 5. As the majority of this result is contained in Choi et al. (2013) we only provide a sketch of the proof. Proofs of well-posedness for subsets of the parameter combinations can also be found in Davis and Norman (1990) and Herczegh and Prokaj (2015). The main innovations compared with Choi et al.

(2013) is that we take a classical approach via the value function and the Hamilton-Jacobi-Bellman equation, whereas Choi et al. construct a solution via the dual problem and the shadow price.

For the parameter combinations listed as leading to a well-posed problem we can construct a positive, C^1-solution n to the auxiliary problem and thence a function G and a candidate value function V^C given by $V^C(X_t, Y_t, \Theta_t, t) = e^{-\beta t}\frac{(x+y\theta)^{1-R}}{1-R}G(\frac{y\theta}{x+y\theta})$. It remains to prove that this candidate value function is the value function V of the optimal consumption/investment problem, and this can be done using a standard verification argument. See Zhu (2015) for details.

In the ill-posed case it is sufficient to exhibit a strategy which yields infinite expected utility. See Choi et al. (2013) for details. □

Proof of Lemma 8. It follows from the proof of Proposition 2 that Σ is decreasing in r. Hence q^* is increasing in ξ, and q_* is decreasing in ξ. □

Appendix C. Asymptotics for small transaction costs

Our goal is to derive higher order expansions for the locations of the upper and lower boundaries of the no-transaction wedge. We do not provide a complete proof but rather explain how an expansion method can be used to generate the leading order terms. A full proof can be constructed by a more careful analysis of the relevant differential equations.

For ease of exposition we assume $R < 1$ and $\mu > 0$ (for example, this allows us to say that solutions to $n' = O(q, n)$ are decreasing) but the general case is similar.

Define $M(r) = m(r+q_M)-m_M$, $L(r) = \ell(r+q_M)-m_M$ and $N(r) = n(r+q_M)-m_M$ and set $\Psi(r) = \frac{N(r)-M(r)}{\delta^2 R(1-R)}$. The Ψ solves an ODE and we want the solution with initial condition $\Psi(-u) = 0$ for small $u > 0$. Write Ψ_u for this solution, which is defined up to the first time $r > -u$ for which $\Psi_u(r) = 0$.

Let $\theta = r/u$ and set $\Theta(\theta) = \Theta_u(\theta) = \Psi_u(\theta u)$. Then $\Theta(-1) = 0$ and Θ solves

$$\Theta'(\theta) = -\frac{2u}{\delta^2 R}\frac{\{m_M + M(u\theta) + \delta^2 R(1-R)\Theta(\theta)\}}{(1-q_M-u\theta)\{q_M(1-q_M) + u\theta(1-2q_M) - u^2\theta^2 - 2R\Theta(\theta)\}}\Theta(\theta) - \theta u^2.$$

Note that $M(u\theta) = \frac{\delta^2}{2}R(1-R)\frac{u^2\theta^2}{2}$. We look for an expansion of Θ in u of the form $\Theta(\theta) = \sum_{k\geq 0} u^k a_k(\theta)$. Proceeding from the power series expressions we find on comparing coefficients of u^0 that $a_0'(\theta) = 0$ and hence $a_0(\theta) = 0$. Further, $a_1'(\theta) = 0$ and $a_1(\theta) = 0$. Then $a_2'(\theta) = -\theta$ (and $a_2(-1) = 0$) which has solution $a_2(\theta) = \frac{1-\theta^2}{2}$. We can easily calculate higher order terms. In particular,

$$a_3'(\theta) = -3\Delta_2 a_2(\theta)$$

and then $a_3(\theta) = -\frac{\Delta_2}{2}(\theta+1)^2(2-\theta)$. Then

$$\Psi_u(r) = \Theta(r/u) = \frac{u^2 - r^2}{2} - \frac{\Delta_2}{2}(r+u)^2(2u-r) + O(u^4)$$

Let $\tilde{\zeta}(u) = \inf\{r > -u : \Psi_u(r) = 0\}$. Then, $\tilde{\zeta}(u) = u - \eta u^2 + O(u^3)$ we find $\eta = 2\Delta_2$. Then we can calculate

$$\ln(1 + \Sigma(q_M - u)) = \Delta_1^{-1}\left\{u^3 - 3\Delta_2 u^4 + O(u^5)\right\}.$$

A derivation of (5.8) now follows in a few lines.

C.1. **The size of the no-transaction wedge in the case $\mu = \sigma^2 R$.** Suppose $\mu = \sigma^2 R$ and that the problem is well-posed for zero transaction costs. Then we have $q_M = 1$ and $m_M > 0$. We want to consider Ψ given by $\Psi(r) = \frac{n(1+r)-m(1+r)}{\delta^2 R(1-R)}$ for $r < 0$. In particular, Ψ_u solves

$$\Psi_u'(r) = -a(r, \Psi(r))\frac{\Psi_u(r)}{r^2} - r \tag{C.1}$$

subject to $\Psi_u(-u) = 0$, where $u > 0$ and

$$a(r,\psi) = \frac{2}{\delta^2 R}\frac{(m_M + M(r) + \delta^2 R(1-R)\psi)}{(1 + r + \frac{2R}{r}\psi)}.$$

We can understand solutions to this equation from the leading order term, which leads us to consider

$$\Phi_u'(r) = -\kappa\frac{\Phi_u(r)}{r^2} - r,$$

subject to $\Phi_u(-u) = 0$, where $\kappa = \frac{2m_M}{\delta^2 R}$. Note that in contrast to (B.2) the inhomogeneous term is not a constant, but rather r, and this leads to different asymptotics near 0.

This ODE has solution $\Phi_u(r) = e^{\frac{\kappa}{r}}\int_{-u}^{r} |s| e^{-\frac{\kappa}{s}} ds$. Note that $e^{\frac{\kappa}{r}} \downarrow 0$ as $r \uparrow 0$. It is easy to see that

$$\Psi_u(r) = \frac{e^{\frac{\kappa}{r}}}{\kappa}\int_{-u}^{r} |s|^3 \frac{\kappa e^{-\frac{\kappa}{s}}}{s^2} ds < e^{\frac{\kappa}{r}}\frac{u^3}{\kappa}\int_{-u}^{r}\frac{\kappa e^{-\frac{\kappa}{s}}}{s^2} ds < \frac{u^3}{\kappa},$$

and, integrating by parts,

$$\Psi_u(r) = -\frac{e^{\frac{\kappa}{r}}}{\kappa}\kappa\int_{-u}^{r} s^3\frac{\kappa e^{-\frac{\kappa}{s}}}{s^2} ds = \left[\frac{|r|^3}{\kappa} - \frac{u^3}{\kappa}e^{\frac{\kappa}{r}+\frac{\kappa}{u}} + \int_{-u}^{r}\frac{3s^2}{\kappa}e^{\frac{\kappa}{r}-\frac{\kappa}{s}} ds\right].$$

It follows that for $-1 < \theta \le 0$, $\Psi_u(u\theta) = \frac{|\theta|^3}{\kappa}u^3 + O(u^4)$.

The above results extend easily to the case where the constant κ is replaced by a strictly positive continuous function $A = A(r)$ with $A(0) = \kappa$, and with a little more work to the case in which κ is replaced by $a = a(r, \psi)$ as in (C.1).

Let $\Lambda(u) = \log(1 + \Sigma(u))$. Then

$$\Lambda(u) = \int_{-u}^{0}\frac{dr}{r(1+r)}\frac{2R\Psi_u(r)}{[r(1+r) + 2R\Psi_u(r)]} \sim u^2\int_{-1}^{0}\frac{d\theta}{\theta^2}2R\frac{|\theta|^3}{\kappa} = \frac{R}{\kappa}u^2.$$

Proposition 12 follows.