Received: 27 September 2017 | Accepted: 1 February 2020

DOI: 10.1111/mafi.12239

ORIGINAL ARTICLE

MATHEMATICAL FINANCE An International Journal of Mathematics, Statistics and Financial Economics WILEY

Shortfall aversion

Paolo Guasoni[1] | **Gur Huberman**[2] | **Dan Ren**[3]

[1]School of Mathematical Sciences, Dublin City University, Dublin, Ireland

[2]Columbia Business School, Columbia University, New York, New York

[3]Department of Mathematics, University of Dayton, Dayton, Ohio

Correspondence
Paolo Guasoni, School of Mathematical Sciences, Dublin City University, Glasnevin, Dublin 9, D09 W6Y4, Ireland.
Email: paolo.guasoni@dcu.ie

Funding information
Partially supported by the ERC (279582), NSF (DMS-14-12529), and SFI (16/IA/4443, 16/SPP/3347).

All authors have contributed equally to this study.

Abstract

Shortfall aversion reflects the higher utility loss of spending cuts from a reference than the utility gain from similar spending increases. Inspired by Prospect Theory's loss aversion and the peak-end rule, this paper posits a model of utility from spending scaled by past peak spending. In contrast to traditional models, which call for spending rates proportional to wealth, the optimal policy in this model implies a constant spending rate equal to the historical peak when wealth is relatively large. The spending rate increases when wealth reaches a model-determined multiple of peak spending. In 1926–2015, shortfall-averse spending is smooth and typically increasing.

KEYWORDS

consumption, endowments, portfolio choice, shortfall aversion

1 | INTRODUCTION

In many circumstances spending adheres to its past peak or, if savings allow, gradually exceeds it, thereby establishing new peaks. Cutting spending, however, leads to a disappointment. Intuitively, this disappointment, given the size of the shortfall, is greater than the pleasure associated with a similar increase in spending relative to the same peak in past spending. This greater sensitivity to shortfall than to gain is familiar—it is loss aversion in consumption, referred to as *shortfall aversion*. It is the focus of this paper.

Loss aversion in wealth, a feature of the one-period Prospect Theory (Kahneman & Tversky, 1979), entails a reference point relative to which losses and gains are assessed. The peak-end rule (Kahneman, 2011) inspires this paper's choice of the reference point to be at the historical peak spending. According to that rule, experience is judged largely based on its most intense point (peak) and its end point. In the words of Dante (Inferno V), "There is no greater pain, than to remember happy times in misery." In recent prose, Bowman, Minehart, and Rabin (1999) note that "a consumer may be wary of developing a luxurious lifestyle because he knows that doing so will make him less happy if he later becomes impoverished."

Assuming a specific form of utility for spending that is rooted in behavioral decision theory, this paper studies its implications for an individual's (or a household's) choice of spending, saving, and portfolio selection. The model also applies to the endowment of a university or a foundation that funds various programs with long-term spending commitments to employees and beneficiaries. Guided by their plans and expectations, these employees and beneficiaries in turn make their spending commitments. Thus, there are layers of reliance on the spending level. A spending cut leads to a waste of resources and a utility loss that is substantially larger than the utility gain associated with an increase in spending of a similar size.

Therefore, it is reasonable, even imperative, to model the benefit that an individual or foundation derives from spending as not merely increasing in spending but as increasing in spending relative to a measure of past spending. The model analyzed here defines the utility of spending as a function of current spending divided by a power of the past spending peak. This ratio is at the heart of the model.

Shortfall aversion shares a key feature with the habit-dependent preferences developed by Sundaresan (1989) and Detemple and Zapatero (1991), namely, that holding other attributes fixed, higher past consumption is associated with lower current utility. But the analogy ends here: in habit preferences, past spending determines the *subsistence* reference level above which utility is measured. With the exception of Detemple and Zapatero (1992) and Detemple and Karatzas (2003), habits are addictive, in that spending cannot equal, let alone fall below, subsistence. In contrast, in the present model, spending is never higher than the reference, which is past peak spending. This implies that shortfall aversion affects choices when spending is at its highest, whereas the strongest deviations of habit based models from the standard models take place when consumption is low. Finally, a driver of this paper's results is the discontinuity of the marginal utility at the endogenously determined past peak spending. Marginal utility is smooth in the habit-based model.

Any model of spending and investment must take a stand on the rate of time preference. Tobin (1974) asserts that "the trustees [of an endowment] are supposed to have a zero subjective rate of time preference." This assumption is adopted in the main part of this paper both because it may be substantively appropriate and because it keeps the model and its analysis clearer. The model's extension to a positive discount rate, developed in Section 9, preserves the main features of the solution, though it leads to less intuitive formulas.

Prospect Theory considers choice among one-period lotteries and posits three deviations from the traditional theory of expected utility: (a) The existence of a reference level of wealth relative to which gains and losses are measured and valued. (b) Higher marginal utility from losses than from gains. (c) The weights of the different prospects are nonlinear functions of their probabilities, in contrast to the usual expected value operator.

Adopting (a) and (b), the model presented here has an internally determined reference point, namely the highest spending rate to date, and higher marginal utility for a reduction of the rate of spending below the reference point than an increase above it. A spending increase above the reference point establishes a new, higher reference point that adversely affects the utility from future spending.

1.1 | A sketch of the model

A specification of the classic Merton (1969, 1971) spending–investment problem is a convenient point of departure. It envisions a decision maker who continuously distributes wealth between current spending and saving (for future spending), which he allocates between a safe and a risky asset. There are no additional income sources (or claims on the wealth) and therefore no hedging motive to affect the spending–investment policy. His objective is to maximize the expected utility of the open-ended spending path. A version of the problem with a rate of zero time preference and a single risky asset is to

$$\max_{c,\pi} \mathbb{E}\left[\int_0^\infty \frac{c_t^{1-\gamma}}{1-\gamma} \mathrm{d}t\right] \tag{1}$$

over the consumption (c) and investment (π) policies, subject to a standard budget constraint.

This problem has a single preference-dependent parameter: the relative risk aversion γ. (The other parameters are market parameters of the joint return distribution of the safe and the risky asset.) The problem is solvable with a zero rate of time preference if $\gamma > 1$. (Without this further specification, prolonged periods of frugality followed by spending binges can achieve infinite welfare.)

According to the solution of the Merton model spending is proportional to wealth, as is the fraction of wealth invested in the risky asset. Explicitly, spending at time t equals to:

$$\hat{c}_t = m\hat{X}_t, \tag{2}$$

and the *Merton consumption fraction*

$$m = \left(1 - \frac{1}{\gamma}\right)\left(r + \frac{\mu^2}{2\gamma\sigma^2}\right), \tag{3}$$

where $\hat{X}_t$ is wealth at time t, r is the safe rate of return, μ is the excess return on the risky asset, and σ is the standard deviation of that return. The fraction of wealth in the risky asset, also constant, is the *Merton risky weight*

$$\pi = \frac{\mu}{\gamma\sigma^2}. \tag{4}$$

The novelty of the present model is that the instantaneous utility depends on spending relative to a function of a target that is past peak spending (Figure 2). Namely, the decision maker's objective is

$$\max_{c,\pi} \mathbb{E}\left[\int_0^\infty \frac{\left(c_t/h_t^\alpha\right)^{1-\gamma}}{1-\gamma} \mathrm{d}t\right] \tag{5}$$

subject to the usual budget constraint, which dictates that current savings finance future spending, without other income sources or uses in the future. The variable h_t is the target relative to which current spending is enjoyed. It equals the maximum past spending rate

$$h_t = \max\left\{\bar{h}, \sup\nolimits_{s\in[0,t]} c_s\right\}, \tag{6}$$

where $\bar{h}$ is the initial target, which represents the status quo inherited by the decision maker. (If no initial target is given, it is sufficient to consider it to be zero. An optimally behaving decision maker immediately increases it.)

The power α is the degree of shortfall aversion. It ranges between zero and one; the case $\alpha = 0$ corresponds to no shortfall aversion, which is the Merton model (1). With $\alpha > 0$, and with spending at the target level, the marginal utility of spending is higher for a slight cut in spending than for a slight increase in spending, which is how shortfall aversion is brought to bear. The ratio between the marginal utility of a spending increase and a spending cut at the target level (and only there!) is $1 - \alpha$. In an experiments-based study, Tversky and Kahneman (1992) suggest that a close analogue to α is about $1/2$.

At first glance, the modeling choice (5) and especially (6) seem to imply a commitment to an outlier in spending even if it took place in the very distant past, as the past peak enters the utility function

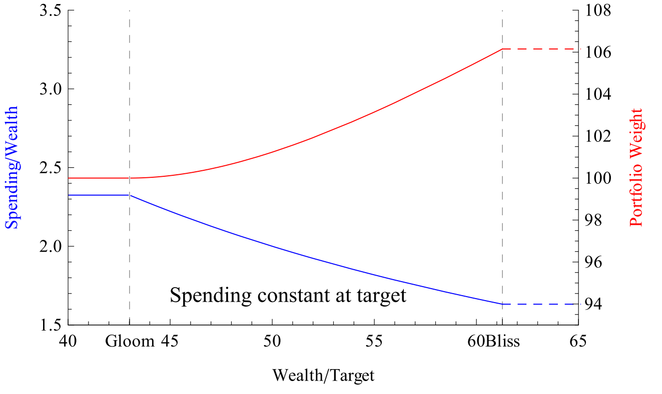

FIGURE 1 Ratio of optimal spending to wealth and portfolio weight of risky asset as functions of the wealth to target ratio, for market parameters the equity premium $\mu = 8\%$, the equity volatility $\sigma = 20\%$, the safe rate $r = 0.65\%$; and for endowment parameters risk aversion $\gamma = 2$ and shortfall aversion $\alpha = 0.5$. (Note that the region on the right of the bliss point is never visited, as the target spending rises immediately to restore the wealth/target ratio to the bliss level) [Color figure can be viewed at wileyonlinelibrary.com]

permanently. This concern is misplaced because the equilibrium spending path derived below is such that when current spending (the status quo) is below past peak spending—regardless how distant it is—then the equilibrium spending and investment policies are those of the Merton model. For past peak spending to affect spending and portfolio choices (which is when wealth is sufficiently high), the current spending is exactly at its historical peak. In fact, current spending is at the historical peak a substantial fraction of the time. (This fraction approximately equals α.)

1.2 | A sketch of the solution

When choosing current spending and saving allocations, a forward looking planner takes into account the impact of current decisions on future benefits through a budget constraint (the standard impact) and also (and crucially) through the possibility that current spending sets a new spending peak and thereby changes the target h that affects the utility derived from future spending.

With preference and market parameters fixed, the decision maker's choice depends at each point on the ratio of wealth to the target. Figure 1 summarizes the main attributes of the closed-form solution, namely the optimal spending and investment rule. It depicts the evolution of the spending rule and of the portfolio weight of the risky asset as a function of the ratio of the wealth to target.

Figure 1 shows two regions and two boundary points. These points correspond to the gloom and bliss wealth to target ratios. The gloom ratio is $g = 1/m$, the inverse of the Merton consumption fraction (3). The bliss ratio is higher and given in (14) below. The target is constant, except at the bliss point, where it increases to match (stochastic) increases in wealth if they take place. The inverse of the bliss point is the *lowest* spending rate as a fraction of wealth.

When the wealth to target ratio is at g $(= 1/m)$ or lower, it is optimal to follow the Merton prescription for spending being proportional to wealth and the portfolio weight of risky asset being a constant, with both parameters equal to those derived by Merton. In this region wealth is so low relative to the established target that the optimal solution is unaffected by the discontinuity in the marginal utility at the point where spending is equal to the target. The inverse of the gloom point is the *highest* spending rate as a fraction of wealth.

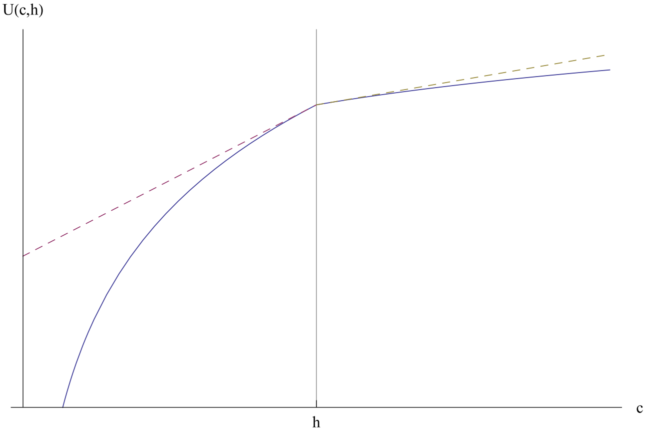

FIGURE 2 Utility from a spending rate c, when the target rate is h [Color figure can be viewed at wileyonlinelibrary.com]

When the wealth to target ratio is higher than g, and as long as it has not reached the bliss point, the optimal behavior entails constant dollar spending that is equal to the established target, and increasing the portfolio weight of the risky asset as the wealth to target ratio increases. This is the normal or target region.

The intuition underlying behavior in this region is straightforward: concern about establishing too high a target that will affect utility from future spending keeps spending from rising above the established target. The response to wealth changes is all in the portfolio weight of the risky asset. It increases with wealth because more wealth makes the target spending rate easier to preserve, thereby making the risky asset more attractive.

These two features of the spending and investment policy, which prevail approximately a fraction α of the time (in a sense to be made precise in Theorem 5.1), are in contrast with the Merton model. Spending is insensitive to changes in wealth in that region, whereas the fraction invested in the risky asset increases if wealth increases.

Behavior at the bliss point is sensitive to the direction of wealth change. Deterioration of wealth takes the decision maker to the normal region and the behavior described above. In contrast, a wealth increase calls for an increase in spending (and thereby an increase in the target), so as to maintain the wealth to target ratio. At this point, the decision maker is so wealthy that the utility from immediate spending overwhelms concerns about the target being too high in the future. Thus, the optimal policy prevents any movement to a wealth to target ratio greater than the bliss ratio. At that point, the optimal portfolio weight of the risky asset is unaffected by wealth increase.

In summary, shortfall aversion induces two spending regimes: in good times (when wealth is between the gloom and bliss ratios) spending is constant at the target level, while in bad times (when wealth is below the gloom ratio) spending is proportional to wealth (in particular, it declines with wealth), so that spending is never higher, as a fraction of wealth, than at the gloom ratio. Vice versa, spending increases above the target when wealth reaches the bliss ratio, so that this ratio is never exceeded.

1.3 | A sketch of the intuition

The classical condition that the marginal utility of optimal spending must match the marginal value of wealth or saving is essential. This condition is more complex in good times, where spending is at the target level and therefore the marginal utility of its increase is lower than that of its decrease. In this region the marginal value of wealth is greater than the marginal utility of spending increase and lower than the marginal utility of spending decrease. Therefore, in that region a small addition to wealth is best used by adding to saving rather than to spending, whereas it is best to subtract from saving rather than from spending a small loss of wealth.

At the boundary between good times and bad times, spending is equal to the target level but its increase and decrease have different implications for its marginal utilities. A decrease in spending means a shift to the bad times region where the marginal utilities of spending increase and decrease coincide, and are equal to the marginal value of wealth. Therefore, in that region any change in wealth is allocated between savings and spending in a way that preserves the equality of their marginal utilities. A wealth increase from the boundary point between the good and bad times regimes means a shift to the good times regime.

At the higher end of the target (good times) region, the marginal utility of an increase in saving is (infinitesimally) lower than the marginal utility of increase in spending (and in the target), and therefore increasing spending is the desired action, rendering the marginal utility of wealth and of spending equal.

1.4 | Applications

Shortfall aversion of the beneficiary of a fund is key to the application. Shortfall aversion may be rooted in preferences, as suggested by the experimental evidence of Kahneman and Tversky, or associated with irreversible investments made in reliance on future spending of the fund's income.

A simple straightforward example is a trust fund to support a recipient into the indefinite future. It requires coherent spending–investment rules. The model studied here lays out the rules for a shortfall-averse recipient in the special case that the endowment and its investment return are the only income sources.

Another simple example is an endowment set up to support various causes that are autonomous and make plans and commitments for multiple years. At the margin, scaling back these plans is more costly than expanding the plans. Therefore, the endowment is shortfall averse. The model offered here suggests how shortfall aversion affects its spending and investment policies.

Yet another possible application is to the payout and investment policy of trust funds set up by wealthy individuals for the benefits of generations of their off-springs. A simple trust has its endowment as its sole source of spending and its spending plans are insensitive to the fortune (or lack thereof) of the beneficiaries.

Interpreting an infinite horizon model as a good approximation to a finite but long horizon model, one can adopt it to approximate the optimal spending and investing plan of a retiree in the early retirement years, if the retiree is shortfall averse. This paper's premise is that shortfall aversion is quite prevalent.

Beyond the applications suggested here, the paper proposes a way to model shortfall aversion and its initial analysis. The present model is normative rather than positive. It does not lend itself to aggregation because variation in wealth or history, let alone preferences, populates the model with individuals who cannot be summarized by a representative agent. Therefore, as it stands, the model is not designed to address questions in asset pricing. Guasoni and Huberman (2016) adopt a representative agent framework to study asset pricing implications of the preferences introduced here.

The next section offers a review of the literature and the one following it presents the model. The closed-form solution and its main properties are in Section 4; long-run properties of the solution are in Section 5. Section 6 entertains the possibility that a shortfall-averse agent would apply the Merton spending–investment policy and concludes that for reasonable parameter values the loss would be the equivalent of 20% or more of initial wealth. Section 7 offers an illustration of the solution and its properties applied to the 1926–2015 market data. It suggests that low risk aversion and high shortfall aversion deliver high spending growth along a smooth path. An informal derivation of the solution is in Section 8, while Sections 9 and 10 consider extensions of the model to a positive discount rate and a finite horizon, respectively. Section 11 concludes. All the proofs are in the Appendix.

2 | LITERATURE REVIEW

Merton (1969) is the classic model of dynamic spending–investment under uncertainty. Assuming a constant relative risk aversion utility of instantaneous spending, its time separability and asset prices following geometric Brownian motion, he argues that spending is a fixed proportion of wealth and that portfolio weights are independent of wealth. Within a similar analytic framework, Merton (1993) explicitly addresses the spending–investment problem of university endowments. He allows for additional income sources and spending needs whose evolution is governed by a Brownian motion. The solution is an adaptation of the earlier work to this more general case.

Building on Ryder and Heal (1973), Sundaresan (1989) considers a utility for instantaneous consumption that depends on the difference between current consumption and habit, defined as a receding weighted average of past consumption. Constantinides (1990), Detemple and Zapatero (1991, 1992), and Campbell and Cochrane (1999) build on, and extend Sundaresan's original construct. (See also Detemple & Karatzas, 2003 on internal habits and Menzly, Santos, and Veronesi, 2004, Santos and Veronesi, 2010 on external habits.) A key distinction between the present model and models of habit formation is that shortfall aversion features discontinuity of the marginal utility for consumption at the target level, whereas preferences are smooth in habit-based models. Furthermore, habit models depart from standard preferences in bad times, while shortfall aversion implies a different behavior in good times.

Loomes and Sugden (1982) and Bell (1982) propose a theory of regret, which incorporates reference dependence by adding to utility a regret–rejoice function that captures the effect of comparing an outcome with an alternative reference. Under appropriate assumptions on the joint distribution of an outcome and its alternative (typically, independence), they show that regret theory can resolve common violations of the expected utility paradigm, such as the Allais paradox. For a recent perspective on regret theory, see Bleichrodt and Wakker (2015).

Koszegi and Rabin (2006, 2007, 2008, 2009) develop a theory of reference-dependent preferences. The theory emphasizes consumption, tension between consumption and beliefs about consumption and possible tension between earlier and later beliefs about consumption. These tensions give rise to a gain–loss utility that is at the core of that work. Koszegi and Rabin do not address a Merton-like problem of consumption and investment. Pagel (2017) builds on their work to study a dynamic model of life cycle consumption. Consumption comes from uncertain labor income and from saving at a safe rate. Pagel (2018) further extends the model to allow for investment in a risky asset and derives the savings rate and risky asset portfolio weight when news regarding one's consumption affects one's well-being.

Reference dependence is also prominent in models of disappointment aversion (Gul, 1991), which specify asymmetric responses (disappointment and elation) to outcomes above and below an endogenously determined certainty equivalent, and have been employed to explain variability in portfolio holdings (Ang, Bekaert, & Liu, 2005) and countercyclical effective risk aversion (Routledge & Zin, 2010). Blavatskyy (2018) shows how the combination of disappointment aversion with sensitivity to expected skewness helps explain a number of behavioral patterns that are inconsistent with expected utility. Shortfall aversion differs from the approaches above in that its reference point is determined intertemporally as the past peak consumption, and therefore cannot be exceeded. When a new peak is established, it does increase utility, but it also sets a higher bar against which future consumption is judged.

Similar to the present paper, also van Bilsen, Laeven, and Nijman (2019) study the optimal spending and investment paths under loss aversion in spending, but model loss aversion differently and therefore derive a very different solution. In their approach, utility from consumption depends on the difference between the consumption rate and a reference rate, the dependence being sensitive to the sign of the difference. The spending–investment model at the heart of the present paper also formalizes loss aversion in spending through reference-dependent preferences. The reference enters multiplicatively and is equal to past peak spending, not expectations about future spending. Thus, the reference is *actual* experience rather than expectations about future experience.

Deusenberry (1949) suggests a ratchet effect of consumption, that is, that consumption is sensitive to its own path. Following an expansion, aggregate income decline is accompanied by a lesser decline in consumption, at the expense of a lower savings rate. Dybvig (1995, 1999) considers a close problem to the one studied here, namely the Merton problem (1) under the intolerance for any decline in the standard of living, that is, that spending cannot decrease over time. This condition in turn implies a spending rate always below the safe rate, as this is the maximum spending rate that can be sustained with certainty. Thillaisundaram (2012) generalizes Dybvig's setup by allowing spending drawdowns. As in Dybvig's, such an approach enforces a lower bound in spending as a fraction of its past maximum, at the price of holding a largely safe portfolio that guarantees such a bound but hinders growth. Similar to our approach, it becomes optimal to hold spending constant before increasing it. The present model, in contrast, acknowledges that a spending shortfall is painful, but does not rule it out. Accordingly, it implies realistic target spending rates well above the safe rate, without encumbering the portfolio with safe holdings to finance a hard floor in spending.

Shortfall aversion entails desirability of sustainability of the spending level but does not guarantee it. The degree of shortfall aversion α captures the effect of the past's peak spending on the utility from today's spending, but does not force the optimal solution to deliver a nondecreasing spending path.

Findings of studies of subjective well-being are consistent with the way preferences are modeled here. De Neve et al. (2018) hypothesize higher individual sensitivity to losses than to gains in economic growth. They report that various polls on measures of subjective well-being are consistent with this hypothesis. Stutzer (2004) considers a survey-based measure of life satisfaction and, based on over 4,000 responses, reports that the measure is negatively correlated with income aspiration that itself is positively correlated with past income levels, especially if they were higher. These results offer empirical support to an interpretation of the preferences modeled here as preferences of people. It is also reasonable to interpret the preferences as those of institutions, especially endowments.

In contrast with the voluminous literature on spending and investment choice, not much has been written on the similar challenge facing endowments. Tobin (1974) is an early piece, titled "What Is Permanent Endowment Income?" It opens, "The trustees of an endowed institution are the guardians of the future against the claims of the present. Their task is to preserve equity among generations. The trustees of an endowed university like my own assume the institution to be immortal. They want to know, therefore, the rate of consumption from endowment that can be sustained indefinitely. Sustainable consumption is their conception of permanent endowment income. In formal terms, the trustees are supposed to have a zero subjective rate of time preference."

Gilbert and Hrdlicka (2012) is an investigation into optimal policies of endowments, focused on the notion of fairness in a stochastic environment. Observing that efficiency considerations combined with attractive but risky rates of return may lead to a preference of future constituents at the expense of the current generation, the authors argue that an increasing preference for stochastic fairness reduces the allocation of endowment assets to risky assets, leading to lower payout rates approaching the safe rate. Gilbert and Hrdlicka (2015) study the appropriate university endowment objective function in the presence of stakeholders with diverse objectives and agency frictions. In contrast, the present paper adheres to the more traditional approach of maintaining tractability, at the expense of not incorporating these issues.

Brown, Dimmock, Kang, and Weisbenner (2014) is an empirical study of the sensitivity of the spending policies of university endowments to changes in the endowments' wealth levels, the changes being mostly due to market returns. They note that following disappointing returns, university endowments tend to spend less than their stated policies. A structural model informed by this paper's analysis may help shed further light on the actual spending and investment choices of university endowments.

3 | THE MODEL

3.1 | Preferences

A decision maker chooses a spending plan $(c_t)_{t\geq 0}$ to maximize the expected utility over an infinite horizon,

$$\mathbb{E}\left[\int_0^\infty U(c_t, h_t)\mathrm{d}t\right].$$

The utility function depends both on current spending c_t and on the *target* h_t, which is the maximum past spending level: if the endowment starts at time zero with a target of $\bar{h}$, at time t the target satisfies $h_t = \max\{\bar{h}, \sup_{0\leq s\leq t} c_s\}$. In particular, utility increases in consumption c but decreases in the target h. Thus, increasing spending above the target implies the simultaneous reset of such target to a new high, which detracts in part from the increased utility. Shortfall aversion is thus defined as the relative deficit of the marginal utility of an increase in spending above the target from the marginal utility of a decrease below the target:[1]

$$SA(h) := 1 - \frac{dU(h,h)/dh}{\partial U(c,h)/\partial c|_{c=h}}. \tag{7}$$

In the concrete setting of isoelastic preferences considered in this paper, define the utility from spending c and target h as

$$U(c,h) = \frac{(ch^{-\alpha})^{1-\gamma}}{1-\gamma}, \tag{8}$$

where $\gamma > 1$ is the usual relative risk aversion $-c\frac{\partial^2 U(c,h)}{\partial c^2}/\frac{\partial U(c,h)}{\partial c}$, and $\alpha \in (0,1)$ is the constant shortfall aversion in (7).[2] Rewrite (8) as

$$U(c,h) = \begin{cases} \frac{(ch^{-\alpha})^{1-\gamma}}{1-\gamma} & c \leq h \\ \frac{c^{(1-\alpha)(1-\gamma)}}{1-\gamma} & c > h \end{cases}. \tag{9}$$

Note that the second expression follows from the first one, as a choice of c above the target h instantly resets h to be equal to c. The risk aversion associated with spending reduction at h is γ. It is natural to consider the risk aversion associated with an increase in spending above h as $\gamma^* = 1 - (1-\alpha)(1-\gamma)$. It is the α-weighted average of one and γ,

$$\gamma^* = \alpha + \gamma(1-\alpha),$$

and it is always lower than γ because $\gamma > 1$.

Equation (9) implies, in particular, that at the target spending rate $c = h$, the marginal utility is $(1-\alpha)h^{-\gamma^*}$ for an increase in spending, whereas it is $h^{-\gamma^*}$ for a decrease. Therefore, when spending is equal to the target, the marginal utility from cutting spending is higher than the marginal utility from increasing spending by a factor of $1/(1-\alpha)$. This discontinuity in the marginal utility when spending equals the target is the key to the distance between the present model's recommendations and those of the earlier work. When $\alpha = 0$ the marginal utilities are equal and the model is identical to the traditional models, and when $\alpha = 1$ the ratio of the two marginal utilities is infinity, and increasing spending above the target has no advantages.

Discontinuity in marginal utility at a reference point is a key feature of Kahneman and Tversky's (1979, 1992) Prospect Theory, where it is called loss aversion. Much of that work of Kahneman and Tversky and the voluminous follow-up work are about aversion to losses of wealth, which presumably are eventually tied to losses in consumption. The present model is about shortfall in spending. The location of the reference point is always a challenging and delicate issue in models based on prospect theory. The choice here is that the reference point, that is, the target, is endogenous and is equal to peak past spending.

Shortfall aversion is irrelevant for monotone spending plans. If spending c_t is increasing (and $\bar{h} = c_0$), then $h_t = c_t$ for all $t \geq 0$, and its utility is the same as in a standard model with risk aversion γ^*. Likewise, if spending is decreasing, then $h_t = \max\{\bar{h}, c_0\}$, and the utility is again the same as in a model with the same risk aversion γ. In particular, in a deterministic setting, where spending is increasing in view of positive interest rates (cf. (3) with $\mu = 0$), shortfall aversion is inconsequential. But it plays a central role in the present model because the endowment can finance its spending with a mix of safe and risky investments.

3.2 | A two-asset investment opportunity set

The financial market includes a safe asset with a fixed interest rate $r \geq 0$ and a risky asset the price of which follows a Brownian motion with excess expected returns μ, and volatility σ. The Brownian motion is defined on a filtered probability space $(\Omega, \mathcal{F}, P)$; $(\mathcal{F}_t)_{t\geq 0}$ is the augmented natural filtration of W. The return on the risky asset satisfies

$$\frac{dS_t}{S_t} = (\mu + r)dt + \sigma dW_t.$$

At each time, the endowment chooses both the spending rate c_t and the fraction of remaining unspent wealth invested in the risky asset π_t. The self-financing condition requires that, with an initial capital $X_0 = x$, total wealth X_t satisfies the dynamics:

$$\mathrm{d}X_t^{c,\pi} = (rX_t^{c,\pi} - c_t)\mathrm{d}t + X_t^{c,\pi}\pi_t(\mu\mathrm{d}t + \sigma\mathrm{d}W_t). \tag{10}$$

The dynamics (10) motivates the definition of admissible spending–investment policies as follows.

Definition 3.1. An admissible strategy is a pair of adapted processes (c_t, π_t), such that $\int_0^t c_s ds < \infty$ and $\int_0^t \pi_s^2 ds < \infty$ a.s. for all $t \geq 0$, and the corresponding wealth process $X^{c,\pi}$ in (10) satisfies $X_t^{c,\pi} \geq 0$ a.s. for all $t \geq 0$.

Denoting the class of admissible strategies by $\mathcal{A}$, the spending–investment problem defines the value function

$$V(x, \bar{h}) = \sup_{(c,\pi)\in\mathcal{A}} \mathbb{E}_{x,\bar{h}}\left[\int_0^\infty U(c_t, h_t)\mathrm{d}t\right], \tag{11}$$

where $\mathbb{E}_{x,\bar{h}}[\cdot]$ denotes for brevity the conditional expectation $\mathbb{E}[\cdot|X_0 = x, h_0 = \bar{h}]$.

3.3 | The optimal policy with no shortfall aversion

With $\alpha = 0$, the model is the classical spending–investment problem considered by Merton, and its solution is the spending rate (2) and risky asset weight (4). Both spending rate and the amount invested in the risky asset are fixed fractions of wealth.

Spending rate being proportional to wealth implies that their relative changes per unit of time are equal, that is, $\frac{dc_t}{c_t} = \frac{dX_t}{X_t}$, and therefore the volatility of consumption equals the volatility of wealth

$$\frac{d\langle c\rangle_t}{c_t^2 dt} = \frac{d\langle X\rangle_t}{X_t^2 dt} = \frac{\mu^2}{\gamma^2\sigma^2}.$$

This implication of the benchmark model is problematic in applications, because highly stable spending is consistent only with very little risk, which in turn implies a consumption rate close to the safe rate (as $\gamma \uparrow \infty$ in (3), $c_t \approx rX_t$). In short, in the benchmark model, a stable consumption is a *small* fraction of wealth, nearly as small as the real interest rate.

As it stands, the benchmark model lacks the flexibility to allow for spending to be much smoother than wealth. But the reluctance to cut spending following wealth reductions appears universal as does avoidance of overspending following unexpected enrichment. These features are modeled through shortfall aversion.

4 | OPTIMAL SPENDING AND INVESTMENT

The introduction already sketches the main attributes of the solution of the spending–investment problem. This section treats it formally, and discusses its main implications. It is qualitatively different from the classical consumption–investment problem because of the discontinuity of the marginal utility function when spending is at the target level.

The expected return of the risky asset and its risk enter the bliss point through its *Sharpe ratio* μ/σ, which is the risky asset's expected return scaled by its risk. On the whole, investment opportunities enter the bliss–gloom ratio through the parameter

$$\rho = \frac{2r}{(\mu/\sigma)^2}. \tag{12}$$

The historical estimates from U.S. data (cf. Section 7) are $r = 0.65\%$, $\mu = 8\%$ and $\sigma = 20\%$ per year. Therefore, a reasonable value for

$$\rho = 8\%. \tag{13}$$

The gloom point g (the wealth to target ratio below which spending cuts are necessary), as a fraction of the bliss point b (the wealth to target ratio at and above which spending should increase), depends on both the preference parameters (α and γ) and on the market parameters through ρ:

$$g = b\frac{(\alpha-1)\gamma^2\rho(\rho+1)}{(\gamma-1)(1-\alpha)^{\rho+1}+(\gamma\rho+1)(\alpha(\gamma-1)(\rho+1)-\gamma(\rho+1)+1)}. \tag{14}$$

(In particular, this expression implies that $g \le b$, as the ratio $g/b \le 1$.) For typical market values, Equation (13) suggests that the gloom to bliss ratio is closely approximated by its limit for $\rho \downarrow 0$:

$$\frac{g}{b} \approx \frac{1-\alpha}{1-\alpha+\frac{\alpha}{\gamma^2}-\frac{1}{\gamma}(1-\frac{1}{\gamma})(1-\alpha)\log(1-\alpha)}.$$

This formula implies that the gloom point is a fraction of the bliss point that is rather insensitive to the exact value of market parameters, as long as the interest rate is small in comparison to the squared Sharpe ratio. This fraction depends both on risk aversion γ and on shortfall aversion α. For example, for γ close to 1, the gloom–bliss ratio approaches the constant $1-\alpha$.

Next is the paper's main result, the optimal spending and portfolio policy in closed form.

Theorem 4.1. *The optimal spending policy is:*[3]

$$\hat{c}_t = \begin{cases} X_t/b & \text{if} \quad b \le X_t/h_t \\ h_t & \text{if} \quad g \le X_t/h_t \le b \\ X_t/g & \text{if} \quad X_t/h_t \le g \end{cases} \tag{15}$$

The optimal weight of the risky asset is the Merton weight (4) when the wealth to target ratio is lower than the gloom point, $X_t/h_t \le g$. Otherwise, that is, when $X_t/h_t \ge g$, the weight of the risky asset is

$$\hat{\pi}_t = \frac{\rho\big(\gamma\rho+(\gamma-1)z^{\rho+1}+1\big)}{(\gamma\rho+1)(\rho+(\gamma-1)(\rho+1)z)-(\gamma-1)z^{\rho+1}}\frac{\mu}{\sigma^2}, \tag{16}$$

where the variable z satisfies the equation:

$$\frac{(\gamma\rho+1)(\rho+(\gamma-1)(\rho+1)z)-(\gamma-1)z^{\rho+1}}{(\gamma-1)(\rho+1)rz(\gamma\rho+1)} = \frac{x}{h}. \tag{17}$$

Equivalent expressions for $\hat{\pi}$ are

$$\hat{\pi}_t = \rho\left[\frac{h_t}{rX_t}\left(1-\frac{1}{(1-\gamma)z}\right)-1\right]\frac{\mu}{\sigma^2} = \frac{2r}{\mu}\left[\frac{h_t}{rX_t}\left(1-\frac{1}{(1-\gamma)z}\right)-1\right]$$

Remark 4.2. For ease of notation, this paper focuses on a single risky asset, but all the results extend immediately to several risky assets with a vector of expected returns μ and a covariance matrix Σ, replacing the term $(\mu/\sigma)^2$ in (3) and (12) with $\mu^\top\Sigma^{-1}\mu$, and the term μ/σ^2 in (16) with $\Sigma^{-1}\mu$. With these substitutions, all the results in the rest of the paper remain valid with multiple risky assets.

Theorem 4.1 identifies the bliss and gloom points and states the optimal solution in the three regions they define. At the gloom point and for lower wealth to target ratios, the solution is as in the benchmark case of Merton. Between the gloom and the bliss point, spending is constant (i.e., insensitive to wealth changes) and equal to the spending rate at the gloom level. The portfolio weight of the risky asset is

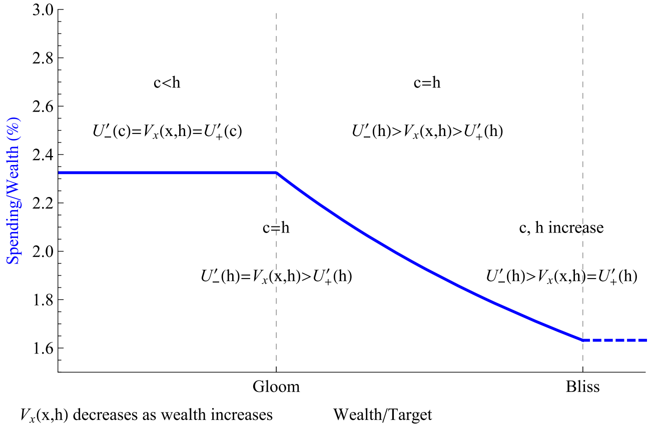

FIGURE 3 Ratio of optimal spending to wealth against the wealth to target ratio. At the left of gloom, spending is proportional to wealth (the Merton proportion). Between gloom and bliss, spending equals its past peak (the target region). Beyond bliss, spending increases instantly, reverting to the bliss point [Color figure can be viewed at wileyonlinelibrary.com]

increasing in wealth in this region. Finally, wealth to target levels higher than that prescribed by the bliss ratio never materialize; if wealth increases at that point, spending increases accordingly, thereby establishing a higher target and keeping the endowment at the bliss point until deterioration in wealth takes the endowment back to the region between the bliss and the gloom point.

Implicitly, Theorem 4.1 covers also what happens initially. Namely, if the decision maker enters with a historical target $\bar{h}$ that is sufficiently high, the theorem covers the behavior at the initial instant and afterward. If he enters with a very low target $\bar{h}$ or none, then initially he spends a fraction $1/b$ of wealth, thereby resetting the initial target to the bliss level.

Unfortunately, with the exception of the benchmark case of $\alpha = 0$, it is impossible to express $\hat{\pi}$ directly in terms of x/h. Theorem 4.1 identifies the investment policy $\hat{\pi}$ in terms of the variable z, which is related to the wealth to target ratio x/h by Equation (17). Equation (17) defines z in terms of primitives of the model. An alternative definition is

$$z = \frac{V_x(x,h)}{h^{-\gamma^*}}, \tag{18}$$

that is the marginal value of wealth $V_x(x,h)$, scaled by the marginal utility $h^{-\gamma^*}$ of target spending at the gloom point. The variable z is a decreasing function of wealth x, at the gloom point it is 1 and at the bliss point it is $1-\alpha$. For wealth levels below the gloom point it can be arbitrarily large if wealth deteriorates. On the whole, along the optimal policy, the scaled marginal value of wealth z_t is a diffusion process with a reflecting boundary at $1-\alpha$. (Compare Subsection 8.2.)

Equation (16) implies that $\hat{\pi}_t$ decreases with z. At $z = 1$ it is equal to the weight of the risky asset in the Merton portfolio, (4) and as wealth increases (and z decreases) it increases. The limiting point $z = 0$ is interesting, corresponding to the bliss point for $\alpha = 1$. At that extreme point $\hat{\pi}_t = \mu/\sigma^2$, which is the Merton risky weight for the log utility function.

A comparison of the marginal value of wealth with that of the marginal utility of spending gives rise to the intuition underlying the optimal spending-saving allocation (15). Figure 3 illustrates this intuition that is made formal at (28) below. At wealth levels below the gloom ratio, the marginal utility of wealth and that of spending are equal and the first-order condition reduces to that of Merton.

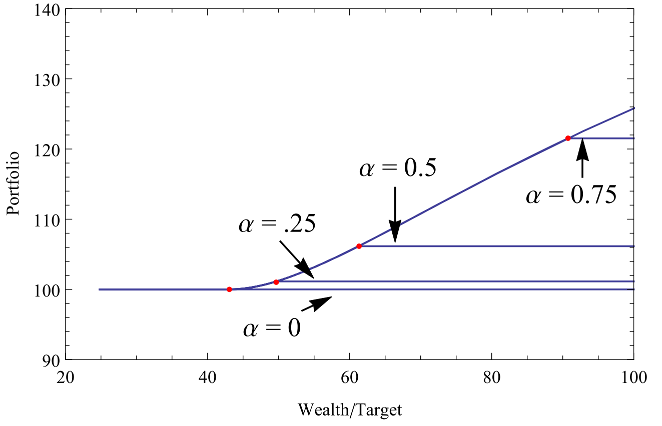

FIGURE 4 Portfolio weight (vertical axis) against wealth to target ratio, for different values of the shortfall aversion α and for risk aversion $\gamma = 2$. $\mu = 8\%, \sigma = 20\%, r = 0.65\%$ [Color figure can be viewed at wileyonlinelibrary.com]

The picture changes as soon as the ratio of wealth to target is at the gloom point because at that point and at higher wealth levels the marginal utility of spending bifurcates into two marginal utilities: A higher one for cutting spending and a lower one for increasing it. Both marginal utilities of spending are constant in the target region because spending is constant in that region. Wealth increases in that region with a movement from the gloom to the bliss ratio. At the gloom ratio the marginal value of wealth is equal to the marginal utility of cutting spending and above the marginal utility of increasing spending. An increase in wealth puts its marginal value between the marginal utilities of increasing and decreasing spending. Hence, at these points, an increase in wealth is associated with an increase in saving rather than in spending, whereas a decrease in wealth is associated with a reduction in saving rather than a reduction in spending.

As wealth increases, its marginal value decreases. At the bliss point the marginal value of wealth is so low that it is equal to the marginal utility of increasing spending. At this point an extra penny delivers higher overall utility if it is split between spending and saving, as to keep the wealth to target ratio at the same level.

The intuition underlying the behavior of the optimal weight of the risky asset, $\hat{\pi}$, parallels that of the intuition underlying the optimal spending behavior. At wealth levels below the gloom ratio, both variables follow the Merton formulas. As consumption reaches the target, the marginal utility of additional spending drops, thereby making the marginal value of risky investments more attractive and leading to a higher portfolio weight, which continues to increase as wealth moves from the gloom to the bliss ratio. At this point, the marginal value of wealth is so low that increased spending becomes attractive again, even as it reduces the utility of future consumption. Another interpretation of investing a lower fraction of savings in the safe asset is through the observation that in the target region spending is locally insensitive to the investment outcome, thereby implicitly increasing the local tolerance for risk. Figure 4 displays the behavior of the portfolio weight of the risky asset as wealth to target varies, for different levels of shortfall aversion.

Next, Theorem 4.3 summarizes the sensitivities of the solution to the parameters.

Theorem 4.3. *The following properties hold:*

1. *The gloom ratio is independent of the shortfall aversion α, and its inverse equals the Merton consumption rate in (3).*
2. *The bliss ratio (defined in (14)) increases as the shortfall aversion α increases.*
3. *Within the target region, the optimal portfolio weight of the risky asset $\hat{\pi}$ is independent of the shortfall aversion α.*
4. *At $\alpha = 0$ the model degenerates to the Merton model and $b = g$, that is, the bliss and the gloom points coincide. At $\alpha = 1$, the bliss point is infinity, that is, shortfall aversion is so strong that the solution calls for no spending increases at all.*
5. *The gloom ratio and bliss ratio both approach the asymptotic value $1/r$ when the risk aversion γ approaches infinity. In particular, if $0 \leq \rho < 1$, the gloom ratio decreases for risk aversion close to one, reaches a minimum, and then increases toward the asymptotic value $1/r$.*
   If $\rho - (1-\alpha)^{\frac{\rho+1}{2}} < 0$, the bliss ratio decreases for risk aversion close to one, reaches a minimum, and then increases toward the asymptotic value $1/r$.
   If $\rho - (1-\alpha)^{\frac{\rho+1}{2}} \geq 0$, the bliss ratio decreases to $1/r$ asymptotically.

Item (5) of the theorem is about the interaction of risk aversion and shortfall aversion. Its first assertion is based on the observation that at very high levels of risk aversion, the portfolio is almost entirely concentrated in the safe asset. Therefore, the spending rate should be close to a fraction r of savings. Beyond the first assertion the item sketches the dependence of the gloom and bliss ratios on risk aversion.

5 | LONG-RUN PROPERTIES OF THE SOLUTION

Theorem 4.1 states the closed-form solution and Theorem 4.3 discusses comparative statics. This section offers further long-run properties of the solution. Theorem 5.1 summarizes them. First, the theorem states that the average fraction of time that the decision maker expects to be in the target region is approximately α. Then the theorem considers an arbitrary starting point and states the expected time the decision maker awaits before reaching the gloom or bliss point. The starting point is described in terms of z, the scaled marginal value of wealth defined in (18) or, equivalently, in (17). This scaled marginal value ranges from $1-\alpha$ at the bliss point, to 1 at the gloom point, to any number above 1 to the left of the gloom point. Applying (17) one interprets the statement in terms of the wealth to target ratio, x/h.

Theorem 5.1.

1. *The long-run average time spent in the target region is a fraction $1-(1-\alpha)^{1+\rho}$ of the total time. (This fraction is approximately α because reasonable values of ρ are close to zero.)*
2. *Starting from $z_0 \in [1-\alpha, 1]$ corresponding to the initial wealth x and target $\bar{h}$, the expected time to reach gloom $\tau_{gloom} = \inf\{t \geq 0 : X_t/h_t = g\}$ is*

$$\mathbb{E}_{x,\bar{h}}[\tau_{gloom}] = \frac{\rho}{(\rho+1)r}\left(\log(z_0) - \frac{(1-\alpha)^{-\rho-1}\left(z_0^{\rho+1}-1\right)}{\rho+1}\right),$$

where

$$\mathbb{E}_{x,\bar{h}}[\tau_{gloom}] = \frac{\rho}{r}\left(\frac{1-z_0}{1-\alpha} + \log z_0\right) + O(\rho^2).$$

In particular, starting from bliss ($z_0 = 1-\alpha$),

$$\mathbb{E}_{x,\bar{h}}[\tau_{gloom}] = \frac{\rho}{r}\left(\frac{\alpha}{1-\alpha} + \log(1-\alpha)\right) + O(\rho^2).$$

3. *Starting from a point $z_0 \in [1-\alpha, 1]$ in the target region, the expected time to reach bliss $\tau_{bliss} = \inf\{t \geq 0 : X_t/h_t = b\}$ is*

$$\mathbb{E}_{x,\bar{h}}[\tau_{bliss}] = \frac{\rho}{r(\rho+1)}\log\left(\frac{z_0}{1-\alpha}\right).$$

In particular, starting from gloom ($z_0 = 1$), the expected time to reach bliss is

$$\mathbb{E}_{x,\bar{h}}[\tau_{bliss}] = \frac{\rho}{r(\rho+1)}\log\left(\frac{1}{1-\alpha}\right).$$

The average fraction of time in the target region being approximately α makes intuitive sense because at $\alpha = 0$ the model reduces to the Merton model in which there is no target region; in contrast, shortfall aversion dominates at $\alpha = 1$ and the decision maker expects to spend almost all the time in the target region. Both the time to reach gloom from bliss and the expected time to reach bliss from gloom increase with α, with both expected times being equal to zero for $\alpha = 0$ to both expected times being equal to infinity when $\alpha = 1$. Next, Theorem 5.2 summarizes the long-run expected growth rate of savings.

Theorem 5.2. *The long-run return on the optimal portfolio, $\tilde{r} = \lim_{T\to\infty} \frac{1}{T}\int_0^T (r + \mu\hat{\pi}_t)dt$, is*

$$\begin{aligned}\mathbb{E}_{x,\bar{h}}[\tilde{r}] =& r - \frac{\mu^2}{\sigma^2}(1-\alpha)^{\rho+1}(\rho+1)\int_{1-\alpha}^{\infty}\frac{q''(z)}{z^{\rho+1}q'(z)}\mathrm{d}z \\ =& r + (1-\alpha)^{\rho+1}\frac{\mu^2}{\gamma\sigma^2} - \frac{\mu^2}{\sigma^2}(1-\alpha)^{\rho+1}(\rho+1)\int_{1-\alpha}^{1}\frac{q''(z)}{z^{\rho+1}q'(z)}\mathrm{d}z, \end{aligned} \tag{19}$$

where the function q is defined in Lemma A.4.

To understand this result, recall that in the Merton model the expected excess return is $\frac{\mu^2}{\gamma\sigma^2}$. (The optimal weight on the risky asset is $\frac{\mu}{\gamma\sigma^2}$.) In a similar vein, the second term in Equation (19) represents the excess return of the Merton portfolio, weighted by the fraction of time that this portfolio is optimal—the time spent in the gloom region. The third term, which does not simplify further, represents the fraction of time spent in the target region (between gloom and bliss), times the corresponding excess return. This average return is higher than the one of the corresponding Merton portfolio because the exposure to the risky asset is higher in the target region.

Numerical calculations show that the long-run growth rate of savings increases with shortfall aversion. There are two sources of the extra growth rate: While the gloom regions of all decision makers are the same, the target region of the more shortfall averse contains that of the less shortfall averse, implying a more frugal spending policy. In addition, in the enlarged target region, the weight of the risky asset is higher.

A comparison across decision makers with different shortfall aversions is instructive. Consider two types with identical initial wealth and levels of risk aversion γ, but with different levels of shortfall aversion, α_1 and α_2. Equality of risk aversion guarantees that their gloom ratios are identical. The one with the higher shortfall aversion has a higher bliss ratio. Assume that they both start with no spending history, so their spending rates move immediately to their respective bliss ratios. Therefore initially, the one with the higher shortfall aversion acts more frugally, spending a smaller fraction of his savings. Moreover, it is likely that before long the gap between their target levels will increase, further enhancing the relative savings of the decision maker with the higher shortfall aversion. In addition, the more shortfall averse puts a higher fraction of his savings in the risky asset, and therefore they will grow at a higher average rate. The higher bliss ratio of the more frugal implies that he saves more and his savings will grow at a higher average rate. Thus, the more frugal eventually will be even spending more in absolute terms, while spending a smaller fraction of wealth.

6 | THE COST OF IGNORING SHORTFALL AVERSION IN THE OPTIMAL POLICY

The Merton spending investment policy is simpler and better known than the optimal policy for the shortfall averse. A natural question is how much a shortfall averse loses by applying the Merton policy rather than the optimal policy.

To address the question, consider an individual with shortfall aversion α, risk aversion γ, initial wealth x, and initial target spending rate h. His expected utility over the open-ended time horizon under the optimal policy is $U^{OP}(x, \alpha, \gamma, h)$; it is $U^M(x, \alpha, \gamma, h)$ under the Merton policy. (The expected utility $U^{OP}(x, \alpha, \gamma, h)$ is an expedient notation for the value function defined in (11).) Trivially, $U^{OP}(x, \alpha, \gamma, h) \geq U^M(x, \alpha, \gamma, h)$ and $U^{OP}(x, 0, \gamma, h) = U^M(x, 0, \gamma, h)$.

Proposition 6.1 derives in closed form the *equivalent fractional wealth loss* L, defined by the solution of $U^{OP}(x(1-L), \alpha, \gamma, h) = U^M(x, \alpha, \gamma, h)$. The solution L is the fraction of initial wealth effectively given up by a shortfall-averse decision maker, who applies the Merton spending–investment policy rather than the optimal one. It depends on α, γ, and h/x.

Proposition 6.1. *The equivalent fractional wealth loss is*

$$L = 1 + \frac{h}{mx} q'(z), \tag{20}$$

where m is the Merton consumption fraction (3) and z solves the equation

$$q(z) - zq'(z) = \frac{(x/h)^{1-\gamma}}{1-\gamma} m^{-\gamma} \left(1 + \frac{\alpha}{1-\alpha} \left(\frac{h}{mx}\right)^{1-\gamma}\right). \tag{21}$$

Table 1 shows the equivalent fractional wealth loss for a range of values of shortfall aversion and for three values of initial target spending rate scaled by the Merton spending rate. An investor with risk aversion $\gamma = 2$ and shortfall aversion $\alpha = 0.5$ (close to the value suggested by Tversky and Kahneman, 1992), who starts with a target spending equal to the Merton spending, is indifferent between (a) using the Merton policy with the initial wealth and (b) using the optimal policy with an initial wealth that is 21.4% lower. This is the cost of ignoring shortfall aversion when choosing the spending–investment strategy.

The larger shortfall aversion the bigger the loss from following the Merton policy rather than the optimal policy. On the other hand, the cost of following the Merton policy decreases as the initial target

TABLE 1 The equivalent fractional wealth loss of the initial capital (percent) from applying the Merton policy instead of the optimal policy, for different levels of shortfall aversion α and for different initial spending targets h scaled by the Merton spending rate mx

	h/x		
α	$1/b$	$m(=1/g)$	$2m$
0.1	0.6	0.6	0.3
0.2	2.8	2.5	1.3
0.3	7.5	6.3	3.3
0.4	15.7	12.5	6.7
0.5	28.2	21.4	12.1
0.6	45.1	33.1	20.4
0.7	64.6	46.8	32.5
0.8	83.5	61.3	49.6
0.9	98.8	76.1	72.5

Note. Risk aversion is $\gamma = 2$. The parameter m $(= 1/g)$ is in (3) and the relation between bliss b and gloom g is in (14). $\mu = 8\%, \sigma = 20\%, r = 0.65\%$.

spending increases relative to the Merton rate: a higher initial target implies that the optimal policy will coincide with the Merton strategy for a longer time, as wealth needs to reach a higher level before the target spending can be achieved, thereby exiting the gloom region. Yet, even for a target rate that is double the Merton spending (rightmost column in Table 1), the equivalent fractional loss is 12.1% of the initial wealth for shortfall aversion equaling 0.5, and higher for higher levels of shortfall aversion. The numbers in Table 1 suggest that the shortfall averse who follow the Merton rather than the optimal policy make a substantial error.

7 | HISTORICAL PERFORMANCE WITH MARKET RETURNS

To illustrate the analysis consider a shortfall-averse decision maker who makes spending and investment choices annually, adhering to the policies derived in Theorem 4.1 and applying the market parameters that prevailed in 1926–2015 on average, namely $\mu = 8\%$ and $\sigma = 20\%$, $r = 0.65\%$. (Data are from the Ibbotson Yearbook. The equity premium μ is estimated as average of returns on large-capitalization equities, the safe rate r as the average real return on Treasury bills.)

Equation (15) describes the fraction of wealth allocated to spending at the beginning of each year and Equation (16) describes the fraction of savings invested in the risky asset. The annual return on the safe and the risky asset are the US Treasury bill rate prevailing at the beginning of the year and the realized market return for that year, respectively.

Figure 5 summarizes the hypothetical wealth accumulation and spending under the Merton problem (the benchmark) and for a shortfall-averse decision maker. The preference parameters underlying Figure 5 are $\gamma = 2$ and α is either 0 or 0.5. Shortfall aversion manifests itself in a smoother spending path. In particular, along the spending path of the shortfall averse only in 25% of the years does spending fall below the target. In the Merton model ($\alpha = 0$) in 64% of the years spending is below its historical peak.

Table 2 summarizes five attributes of similar simulations done for various levels of risk aversion and shortfall aversion using the 1926–2015 data. The top two panels show average input variables—the average weight of the risky asset and the average spending rate. For each level of shortfall aversion,

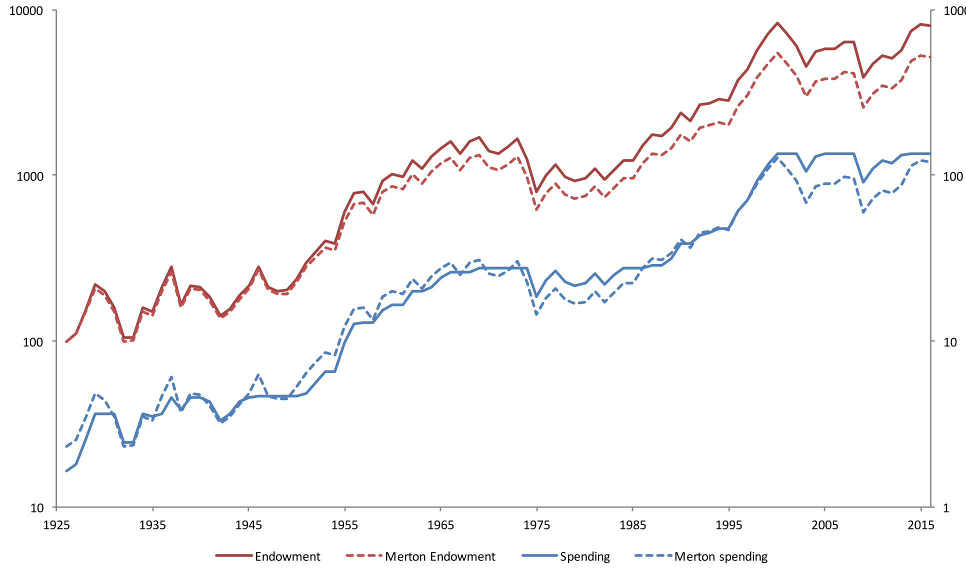

FIGURE 5 Savings (red) and annual spending rate (blue) with shortfall aversion $\alpha = 0.5$ (solid) versus the Merton benchmark of $\alpha = 0$ (dashed). Initial savings are at 100 and risk aversion $\gamma = 2$. $\mu = 8\%, \sigma = 20\%, r = 0.65\%$ [Color figure can be viewed at wileyonlinelibrary.com]

both decrease with risk aversion. In contrast, for a fixed level of risk aversion, both the average weight of the risky asset and the saving rate increase with shortfall aversion. The monotonicity of the average portfolio return in α and in γ reflects the monotonicity of the weight of the risky asset in these parameters.

High shortfall aversion and low risk aversion are associated with high savings rate and high exposure to the risky asset. This combination leads to higher savings growth rate that in the long run leads to higher spending levels. The fraction of years in which shortfall is experienced decreases with α; an extreme example is available for $\gamma = 2$ and $\alpha = 0.75$ where in 94% of the years spending does not fall below its historical peak. Moreover, conditioned on a shortfall, its magnitude decreases for higher levels of shortfall aversion.

8 | DERIVATION OF THE SOLUTION

This section sketches the control arguments underlying the derivation of the optimal spending and investment policies. The strategy is to derive the first-order conditions governing the value function, then to impose boundary and smoothness constraints informed by economic intuition, obtain the tentative closed-form solution, and finally—in the Appendix—verify rigorously that the tentative solution is indeed optimal and therefore the solution.

The first subsection derives the relevant partial differential equation (PDE) and mentions some of its properties, recalling that the utility from spending depends also on the historical peak spending. The second subsection narrows the discussion to the utility function $U(c,h) = (c/h^{\alpha})^{1-\gamma}/(1-\gamma)$ and reduces the PDE to a single ordinary differential equation (ODE) with a known solution.

Crucial in the second subsection is the observation that the utility function $U(c,h)$ retains the usual scaling property with respect to wealth, whereby doubling simultaneously wealth and spending (including the target) leads to a scaling of the utility by a constant. This property is key to arguing that the

TABLE 2 Summary statistics for spending and investments paths with the 1926–2015 realized returns for various values of risk and shortfall aversion

I: Average portfolio weight of risky asset α				
γ	**0**	**0.25**	**0.5**	**0.75**
2	100	101	103	114
5	40	40	41	46
10	20	20	20	21
II: Average spending rate α				
γ	**0**	**0.25**	**0.5**	**0.75**
2	2.3	2.2	1.9	1.5
5	1.8	1.8	1.7	1.5
10	1.3	1.3	1.3	1.2
III: Average portfolio return α				
γ	**0**	**0.25**	**0.5**	**0.75**
2	8.5	8.5	8.7	9.4
5	3.8	3.8	3.9	4.2
10	2.3	2.3	2.3	2.4
IV: Percentage of years at target spending α				
γ	**0**	**0.25**	**0.5**	**0.75**
2	36	50	75	94
5	32	46	62	80
10	23	26	32	45
V: Average shortfall when below target α				
γ	**0**	**0.25**	**0.5**	**0.75**
2	26	21	16	7
5	16	14	12	7
10	20	19	17	15

Note. $\mu = 8\%, \sigma = 20\%, r = 0.65\%$.

wealth to target ratio is an autonomous diffusion along the optimal path and deriving the analytic solution of that path.

8.1 | Considerations for general utility functions

This subsection derives the Hamilton–Jacobi–Bellman PDE ((22)–(23) below) for the value function $V(X_t, h_t)$ defined in (11). Such equation is a nonlinear PDE (24) for the value function V. However, passing to its convex conjugate $\tilde{V}$ yields a linear PDE (26), as in Karatzas, Lehoczky, Shreve, and Xu (1991). By Itô's formula, the dynamics of the value function for the wealth process X_t corresponding to the strategy π, c satisfies

$$\mathrm{d}V(X_t, h_t) = L(X_t, \pi_t, c_t, h_t)\mathrm{d}t + V_h(X_t, h_t)\mathrm{d}h_t + V_x(X_t, h_t)X_t\pi_t\sigma\mathrm{d}W_t,$$

where subscripts in x and h denote partial derivatives with respect to these variables and

$$L(X_t, \pi_t, c_t, h_t) = U(c_t, h_t) + (X_t r - c_t + X_t\pi_t\mu)V_x(X_t, h_t) + \frac{V_{xx}(X_t, h_t)}{2}X_t^2\pi_t^2\sigma^2.$$

As the process h_t increases only on the set $\{c_t = h_t\}$, the martingale principle of optimal control (Davis & Varaiya, 1973) requires that on the set $\{c_t < h_t\}$ the drift coefficient $L(X_t, \pi_t, c_t, h_t)$ is less than or equal to zero for any strategy, and equal to zero for the optimal strategy,

$$\sup_{c,\pi} L(x, \pi, c, h) = 0 \qquad \text{for } c < h. \tag{22}$$

For $c = h$, the martingale principle requires that the coefficient of dh_t vanishes,

$$V_h(x, h) = 0 \qquad \text{for } c = h. \tag{23}$$

The first step to solve the Hamilton–Jacobi–Bellman Equations (22) and (23) is to simplify the supremum in (22). Maximizing $L(x, \pi, c, h)$ with respect to π and c, and defining the convex conjugate of U as $\tilde{U}(y, h) = \sup_{c\geq 0}[U(c, h) - cy]$, Equation (22) becomes:

$$\tilde{U}(V_x, h) + xrV_x(x, h) - \frac{V_x^2(x, h)}{2V_{xx}(x, h)} \frac{\mu^2}{\sigma^2} = 0. \tag{24}$$

This nonlinear differential equation for $V(x, h)$ does not admit explicit solutions. Thus, consider the convex conjugate of the value function $\tilde{V}(y, h) = \sup_{x\geq 0}[V(x, h) - xy]$, which depends on the current target h and on the variable y, interpreted as the current marginal utility of wealth. The definition of $\tilde{V}$ implies the following relations with the value function V:

$$V_x(x, h) = y,\ x = -\tilde{V}_y(y, h),\ V_{xx}(x, h) = -1/\tilde{V}_{yy}(y, h),\ V(x, h) = \tilde{V}(y, h) - y\tilde{V}_y(y, h). \tag{25}$$

Rewriting (24) in terms of $\tilde{V}$ and its derivatives, it becomes a linear differential equation:

$$\frac{\mu^2}{2\sigma^2} y^2 \tilde{V}_{yy} - ry\tilde{V}_y = -\tilde{U}(y, h). \tag{26}$$

8.2 | Scaling properties with power utility

The conjugate value function $\tilde{V}(y, h)$ depends both on the marginal utility of wealth y and on the target spending rate h. The next step is to exploit the scaling property of the utility function $U(c, h)$ to reduce this dependence to the single variable z, which is the marginal utility of wealth scaled by the marginal utility of spending at the gloom point.

This subsection applies the power utility (9) to the convex conjugate of the utility function U in (27), and, in (29) reduces the dimensionality of the conjugate value function. The latter and a change of variables lead to the simplified version of the Hamilton–Jacobi–Bellman equation as the second-order ODE (31), with the available solution (33).

To proceed, recall (9) and calculate the conjugate utility $\tilde{U}$, which represents the maximum utility from spending, for marginal value of wealth y and target h. It is

$$\tilde{U}(y, h) = \sup_{c\geq 0}[U(c, h) - cy] = \begin{cases} \dfrac{h^{1-\gamma^*}}{1-\gamma} - hy & (1-\alpha) \leq yh^{\gamma^*} \leq 1, \\ \dfrac{\gamma}{1-\gamma} y^{1-1/\gamma} h^{-\alpha(1-\gamma)/\gamma} & yh^{\gamma^*} > 1, \end{cases} \tag{27}$$

while the $\tilde{U}(y,h)$-maximizing spending rate c equals to

$$c = \begin{cases} h & (1-\alpha) \leq yh^{\gamma^*} \leq 1, \\ y^{-1/\gamma} h^{-\alpha(1-\gamma)/\gamma} & yh^{\gamma^*} > 1. \end{cases} \tag{28}$$

In summary, the optimal spending and its utility have two different forms, depending on whether $z = yh^{\gamma^*}$ (the marginal utility or wealth relative to the marginal utility of the target's spending rate) lies in $[1-\alpha, 1]$ or in $(1,\infty)$. Further, the above power utility has the scaling property $U(\lambda c, \lambda h) = \lambda^{1-\gamma^*} U(c,h)$ for $\lambda > 0$. As in the Merton setting, this property is inherited by the value function, that is, $V(\lambda x, \lambda h) = \lambda^{1-\gamma^*} V(x,h)$, and the conjugate function $\tilde{V}$ is also homogeneous, in that

$$\tilde{V}(\lambda^{-\gamma^*} y, \lambda h) = \sup_{x\geq 0}[V(\lambda x, \lambda h) - (\lambda x)(\lambda^{-\gamma^*} y)] = \sup_{x\geq 0}[\lambda^{1-\gamma^*} V(x,h) - \lambda^{1-\gamma^*} xy] = \lambda^{1-\gamma^*} \tilde{V}(y,h),$$

and with $\lambda = 1/h$, it follows that

$$\tilde{V}(y,h) = h^{1-\gamma^*} \tilde{V}(yh^{\gamma^*}, 1). \tag{29}$$

Denoting by $z = yh^{\gamma^*}$ the scaled marginal utility of wealth, and setting

$$q(z) = \tilde{V}(z,1), \tag{30}$$

the above equation means that $\tilde{V}(y,h) = h^{1-\gamma^*} q(z)$ for some function q of the single variable z. Likewise, $\tilde{U}(y,h) = h^{1-\gamma^*} \tilde{U}(z,1)$. Exploiting these properties, the PDE (26) reduces to the following ODE, defined piecewise

$$\frac{\mu^2}{2\sigma^2} z^2 q''(z) - rzq'(z) = -\tilde{U}(z,1) = \begin{cases} z - \dfrac{1}{1-\gamma} & 1-\alpha \leq z \leq 1, \\ -\dfrac{\gamma}{1-\gamma} z^{1-1/\gamma} & z > 1, \end{cases} \tag{31}$$

and in terms of q, the optimal policy is

$$c = \begin{cases} h & 1-\alpha \leq z \leq 1, \\ hz^{-1/\gamma} & z > 1, \end{cases} \qquad \pi = -\frac{zq''(z)}{q'(z)} \frac{\mu}{\sigma^2}. \tag{32}$$

For $r \neq 0$, the general solution to Equation (31) is

$$q(z) = \begin{cases} C_1 + C_2 z^{1+\rho} - \frac{z}{r} + \frac{2\log z}{(2r+\frac{\mu^2}{\sigma^2})(1-\gamma)} & 0 \leq z \leq 1, \\ C_3 + C_4 z^{1+\rho} + \frac{\gamma}{(1-\gamma)m} z^{1-1/\gamma} & z > 1, \end{cases} \tag{33}$$

where C_1, C_2, C_3, C_4 are constants to be determined. The case $r = 0$ leads to a similar formula.

To identify the four constants in (33), four boundary conditions are needed. The first one is at $z = 1-\alpha$, the optimality condition $V_h(x,h) = 0$ in (23). Recalling that $\tilde{V}_h(y,h) = -x$, and denoting by $\hat{y}(h)$ value of y such that $V(x,h) = \tilde{V}(\hat{y},h) + x\hat{y}(h)$, it follows that

$$V_h(x,h) = \frac{\partial}{\partial h}\left(\tilde V(\hat y, h) + x\hat y(h)\right) = \tilde V_h(y,h) + \left(\tilde V_y(y,h) + x\right)\frac{\mathrm{d}\hat y(h)}{\mathrm{d}h} = \tilde V_h(y,h)$$
$$= \frac{\mathrm{d}}{\mathrm{d}h}\left(h^{1-\gamma^*} q(z)\right) = h^{-\gamma^*}\left[(1-\gamma^*)q(z) + \gamma^* z q'(z)\right],$$

where $z = yh^{\gamma^*}$. Thus, Equation (23) in terms of $q(z)$ reduces to

$$(1-\gamma^*)q(z) + \gamma^* z q'(z) = 0. \tag{34}$$

The definition of the convex conjugate of the value function $\tilde V$ implies (30), the definition of the function q, from which it naturally follows that the function $q(z)$ should be continuously differentiable, including at the point $z = 1$ where the two regions meet. The two requirements that follow are: value-matching ($q(1_-) = q(1_+)$), and smooth-pasting ($q'(1_-) = q'(1_+)$).

The fourth condition is obtained as z increases to infinity. Recall that z represents marginal utility relative to the target, and the natural condition is that marginal utility is infinite when wealth declines to zero. As $q'(z) = \frac{\tilde V_y(y,h)}{h} = -\frac{x}{h}$ by (25), the fourth condition is

$$\lim_{z\to\infty} q'(z) = 0. \tag{35}$$

As a result, the boundary conditions (34) at the bliss point $1-\alpha$ and (35) at bankruptcy, combined with the value-matching and smooth-pasting conditions at gloom, yield the four equations that identify the constants in the general form of the value function in (33).

9 | POSITIVE DISCOUNT RATE

This section extends the results in the main body of the paper to the presence of a positive discount rate, defined by the objective

$$V(x,\bar h) = \sup_{(c,\pi)\in\mathcal{A}} \mathbb{E}_{x,\bar h}\left[\int_0^\infty e^{-\beta t} U(c_t, h_t)\mathrm{d}t\right],$$

where the positive β is the subjective discount rate that summarizes the decision maker's time preference. The Hamiton–Jacobi–Bellman equation for this problem has the same form as in Equation (22) and (23), however, with a slightly changed operator L:

$$L(x,\pi,c,h) = U(c,h) + (xr - c + x\pi\mu)V_x(x,h) + \frac{V_{xx}(x,h)}{2}x^2\pi^2\sigma^2 - \beta V(x,h).$$

For the power utility (9) with $\gamma > 0$ and $\gamma \neq 1$, following the similar arguments as in the previous section, Equation (22) reduces to a similar piecewise ODE as in Equation (31):

$$\frac{\mu^2}{2\sigma^2} z^2 q''(z) + (\beta - r) z q'(z) - \beta q(z) = -\tilde U(z,1) = \begin{cases} z - \dfrac{1}{1-\gamma} & 1-\alpha \le z \le 1, \\ -\dfrac{\gamma}{1-\gamma} z^{1-1/\gamma} & z > 1, \end{cases} \tag{36}$$

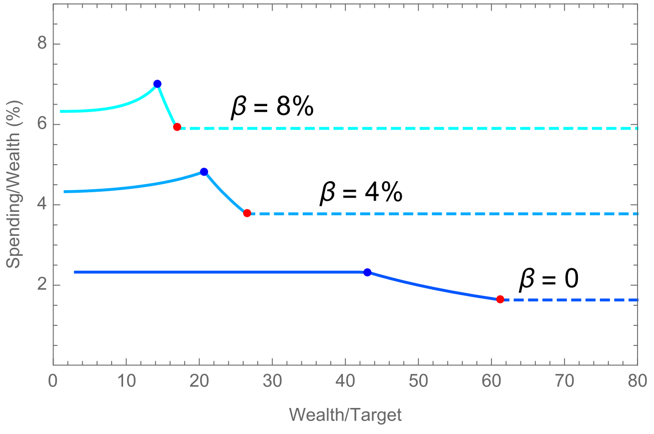

FIGURE 6 Ratio of optimal spending to wealth (vertical axis) against wealth to target ratio (horizontal), for different values of the discount rate β, with $\alpha = 0.5, \gamma = 2$. For each β value, the blue dot identifies gloom and the red dot bliss. Beyond bliss (dashed line beyond bliss), spending rises instantly to revert to bliss. $\mu = 8\%, \sigma = 20\%, r = 0.65\%$ [Color figure can be viewed at wileyonlinelibrary.com]

and in terms of q, the optimal policy is also described by (32). Complementing the solution to Equation (36) with the four boundary conditions,

$$(1-\gamma^*)q(z) + \gamma^* z q'(z) = 0, \quad q(1_-) = q(1_+), \quad q'(1_-) = q'(1_+), \quad \lim_{z\to\infty} q'(z) = 0,$$

leads to the explicit solution in Lemma A.9.

Next is the main result for the case of a positive discount rate $\beta > 0$. Note that the optimal spending and portfolio policy has the similar closed form as in Theorem 4.1, except that the values of gloom point g and bliss point b are different, as both of them depend on the discount rate β.

Theorem 9.1. *Let $q(z)$ be defined by the explicit formula in Lemma A.9. Then the optimal spending and investment policy is:*

$$\hat{c}_t = \begin{cases} X_t/b & b < X_t/h_t, \\ h_t & g < X_t/h_t \le b, \\ h_t\left[p\left(-\frac{X_t}{h_t}\right)\right]^{-1/\gamma} & X_t/h_t \le g. \end{cases} \qquad \hat{\pi}_t = \begin{cases} -\frac{(1-\alpha)q''(1-\alpha)}{q'(1-\alpha)}\frac{\mu}{\sigma^2} & b < X_t/h_t, \\ \frac{h_t p\left(-\frac{X_t}{h_t}\right)}{X_t p'\left(-\frac{X_t}{h_t}\right)}\frac{\mu}{\sigma^2} & X_t/h_t \le b, \end{cases}$$

where p is the inverse function of q', and

$$b = -q'(1-\alpha), \; g = -q'(1).$$

Figures 6 and 7 plot the optimal spending and investment policies, respectively, for different values of the discount rate β. Two features are apparent in the spending policy: First, a positive discount rate significantly reduces both the gloom and bliss ratios, as the urge to spend more and sooner overwhelms the concern about the distant future. Even a modest $\beta = 4\%$ is sufficient to halve both ratios, thereby doubling spending rates. Second, attachment to target spending becomes especially pronounced in the target region near the gloom point, leading to spending rates relative to wealth that are well above the Merton rate that prevails far below gloom.

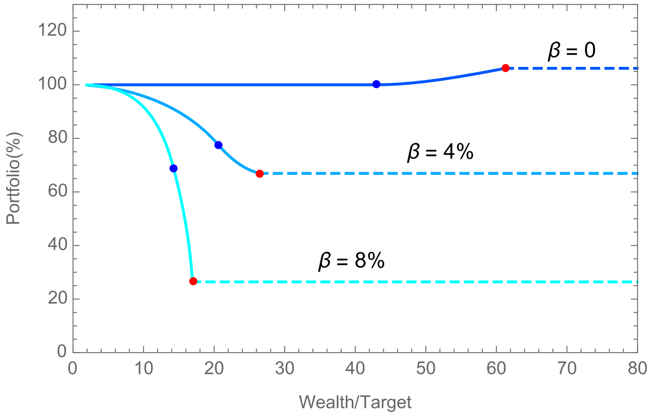

FIGURE 7 Optimal portfolio weight (vertical axis) against wealth to target ratio (horizontal), for different values of the discount rate β, with $\alpha = 0.5, \gamma = 2$. For each β value, the red point refers to the bliss point, and blue point refers to the gloom point. Beyond bliss (the dashed part beyond the red point in the figure), portfolio weight is adjusted instantly, reverting to the bliss point. $\mu = 8\%, \sigma = 20\%, r = 0.65\%$ [Color figure can be viewed at wileyonlinelibrary.com]

In particular, spending at gloom does not coincide with Merton spending (in contrast to the baseline $\beta = 0$) but is higher. The interpretation is straightforward: as current spending becomes increasingly large relative to wealth, the decision maker needs to trade off the present pain of immediate cuts against the future pain of lower spending. The higher the discount rate, the higher the reluctance to cut spending below the target, thus the propensity to forego future spending. However, when wealth deteriorates and cuts become necessary, they must be aggressive to restore the spending ratio near the Merton proportion that prevails well below gloom.

The resulting investment policy reflects the spending plan, making the decision maker more prudent in good times to budget for future target spending and to reduce the incidence of painful cuts, even if this means sacrificing growth and hence higher future spending. In contrast to perfect patience, which implies a proportion of wealth in the risky asset that increases from gloom to bliss, as the focus shifts to future growth, a positive discount rate can offset such growth motive with a stability motive: as bliss approaches, the decision maker becomes increasingly attached to the possibility of guaranteeing the current target for longer periods of time. Vice versa, the portfolio weight reverts to the higher Merton level as wealth declines and target spending becomes a more distant and less likely prospect.

10 | FINITE HORIZON

This section considers the effect of a finite horizon $T > 0$, thereby solving for the value function corresponding to the optimization problem:

$$V(x, \bar{h}, t) = \sup_{(c,\pi)\in\mathcal{A}} \mathbb{E}_{x,\bar{h}}\left[\int_t^T e^{-\beta(s-t)} U(c_s, h_s)\mathrm{d}s\right].$$

The Hamiton–Jacobi–Bellman equation for this problem is:

$$\sup_{c,\pi} L(x,\pi,c,h,t) = 0 \qquad \text{for } c < h;$$

$$V_h(x,h,t) = 0 \qquad \text{for } c = h$$

with L defined as:

$$L(x,\pi,c,h,t) = U(c,h) + (rx - c + \mu x\pi)V_x(x,h,t) + \frac{V_{xx}(x,h,t)}{2}x^2\pi^2\sigma^2 + V_t(x,h,t) - \beta V(x,h,t).$$

In particular, for power utility (9) with $\gamma > 0$ and $\gamma \neq 1$, the HJB Equations (37) and (37) become:

$$\frac{\mu^2}{2\sigma^2}z^2 q_{zz}(z,t) - (r-\beta)zq_z(z,t) + q_t(z,t) - \beta q(z,t) = \begin{cases} z - \dfrac{1}{1-\gamma} & 1-\alpha \leq z \leq 1, \\ -\dfrac{\gamma}{1-\gamma}z^{1-1/\gamma} & z > 1, \end{cases} \tag{37}$$

$$(1-\gamma^*)q(z,t) + \gamma^* zq_z(z,t) = 0, \quad \text{at } z = 1-\alpha,\ \forall t \in [0,T], \tag{38}$$

with

$$q(z,t) = h^{\gamma^*-1}\tilde{V}(y,h,t) = h^{\gamma^*-1}\sup_{x\geq 0}[V(x,h,t) - xy], (y \geq 0).$$

The optimal policy is:

$$c = \begin{cases} h & 1-\alpha \leq z \leq 1, \\ hz^{-1/\gamma} & z > 1. \end{cases} \qquad \pi = -\frac{\mu}{\sigma^2}\frac{zq_{zz}(z,t)}{q_z(z,t)}. \tag{39}$$

Using Laplace transforms, the solution to the PDE in (39) with (40) follows explicitly and is reported in the Appendix. For brevity, its lengthy derivation is omitted.

As the finite-horizon problem is time-dependent, the bliss and gloom ratios depend on the residual horizon, and are displayed in Figure 8. When the horizon is still several decades away, the bliss and gloom ratios are close to their infinite-horizon limits, which therefore are appropriate for funds charged to sustain spending programs over several generations, such as sovereign funds, endowments of universities and charitable institutions, and family trusts.

As the horizon nears, the optimal spending policy becomes increasingly driven by the limited scope for further growth, which steers spending toward $X_t/(T-t)$, that is, current wealth divided by the residual horizon. Accordingly, for short horizons the wealth-spending ratios, at both bliss and gloom, converge to $X_t/(X_t/(T-t)) = T-t$, as clear from the leftmost part of the figure. In other words, reference dependence becomes weaker as the horizon nears because the budget constraint becomes an overriding factor that tends to equalize spending over the remaining time.

Figure 9 sheds further light on the effect of the horizon on the optimal policy, and shows that the effect of a finite horizon is akin to that of a positive discount rate (cf. Figure 6). As the horizon nears, spending increases as a proportion of residual wealth, and the gap between bliss and gloom shrinks. At the same time, the decision maker is willing to increase spending well above the Merton level in the attempt to cling to the target, but once cuts become necessary, they are made swiftly. The corresponding investment policy is also similar to Figure 7 and thus omitted.

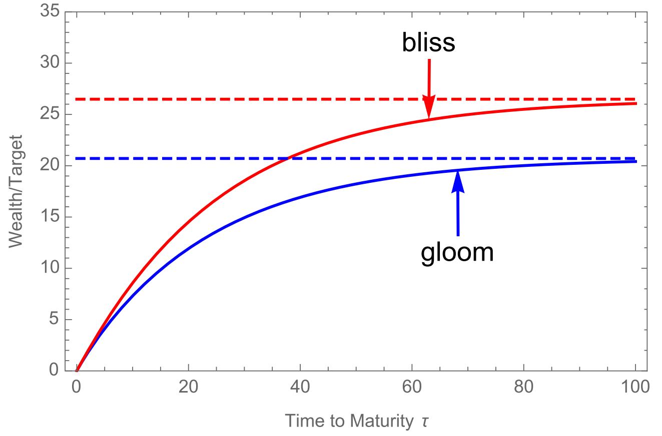

FIGURE 8 *Bliss* (red solid curve) and *gloom* (blue) as wealth to target ratios (vertical axis), against residual horizon (horizontal), for $\alpha = 0.5, \beta = 0.04, \gamma = 2$. Each curve increases with the residual horizon, asymptotically approaching the corresponding ratio in the infinite-horizon problem (dashed lines). $\mu = 8\%, \sigma = 20\%, r = 0.65\%$ [Color figure can be viewed at wileyonlinelibrary.com]

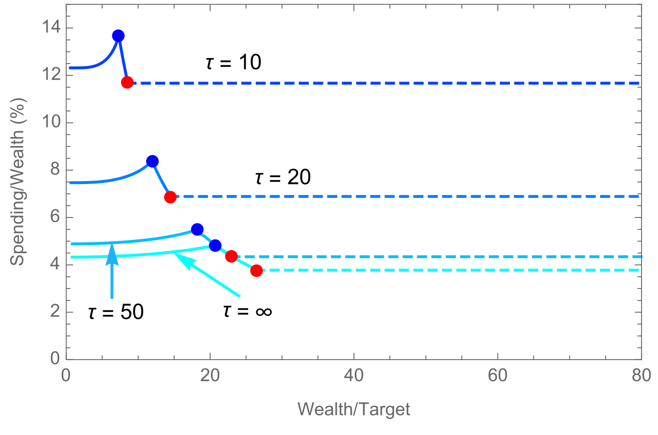

FIGURE 9 Ratio of optimal spending to wealth (vertical axis) against wealth to target ratio, for different values of time to maturity τ, with $\alpha = 0.5, \beta = 0.04, \gamma = 2$. For each τ value, the red point refers to the bliss point, and blue point refers to the gloom point. $\mu = 8\%, \sigma = 20\%, r = 0.65\%$ [Color figure can be viewed at wileyonlinelibrary.com]

11 | CONCLUDING REMARKS

It appears far easier to accustom oneself to a higher standard of living than to a lower one. The model presented here focuses on this feature of preferences by scaling the utility of spending by a fractional power of past peak spending that is the target spending. When spending exceeds its own past peak, the target adjusts accordingly. Thus, the target is endogenous.

The analysis focuses on the ratio of wealth to target, which is a diffusion process with a reflecting barrier. At the lower levels of wealth to target the optimal spending rate is proportional to the wealth (and therefore spending is cut as wealth goes down in this region). Moreover, in this region the fraction of savings invested in the risky asset is fixed.

At the higher levels of wealth to target the spending rate is a constant, unaffected by changes in wealth. All changes in wealth translate immediately to commensurate changes in savings; moreover, increases (decreases) in savings translate to increases (decreases) in the fraction of savings invested in the risky asset. In this region the *payout rate* moves in the *opposite* direction of wealth, that is, it increases as wealth decreases. At the reflecting barrier positive returns are partially consumed, raising the target spending rate.

An inspiration for the present model comes from the literature on Prospect Theory, which prominently features a reference point and an asymmetric reaction to changes up or down relative to the reference point, the down change being felt more strongly. Much of that literature is about wealth and changes thereof rather than consumption or spending, whereas the focus here is on spending. (Recall the title of Markowitz, 1952 precursor of that literature, “The Utility of Wealth.”)

The location of the reference point is always a thorny issue, with many favoring the status quo as the reference point. In this paper the formal choice is to consider the historically highest spending rate as the reference point, which may seem very different from the status quo. However, in equilibrium, where it counts—in the target or normal region—the status quo is the historically highest spending rate.

The transition from the higher wealth to the target region to the lower one entails a reduction in payout rate that recalls the observations of Brown et al. (2014) regarding University endowments. They report

> *university endowments exhibit an asymmetric response to contemporaneous positive and negative financial shocks. Specifically, following positive shocks endowments tend to follow their own stated payout policies (e.g., pay out 5% of the past three- year average of endowment values). Whereas following contemporaneous negative shocks, many endowments actively deviate from their stated payout policies, actually reducing payout rates to a level below that implied by their standard smoothing rules.*

Thus, Brown et al. (2014) observe two regions of behavior with payout policy being lower following negative financial shocks. Similarly, also the present model suggests that behavior should be different following negative financial shocks. However, the model suggests lower payout levels following a shock, but higher payout ratios. Moreover, the model suggests that in normal times payout levels should be constant (implying that payout rates should decrease with wealth) until they are increased due to a substantial wealth increase.

Another payout policy that recalls the results of the present model is that used by corporations to determine their dividends. Summarizing a 7-year, 28 company study, Lintner (1956) writes,

> *With the possible exception of 2 companies… [c]onsideration of what dividends should be paid at any given time turned, first and foremost in every case, on the question of whether the existing rate of payment should be changed…*
>
> *We found no instance in which the question of how much should be paid in a given quarter or year was considered without regard to the existing rate as an optimum problem in terms of the interests of the company and/or its stockholders at the given time…*
>
> *…serious consideration of the second question of just how large the change in dividend payments should be only after management had satisfied itself that a change in the existing rate would be positively desirable. Even then, the companies’ existing dividend rate continued to be a central bench mark for the problem in management’s eyes…*

… these elements of inertia and conservatism… were strong enough that most managements sought to avoid making changes in their dividend rates that might have to be reversed within a year or so.

Five decades later, Brav, Graham, Harvey, and Michaely (2005) perform a survey of nearly four-hundred companies and find that such concerns are still relevant:

Our analysis indicates that maintaining the dividend level is a priority on par with investment decisions. Managers express a strong desire to avoid dividend cuts, except in extraordinary circumstances.

The general challenge is the reconciliation of a preference for smooth spending, and reluctance to cut back on spending with the desire to enjoy the higher returns associated with investments in risky assets. The model presented here has these properties; central and novel among these is the relative pain associated with a shortfall in spending.

CONFLICT OF INTEREST

The authors do not declare any conflicts of interest.

DATA AVAILABILITY STATEMENT

This paper did not generate new data sets. The historical analysis uses data from the Ibbotson yearbook, which is published annually by Ibbotson and Associates.

ACKNOWLEDGMENTS

For helpful comments, we thank Andrew Ang, Kerry Back, Davide Fiaschi, Yiting Xu; seminar participants at the London School of Economics, King's College London, Rutgers University, the University of Michigan, the University of Pisa, the University of Vienna, CEMFI, Columbia University; and attendees at the Paul Woolley Conference in Sydney on Capital Market Dysfunctionality.

ENDNOTES

[1] Qualitatively, these features are reminiscent of the habit model of Bowman et al. (1999, Assumptions A4, A5, B2), with two differences. First, in the present model the reference is set at the past maximum, hence cannot be exceeded. Second, the asymmetry between gains and losses here is multiplicative rather than additive, in that utility depends on the ratio of consumption to its target, rather than the difference.

[2] With zero subjective time preference, $\gamma > 1$ is a necessary and sufficient condition for the problem to be well-posed. In its absence, expected utility can grow arbitrarily large by postponing spending. The extension to a positive discount rate in Section 9 entertains even values of $\gamma \in (0, 1)$.

[3] Note that the first case $X_t/h_t \geq b$ may only materialize at $t = 0$, as subsequent values of $h_t = h_0 \vee \sup_{0\leq s\leq t} \hat{c}_t$ satisfy by construction $X_t/h_t \leq b$.

ORCID

Paolo Guasoni https://orcid.org/0000-0002-8562-3658
Dan Ren https://orcid.org/0000-0002-6829-8945

REFERENCES

Ang, A., Bekaert, G., & Liu, J. (2005). Why stocks may disappoint. *Journal of Financial Economics*, *76*(3), 471–508.

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research*, *30*(5), 961–981.

Blavatskyy, P. (2018). A second-generation disappointment aversion theory of decision making under risk. *Theory and Decision*, *84*(1), 29–60.

Bleichrodt, H., & Wakker, P. P. (2015). Regret theory: A bold alternative to the alternatives. *The Economic Journal*, *125*(583), 493–532.

Borodin, A. N., & Salminen, P. (2002). *Handbook of Brownian motion: Facts and formulae*. Berlin: Springer.

Bowman, D., Minehart, D., & Rabin, M. (1999). Loss aversion in a consumption–savings model. *Journal of Economic Behavior & Organization*, *38*(2), 155–178.

Brav, A., Graham, J. R., Harvey, C. R., & Michaely, R. (2005). Payout policy in the 21st century. *Journal of Financial Economics*, *77*(3), 483–527.

Brown, J., Dimmock, S., Kang, J., & Weisbenner, S. (2014). How university endowments respond to financial market shocks: Evidence and implications. *American Economic Review*, *104*, 931–962.

Campbell, J. Y., & Cochrane, J. H. (1999). By force of habit: A consumption based explanation of aggregate stock market behavior. *Journal of Political Economy*, *107*(2), 205–251.

Constantinides, G. (1990). Habit formation: A resolution of the equity premium puzzle. *The Journal of Political Economy*, *98*(3), 519–543.

Davis, M., & Varaiya, P. (1973). Dynamic programming conditions for partially observable stochastic systems. *SIAM Journal on Control*, *11*(2), 226–261.

De Neve, J. E., Ward, G., De Keulenaer, F., Van Landeghem, B., Kavetsos, G., & Norton, M. I. (2018). The asymmetric experience of positive and negative economic growth: Global evidence using subjective well-being data. *Review of Economics and Statistics*, *100*(2), 362–375.

Detemple, J. B., & Karatzas, I. (2003). Non-addictive habits: Optimal consumption-portfolio policies. *Journal of Economic Theory*, *113*(2), 265–285.

Detemple, J. B., & Zapatero, F. (1991). Asset prices in an exchange economy with habit formation. *Econometrica: Journal of the Econometric Society*, *59*(6), 1633–1657.

Detemple, J. B., & Zapatero, F. (1992). Optimal consumption-portfolio policies with habit formation1. *Mathematical Finance*, *2*(4), 251–274.

Deusenberry, J. (1949). *Income, saving and the theory of consumer behaviour*. Cambridge : Harvard University Press.

Dybvig, P. H. (1995). Dusenberry's ratcheting of consumption: Optimal dynamic consumption and investment given intolerance for any decline in standard. *Review of Economic Studies*, *62*, 287–313.

Dybvig, P. H. (1999). Using asset allocation to protect spending. *Financial Analysts Journal*, *55*, 49–62.

Gilbert, T., & Hrdlicka, C. (2012). *Fairness and risk-sharing across generations: An application to university and nonprofit endowments*. University of Washington Working Paper.

Gilbert, T., & Hrdlicka, C. (2015). Why are university endowments large and risky? *The Review of Financial Studies*, *28*(9), 2643–2686.

Guasoni, P., & Huberman, G. (2016). *Asset pricing with shortfall aversion*. Working Paper.

Gul, F. (1991). A theory of disappointment aversion. *Econometrica*, *59*(3), 667–686.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Macmillan.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, *47*(2), 263–291.

Karatzas, I., Lehoczky, J. P., Shreve, S. E., & Xu, G.-L. (1991). Martingale and duality methods for utility maximization in an incomplete market. *SIAM Journal on Control and optimization*, *29*(3), 702–730.

Karatzas, I., & Shreve, S. E. (1991). *Brownian motion and stochastic calculus (Graduate texts in mathematics) (2nd ed., Vol. 113)*. New York: Springer.

Kőszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics*, *121*(4), 1133–1165.

Kőszegi, B., & Rabin, M. (2007). Reference-dependent risk attitudes. *The American Economic Review*, *97*(4), 1047–1073.

Kőszegi, B., & Rabin, M. (2008). Choices, situations, and happiness. *Journal of Public Economics*, *92*(8), 1821–1832.

Kőszegi, B., & Rabin, M. (2009). Reference-dependent consumption plans. *American Economic Review*, *99*(3), 909–936.

Lintner, J. (1956). Distribution of incomes of corporations among dividends, retained earnings, and taxes. *American Economic Review*, *46*(2), 97–113.

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal*, *92*(368), 805–824.

Markowitz, H. (1952). The utility of wealth. *Journal of Political Economy*, *60*(2), 151–158.

Menzly, L., Santos, T., & Veronesi, P. (2004). Understanding predictability. *Journal of Political Economy*, *112*(1), 1–47.

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

Merton, R. C. (1993). Optimal investment strategies for university endowment funds. In *Studies of supply and demand in higher education* (pp. 211–242). University of Chicago Press.

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

Pagel, M. (2017). Expectations-based reference-dependent life-cycle consumption. *Review of Economic Studies*, *84*(2), 885–934.

Pagel, M. (2018). A news-utility theory for inattention and delegation in portfolio choice. *Econometrica*, *86*(2), 491–522.

Rogers, L. C. G., & Williams, D. (2000). *Diffusions, Markov processes, and martingales: Volume 1, foundations*. Cambridge: Cambridge University Press.

Routledge, B. R., & Zin, S. E. (2010). Generalized disappointment aversion and asset prices. *The Journal of Finance*, *65*(4), 1303–1332.

Ryder, H. E., & Heal, G. M. (1973). Optimal growth with intertemporally dependent preferences. *The Review of Economic Studies*, *40*(1), 1–31.

Santos, T., & Veronesi, P. (2010). Habit formation, the cross section of stock returns and the cash-flow risk puzzle. *Journal of Financial Economics*, *98*(2), 385–413.

Stutzer, A. (2004). The role of income aspirations in individual happiness. *Journal of Economic Behavior & Organization*, *54*(1), 89–109.

Sundaresan, S. M. (1989). Intertemporally dependent preferences and the volatility of consumption and wealth. *Review of financial Studies*, *2*(1), 73–89.

Thillaisundaram, A. (2012). *The Merton problem with a drawdown constraint on consumption* (Technical report). Cambridge: University of Cambridge.

Tobin, J. (1974). What is permanent endowment income? *American Economic Review*, *64*(2), 427–432.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*(4), 297–323.

van Bilsen, S., Laeven, R., & Nijman, T. (2019). Consumption and portfolio choice under loss aversion and endogenous updating of the reference level. *Management Science*.

How to cite this article: Guasoni P, Huberman G, Ren D. Shortfall aversion. *Mathematical Finance*. 2020;30:869–920. https://doi.org/10.1111/mafi.12239

APPENDIX A: PROOFS

This section contains the proofs of all the statements in the paper. The first subsection constructs the spending and investment policy explicitly, and proves that it is the unique optimal solution to the intertemporal utility maximization problem. In particular, Lemma A.1 uses a martingale argument to obtain an a priori upper bound on the utility of any spending–investment policy, while Lemma A.2 shows that this bound is achieved by the candidate optimal policy. Proposition A.3 and its auxiliary Lemmas A.4, A.5, A.8 link the optimal spending policy identified through the martingale argument to the solution of the ODE in the paper, thereby providing the closed-form solution in Theorem 4.1.

The second subsection investigates the long-run properties of the optimal policy. Lemma A.10 identifies the invariant distribution of the marginal utility relative to the target, and hence the distribution of the spending rate relative to the target spending. Theorem A.12 proves that the risky asset weight is increasing in the target–wealth ratio.

A.1 The optimal spending and investment policy

The first task is to construct the optimal spending policy $\hat{c}_t$. For any constant $y \geq 0$, recall the dual of the utility function, $\tilde{U}(y,h) = \sup_{c\geq 0}(U(c,h) - cy)$, in (27):

$$\tilde{U}(y,h) = \begin{cases} -(1-\alpha)^{1/\gamma^*}\frac{y^{1-\frac{1}{\gamma^*}}}{1-\frac{1}{\gamma^*}} & y \leq (1-\alpha)h^{-\gamma^*} \\ \frac{h^{(1-\alpha)(1-\gamma)}}{1-\gamma} - hy & (1-\alpha)h^{-\gamma^*} < y \leq h^{-\gamma^*} \\ -\frac{y^{1-\frac{1}{\gamma}}}{1-\frac{1}{\gamma}}h^{\alpha(1-\frac{1}{\gamma})} & y > h^{-\gamma^*} \end{cases}, \tag{A.1}$$

while the $\tilde{U}(y,h)$-maximizing spending rate $\hat{c}$ equals to

$$\hat{c} = \begin{cases} (y/(1-\alpha))^{-1/\gamma^*} & y \leq (1-\alpha)h^{-\gamma^*}, \\ h & (1-\alpha)h^{-\gamma^*} < y \leq h^{-\gamma^*}, \\ y^{-1/\gamma}h^{-\alpha(1-\gamma)/\gamma} & y > h^{-\gamma^*}. \end{cases}$$

In particular, the expression $U(c,h) - cy$ attains its maximum for $c > h$ in the first case, for $c = h$ in the second case, and for $c < h$ in the third case. (Note that in our previous argument in Subsection 8.2, the first case is merged to the second one, because once $c > h$, the target h is updated and then $h = c$ as in the second case.) Note also that, when the first case holds, the value of $\tilde{U}$ does not depend on h. Denote by M_t the *stochastic discount factor*

$$M_t := e^{-\left(r+\frac{\mu^2}{2\sigma^2}\right)t-\frac{\mu}{\sigma}W_t}$$

and recall that any spending–investment policy (c_t, π_t) starting from initial capital x satisfies the condition $\mathbb{E}_{x,\bar{h}}[\int_0^\infty c_t M_t] \leq x$. The next result uses a martingale argument to find a family of upper bounds for the expected utility of any such spending plan.

Lemma A.1. *Any spending plan $(c_t)_{t\geq 0}$ such that $\mathbb{E}_{x,\bar{h}}[\int_0^\infty c_t M_t dt] \leq x$ satisfies, for any $y > 0$:*

$$\mathbb{E}_{x,\bar{h}}\left[\int_0^\infty \frac{(c_t h_t^{-\alpha})^{1-\gamma}}{1-\gamma}dt\right] \leq \mathbb{E}_{x,\bar{h}}\left[\int_0^\infty \tilde{U}\big(yM_t, \hat{h}_t(y)\big)dt\right] + xy, \tag{A.2}$$

where $h_t = \bar{h} \vee \sup_{s\leq t} c_s$ and

$$\hat{h}_t(y) = \bar{h} \vee \left(y \inf_{s\leq t} M_s/(1-\alpha)\right)^{-1/\gamma^*}. \tag{A.3}$$

Proof. First, note that for any $y > 0$ and any $(c_t)_{t\geq 0}$ such that $\mathbb{E}_{x,\bar{h}}[\int_0^\infty c_t M_t] \leq x$,

$$\mathbb{E}_{x,\bar{h}}\left[\int_0^\infty U(c_t,h_t)dt\right] = \mathbb{E}_{x,\bar{h}}\left[\int_0^\infty (U(c_t,h_t) - yM_tc_t)dt\right] + y\mathbb{E}_{x,\bar{h}}\left[\int_0^\infty c_tM_tdt\right]$$
$$\leq \mathbb{E}_{x,\bar{h}}\left[\int_0^\infty (U(c_t,h_t) - yM_tc_t)dt\right] + xy \leq \mathbb{E}_{x,\bar{h}}\left[\int_0^\infty \tilde{U}(yM_t,h_t)dt\right] + xy.$$

The proof is now completed by showing that $\tilde{U}(yM_t,h_t) \leq \tilde{U}(yM_t,\hat{h}_t(y))$ for all $t \geq 0$.

Now, suppose that $h_t(\omega)$ strictly increases at t (henceforth ω is dropped for brevity), in that for any $\varepsilon > 0$ there is $u \in (t, t+\varepsilon)$ such that $h_u \geq c_u > h_t$. For any such u, the first case in (A.1) holds, and therefore

$$U(c_t,h_t) - yM_uc_t \leq \tilde{U}(yM_u,h_t) = \tilde{U}(yM_u,\hat{c}_u),$$

where $\hat{c}_u > h_t$ satisfies the first-order condition $\hat{c}_u = (yM_u/(1-\alpha))^{-1/\gamma^*}$. Passing to the limit as $u \downarrow t$, by the continuity of $M_\cdot$ and $\tilde{U}(\cdot,\cdot)$, it follows that

$$\tilde{U}(yM_t,h_t) = \tilde{U}(yM_t,\hat{c}_t).$$

Note that, in pathwise sense, for any t, $h_t = \bar{h} \vee \sup_{s\in\mathcal{I}_t} h_s$, where $\mathcal{I}_t = \{s \leq t, h \text{ strictly increases at } s\}$. Then, $h_t = \bar{h} \vee \sup_{s\in\mathcal{I}_t} \hat{c}_s = \bar{h} \vee \sup_{s\leq t} \hat{c}_s = \hat{h}_t(y)$, where the second equality follows by observing that $\hat{c}_s < h_s$ for all $s \notin \mathcal{I}_t$ (as seen from the expression of $\hat{c}_u$ with $yM_u > (1-\alpha)h_u^{-\gamma^*}$). Hence,

$$U(c_t,h_t) - yM_tc_t \leq \tilde{U}(yM_t,h_t) = \tilde{U}(yM_t,\bar{h} \vee \sup_{s\in\mathcal{I}_t} h_s) = \tilde{U}(yM_t,\bar{h} \vee \sup_{s\in\mathcal{I}_t} \hat{c}_s)$$
$$= \tilde{U}(yM_t,\bar{h} \vee \sup_{s\leq t} \hat{c}_s) = \tilde{U}\left(yM_t,\hat{h}_t(y)\right).$$

□

In the above argument, the utility of a spending plan $(c_t)_{t\geq 0}$ achieves the upper bound if both the *first-order condition* (A.4) and the *saturation condition* (A.5) hold:

$$\tilde{U}(yM_t,h_t) = U(c_t,h_t) - yM_tc_t, \tag{A.4}$$

$$\mathbb{E}_{x,\bar{h}}\left[\int_0^\infty M_tc_t\mathrm{d}t\right] = x. \tag{A.5}$$

The next lemma shows that for any initial capital x there exists a constant $\hat{y} > 0$ and a spending plan $\hat{c}$ such that both conditions are satisfied, thereby proving the existence of an optimal policy.

Lemma A.2. *There exists a unique $\hat{y} > 0$ and a spending plan $\hat{c}_t(\hat{y})$ such that (A.2) holds as an equality. Hence, such plan is optimal.*

Proof. First, define the spending plan $\hat{c}_t(y)$ as in (28) for $\hat{h}_t(y) = \bar{h} \vee (\frac{\inf_{s\leq t} Y_s}{1-\alpha})^{-1/\gamma^*}$, with $Y_t = yM_t$, so that it satisfies the first-order condition. Thus, define $\hat{c}_t(y) = \hat{h}_t(y)F(Y_t)$, where

$$\begin{aligned}F(Y_t) =& 1_{\{(1-\alpha)(\hat{h}_t(y))^{-\gamma^*}\le Y_t\le(\hat{h}_t(y))^{-\gamma^*}\}} + \left(Y_t(\hat{h}_t(y))^{\gamma^*}\right)^{-1/\gamma} 1_{\{Y_t>(\hat{h}_t(y))^{-\gamma^*}\}}\\ =& 1+\left[\left(Y_t(\hat{h}_t(y))^{\gamma^*}\right)^{-1/\gamma}-1\right]1_{\{Y_t>(\hat{h}_t(y))^{-\gamma^*}\}}\\ =& 1+\left[\left(h^{-\gamma^*/\gamma}Y_t^{-1/\gamma}\wedge\left(\frac{\inf_{s\le t}Y_s}{(1-\alpha)Y_t}\right)^{1/\gamma}\right)-1\right]1_{\left\{h^{-\gamma^*/\gamma}Y_t^{-1/\gamma}\wedge\left(\frac{\inf_{s\le t}Y_s}{(1-\alpha)Y_t}\right)^{1/\gamma}<1\right\}}.\end{aligned}$$

This spending plan satisfies (A.4) by construction. To show that for some $y > 0$ it also satisfies (A.5), observe that when y increases, both $\hat{h}_t(y)$ and $F(Y_t)$ strictly decrease. Therefore:

1. The following expectation is strictly decreasing with respect to y:

$$\mathbb{E}_{x,\bar{h}}\left[\int_0^\infty M_t\hat{c}_t(y)\mathrm{d}t\right] = \mathbb{E}_{x,\bar{h}}\left[\int_0^\infty M_t\hat{h}_t(y)F(Y_t)\mathrm{d}t\right].$$

2. If $y\downarrow 0$, then $\hat{h}_t(y)\uparrow\infty$ and $F(Y_t)>0$, implying $\mathbb{E}_{x,\bar{h}}[\int_0^\infty M_t\hat{c}_t(y)\mathrm{d}t]\uparrow\infty$.
3. If $y\uparrow\infty$, then $\hat{h}_t(y)\downarrow h$ and $F(Y_t)\downarrow 0$, implying $\mathbb{E}_{x,\bar{h}}[\int_0^\infty M_t\hat{c}_t(y)\mathrm{d}t]\downarrow 0$.

Moreover, as $\mathbb{E}_{x,\bar{h}}[\int_0^\infty M_t\hat{c}_t(y)\mathrm{d}t]$ is continuous in y, for any $x>0$ there exists a unique $\hat{y}$ such that also the saturation condition holds. As this policy $\hat{c}_t$ satisfies both the first-order and saturation conditions, and in addition $\hat{h}\vee\sup_{s\le t}\hat{c}_s(y)=\hat{h}_t(y)$, it follows that (A.2) holds as an equality, which means that $\hat{c}_t(y)$ is optimal. □

The previous statements prove the existence of an optimal spending plan, and identify the value function as the minimum over y of the right-hand side of (A.2). The next step is to link this expression of the value function with $\tilde{V}(y,h)$, defined as the solution of the PDE (26) and its boundary conditions. By the homogeneity of the value function, for any $y\ge 0$, let $\tilde{V}(y,h)=h^{1-\gamma^*}q(z)$ with $z=h^{\gamma^*}y$, where q is defined in Lemma A.4.

Proposition A.3. *For any constant $y\ge 0$ and $\hat{h}_t(y)$ defined in (A.3),*

$$\mathbb{E}_{x,\bar{h}}\int_0^\infty\tilde{U}(yM_t,\hat{h}_t(y))\mathrm{d}t=\tilde{V}(y,\bar{h}).$$

The proof of Proposition A.3 hinges on a few technical Lemmas, which are proved afterward.

Proof. Recall that for any $y\ge 0$, function $\tilde{V}(y,\bar{h})$ satisfies the PDE:

$$\frac{(\mu/\sigma)^2}{2}y^2\tilde{V}_{yy}-ry\tilde{V}_y+\tilde{U}(y,\bar{h})=0.$$

Applying Itô's formula to $\tilde{V}(yM_t,\hat{h}_t(y))$, and using the equation above, it follows that

$$\begin{aligned}\mathrm{d}\big(\tilde{V}(yM_t,\hat{h}_t(y))\big)=&-\tilde{U}(yM_t,\hat{h}_t(y))\mathrm{d}t+\tilde{V}_h(yM_t,\hat{h}_t(y))\mathrm{d}\hat{h}_t(y)\\ &-\frac{\mu}{\sigma}yM_t\tilde{V}_1(yM_t,\hat{h}_t(y))\mathrm{d}W_t.\end{aligned}$$

Define $\tau_n=\inf\{t\ge 0\mid yM_t\ge n,\hat{h}_t(y)\ge[(1-\alpha)n]^{1/\gamma^*}\}$. By the form of $\hat{h}(y)$ in (A.3), it follows that $\inf_{s\le t}(yM_s)\ge 1/n$ for all $t\le\tau_n$. Now take any $T\in(0,\infty)$, and integrate the above from 0 to $T\wedge\tau_n$. Then, note that

1. The integral of the $d\hat{h}_t(y)$ term vanishes, because $\hat{h}_t(y)$ increases only if $\hat{c}_t(y) = \hat{h}_t(y)$, at which case $\tilde{V}_h = V_h = 0$ by the Neumann boundary condition in (23).
2. The stochastic integral

$$\int_0^{T\wedge\tau_n} \frac{\mu}{\sigma} yM_t\tilde{V}_1(yM_t, \hat{h}_t(y))dW_t = \int_0^{T\wedge\tau_n} \frac{\mu}{\sigma} yM_t\hat{h}_t(y)q'\Big(yM_t(\hat{h}_t(y))^{\gamma^*}\Big)dW_t$$

is a local martingale in T. In addition, by the continuity of the function q' (Lemma A.4) and the definition of τ_n, the expectation of the quadratic variation of this local martingale is finite, hence it is a martingale.

Then, taking expectations, it follows that:

$$\tilde{V}(y, \bar{h}) = \mathbb{E}_{x,\bar{h}}\left[\int_0^{T\wedge\tau_n} \tilde{U}(yM_t, \hat{h}_t(y))dt\right] + \mathbb{E}_{x,\bar{h}}\left[\tilde{V}(yM_{T\wedge\tau_n}, \hat{h}_{T\wedge\tau_n}(y))\right] \tag{A.6}$$

$$= \mathbb{E}_{x,\bar{h}}\left[\int_0^{T\wedge\tau_n} \tilde{U}(yM_t, \hat{h}_t(y))dt\right] + \mathbb{E}_{x,\bar{h}}\left[\tilde{V}(yM_{\tau_n}, \hat{h}_{\tau_n}(y))1_{\{\tau_n\le T\}}\right] + \mathbb{E}_{x,\bar{h}}\left[\tilde{V}(yM_T, \hat{h}_T(y))1_{\{\tau_n> T\}}\right]. \tag{A.7}$$

Now consider the three expectations in (A.6) separately. For the first one, note that $\tilde{U} < 0$ because $\gamma > 1$, then as $n \to \infty$ the first term converges to $\mathbb{E}_{x,\bar{h}} \int_0^T \tilde{U}(yM_t, \hat{h}_t(y))dt$ by the monotone convergence theorem. The second term is bounded in absolute value by

$$\mathbb{E}_{x,\bar{h}}\left[\left|\tilde{V}(yM_{\tau_n}, \hat{h}_{\tau_n}(y))\right|1_{\{\tau_n\le T\}}\right] = \mathbb{E}_{x,\bar{h}}\left[\Big(\hat{h}_{\tau_n}(y)\Big)^{1-\gamma^*}\left|q\Big(yM_{\tau_n}(\hat{h}_{\tau_n}(y))^{\gamma^*}\Big)\right|1_{\{\tau_n\le T\}}\right]. \tag{A.8}$$

As $\gamma^* > 1$, and by the definition of τ_n, it follows that $(\hat{h}_{\tau_n}(y))^{1-\gamma^*} \le \bar{h}^{1-\gamma^*}$, $yM_{\tau_n}(\hat{h}_{\tau_n}(y))^{\gamma^*} \le (1-\alpha)n^2$; and recall that (Lemma A.4)

$$q(z) = O(z^{1-1/\gamma}) \quad \text{as } z \to \infty,$$

then

$$\Big(\hat{h}_{\tau_n}(y)\Big)^{1-\gamma^*}\left|q\Big(yM_{\tau_n}(\hat{h}_{\tau_n}(y))^{\gamma^*}\Big)\right| = O\Big(n^{2(1-1/\gamma)}\Big) \text{ for } n \to \infty. \tag{A.9}$$

Moreover, by Chebyshev's inequality and Karatzas and Shreve (1991, 5.3.17), there exists some constant C depending on m such that

$$\mathbb{E}_{x,\bar{h}}\Big(1_{\{\tau_n\le T\}}\Big) = \mathbb{P}_{x,\bar{h}}\big(\tau_n \le T\big) \le \mathbb{P}_{x,\bar{h}}\left(\{\sup_{t\in[0,T]} yM_t \ge n\} \bigcup \{\inf_{t\in[0,T]} yM_t \le 1/n\}\right)$$

$$\le \mathbb{P}_{x,\bar{h}}\left(\sup_{t\in[0,T]} yM_t \ge n\right) + \mathbb{P}_{x,\bar{h}}\left(\inf_{t\in[0,T]} yM_t \le 1/n\right)$$

$$\le \mathbb{P}_{x,\bar{h}}\left(\sup_{t\in[0,T]} yM_t \ge n\right) + \mathbb{P}_{x,\bar{h}}\left(\sup_{t\in[0,T]} y^{-1}M_t^{-1} \ge n\right)$$

$$\leq n^{-2\kappa}\mathbb{E}_{x,\bar{h}}\left[\sup_{t\in[0,T]}\left(yM_t\right)^{2\kappa}\right]+n^{-2\kappa}\mathbb{E}_{x,\bar{h}}\left[\sup_{t\in[0,T]}\left(yM_t\right)^{-2\kappa}\right]=O\left(n^{-2\kappa}(1+y^{2\kappa})e^{CT}\right) \tag{A.10}$$

for any $\kappa \geq 1$. Then (A.8) converges to 0 as $n \to \infty$, by (A.9) and (A.10). That is, the second expectation in (A.6) also converges to 0 as $n \to \infty$. As $n \to \infty$, the third expectation in (A.6) converges to $\mathbb{E}_{x,\bar{h}}[\tilde{V}(yM_T, \hat{h}_T(y))]$. Thus, passing to the limit as $n \to \infty$ in (A.6), it follows that

$$\tilde{V}(y,\bar{h})=\mathbb{E}_{x,\bar{h}}\left[\int_0^T \tilde{U}(yM_t,\hat{h}_t(y))\mathrm{d}t\right]+\mathbb{E}_{x,\bar{h}}\left[\tilde{V}(yM_T,\hat{h}_T(y))\right]. \tag{A.11}$$

The proof is completed by observing that, as $T \to \infty$, the first expectation on the right-hand side of (A.11) converges to $\mathbb{E}_{x,\bar{h}}[\int_0^\infty \tilde{U}(yM_t, \hat{h}_t(y))\mathrm{d}t]$ by monotone convergence, and the second expectation converges to 0 by Lemma A.8. □

The next lemma computes the explicit solution to the main ODE (31) and states its properties.

Lemma A.4. *The C^2 function $q : (1-\alpha, \infty) \to \mathbb{R}$ defined in the following cases, is convex and nonincreasing on $(1-\alpha,\infty)$. (Recall that $m = (1-\frac{1}{\gamma})(r+\frac{\mu^2}{2\gamma\sigma^2})$ denotes the Merton consumption.)*

Case 1*: $r \neq 0$.*

$$q(z)=\begin{cases} C_1+C_2z^{1+\rho}-\frac{z}{r}+\frac{\rho\log z}{r(1-\gamma)(\rho+1)} & 1-\alpha<z\leq 1,\\ C_3+\frac{\gamma}{(1-\gamma)m}z^{1-1/\gamma} & z>1,\end{cases}$$

where

$$C_1=-\frac{\gamma^*\rho+1}{1-\gamma}(1-\alpha)^\rho C_2+\frac{1}{r(1-\gamma)}\left[1-\frac{\rho}{\rho+1}\left(\frac{\gamma^*}{1-\gamma^*}+\log(1-\alpha)\right)\right]$$

$$C_2=\frac{1}{r(\rho+1)^2(\gamma\rho+1)}C_3=C_1+C_2+\frac{1}{r}\left[\frac{\rho\gamma^3}{(\gamma-1)^2(\gamma\rho+1)}-1\right]$$

Case 2*: $r = 0$.*

$$q(z)=\begin{cases} C_1+C_2z+\frac{2z\log z}{(\mu/\sigma)^2}+\frac{2\log z}{(1-\gamma)(\mu/\sigma)^2} & 1-\alpha<z\leq 1,\\ C_3+\frac{\gamma}{(1-\gamma)m}z^{1-1/\gamma} & z>1,\end{cases}$$

where

$$C_1=\frac{2}{(1-\gamma)(\mu/\sigma)^2}\left[\alpha(\gamma-1)+3-\frac{1}{1-\gamma^*}-2\log(1-\alpha)\right],$$

$$C_2=-\frac{2(\gamma+2)}{(\mu/\sigma)^2}, C_3=\frac{2}{(1-\gamma)(\mu/\sigma)^2}\left[\alpha\left(\gamma-1-\frac{1}{1-\gamma^*}\right)-2\log(1-\alpha)\right].$$

Proof. For brevity, consider $r \neq 0$, as the argument for $r = 0$ is analogous. To see that q is convex, consider its second derivative

$$q''(z) = \begin{cases} C_2(1+\rho)\rho z^{\rho-1} + \frac{\rho}{r(\gamma-1)(\rho+1)z^2} & 1-\alpha < z \leq 1, \\ \frac{1}{\gamma m} z^{-1/\gamma-1} & z > 1. \end{cases}$$

As C_2 is positive, $1 < 1+\rho$, and q is C^2, it follows that q is convex on $(1-\alpha, \infty)$. As q is convex, its derivative q' is nondecreasing, and as $q'(z) < 0$ for $z > 1$, it follows that q is nonincreasing for any $z > 1-\alpha$. □

The next Lemma is used to prove Lemma A.8.

Lemma A.5. *Let $(B_t)_{t\geq 0}$ be a standard Brownian motion under the probability measure $\mathbb{P}$, and denote by $B_t^* = \sup_{0\leq s\leq t} B_s$ its running maximum. Then, for any constants a, b, k with $2a+b \neq 0, k \geq 0$*

$$\mathbb{E}\left[e^{aB_T+bB_T^*}1_{\{B_T^*>k\}}\right] = \frac{2(a+b)}{2a+b} \exp\left\{\frac{(a+b)^2}{2}T\right\}\Phi\left((a+b)\sqrt{T} - \frac{k}{\sqrt{T}}\right) + \frac{2a}{2a+b}\exp\left\{(2a+b)k + \frac{a^2}{2}T\right\}\Phi\left(-a\sqrt{T} - \frac{k}{\sqrt{T}}\right),$$

and hence

$$\lim_{T\to\infty} \frac{1}{T} \log \mathbb{E}\left[e^{aB_T+bB_T^*}1_{\{B_T^*>k\}}\right] = \begin{cases} \frac{(a+b)^2}{2} & a+b > 0, 2a+b > 0, \\ \frac{a^2}{2} & a < 0, 2a+b < 0, \\ 0 & a+b \leq 0, a \geq 0, \end{cases}$$

where $\Phi(\cdot)$ is the standard normal distribution function.

Proof. Recall the joint probability density of (B_T, B_T^*) is

$$f_{B_T,B_T^*}(x,y) = \frac{2(2y-x)}{\sqrt{2\pi t^3}} e^{-\frac{(2y-x)^2}{2t}}, \text{ for } y \geq 0, \ x \leq y.$$

Hence, the expectation is:

$$\mathbb{E}\left[e^{aB_T+bB_T^*}1_{\{B_T^*>k\}}\right] = \int_k^\infty \int_{-\infty}^y e^{ax+by} \frac{2(2y-x)}{\sqrt{2\pi T^3}} e^{-\frac{(2y-x)^2}{2T}} \mathrm{d}x\mathrm{d}y$$
$$= 2e^{\frac{(a+b)^2 T}{2}} \Phi\left((a+b)\sqrt{T} - \frac{k}{\sqrt{T}}\right) - 2a\int_k^\infty e^{(2a+b)y + \frac{a^2 T}{2}} \Phi\left(-a\sqrt{T} - \frac{y}{\sqrt{T}}\right)\mathrm{d}y,$$

where the second term follows by integration by parts

$$-2a\int_k^\infty e^{(2a+b)y+\frac{a^2T}{2}}\Phi\left(-a\sqrt{T} - \frac{y}{\sqrt{T}}\right)\mathrm{d}y$$
$$= \frac{2a}{2a+b} e^{\frac{a^2T}{2}+(2a+b)k}\Phi\left(-a\sqrt{T} - \frac{k}{\sqrt{T}}\right) - \frac{2a}{2a+b} e^{\frac{(a+b)^2T}{2}}\Phi\left((a+b)\sqrt{T} - \frac{k}{\sqrt{T}}\right)$$

and the lemma follows. □

Remark A.6. Note that when $k < 0$, then $\mathbb{E}_{x,\bar{h}}[e^{aB_T + bB_T^*} 1_{\{B_T^* > k\}}] = \mathbb{E}_{x,\bar{h}}[e^{aB_T + bB_T^*}]$, and the corresponding limit $\lim_{T\to\infty} \frac{1}{T} \log \mathbb{E}_{x,\bar{h}}[e^{aB_T + bB_T^*} 1_{\{B_T^* > k\}}]$ is the same as in the above Lemma. That is, even if k is restricted to be nonnegative in Lemma A.5 for the expectation calculation, there is no restriction for k in the limit.

Corollary A.7. *Let $B_t^{(\zeta)} = B_t + \zeta t$, where B is a standard Brownian motion under probability measure $\mathbb{P}$, $(B_t^{(\zeta)})^*$ be the running maximum of $B_t^{(\zeta)}$. Then for any constants a, b, k with $2a + b + 2\zeta \neq 0, k \geq 0$, the following expectation under $\mathbb{P}$ is:*

$$
\begin{aligned}
\mathbb{E}\left[e^{aB_T^{(\zeta)} + b\left(B_T^{(\zeta)}\right)^*} 1_{\left\{\left(B_T^{(\zeta)}\right)^* > k\right\}}\right] = \frac{2(a+b+\zeta)}{2a+b+2\zeta} \exp\left\{\frac{(a+b)(a+b+2\zeta)}{2} T\right\} \\
\Phi\left((a+b+\zeta)\sqrt{T} - \frac{k}{\sqrt{T}}\right) + \frac{2(a+\zeta)}{2a+b+2\zeta} \\
\exp\left\{(2a+b+2\zeta)k + \frac{a(a+2\zeta)}{2} T\right\} \Phi\left(-(a+\zeta)\sqrt{T} - \frac{k}{\sqrt{T}}\right).
\end{aligned}
$$

Proof. Find the probability measure $\mathbb{Q}$ by Girsanov's theorem so that $(B_t^{(\zeta)})_{t\geq 0}$ is a standard Brownian motion under $\mathbb{Q}$, and then calculate the expectation under $\mathbb{Q}$. □

Lemma A.8. *For any $y \geq 0$, $\lim_{T\to\infty} \mathbb{E}_{x,\bar{h}}[\tilde{V}(yM_T, \hat{h}_T(y))] = 0$.*

Proof. By the definition of q, it follows that, for z large, $q(z) = O(z^{1-1/\gamma})$. Therefore,

$$
\begin{aligned}
\mathbb{E}_{x,\bar{h}}\left[\tilde{V}(yM_T, \hat{h}_T(y))\right] &= \mathbb{E}_{x,\bar{h}}\left[(\hat{h}_T(y))^{1-\gamma^*} q\left(yM_T(\hat{h}_T(y))^{\gamma^*}\right)\right] \\
&= O\left(\mathbb{E}_{x,\bar{h}}\left[(\hat{h}_T(y))^{1-\gamma^*}\left(yM_T(\hat{h}_T(y))^{\gamma^*}\right)^{1-1/\gamma}\right]\right) \\
&= O\left(\mathbb{E}_{x,\bar{h}}\left[(\hat{h}_T(y))^{1-\gamma^*/\gamma}\left(M_T\right)^{1-1/\gamma}\right]\right).
\end{aligned} \tag{A.12}
$$

Moreover,

$$
\begin{aligned}
&\mathbb{E}_{x,\bar{h}}\left[(\hat{h}_T(y))^{1-\gamma^*/\gamma}\left(M_T\right)^{1-1/\gamma}\right] = \mathbb{E}_{x,\bar{h}}\left[\left(\bar{h} \vee \left(\frac{y \inf_{s\leq t} M_s}{1-\alpha}\right)^{-1/\gamma^*}\right)^{1-\gamma^*/\gamma}\left(M_T\right)^{1-1/\gamma}\right] \\
&\leq \bar{h}^{1-\gamma^*/\gamma} \mathbb{E}_{x,\bar{h}}\left[\left(M_T\right)^{1-1/\gamma}\right] + \mathbb{E}_{x,\bar{h}}\left[\left(\frac{y \inf_{s\leq t} M_s}{1-\alpha}\right)^{-1/\gamma^*+1/\gamma}\left(M_T\right)^{1-1/\gamma} 1_{\left\{\left(\frac{y \inf_{s\leq t} M_s}{1-\alpha}\right)^{-1/\gamma^*} \geq \bar{h}\right\}}\right] \\
&= O\left(\mathbb{E}_{x,\bar{h}}\left[\left(M_T\right)^{1-1/\gamma}\right]\right) + O\left(\mathbb{E}_{x,\bar{h}}\left[\left(\inf_{s\leq t}(M_s)\right)^{-1/\gamma^*+1/\gamma}\left(M_T\right)^{1-1/\gamma} 1_{\{\inf_{s\leq t}(M_s) \leq \frac{(1-\alpha)(\bar{h})^{-\gamma^*}}{y}\}}\right]\right).
\end{aligned} \tag{A.13}
$$

The inequality above holds by observing that

$$\left(\bar{h} \vee \left(\frac{y \inf_{s\le t} M_s}{1-\alpha}\right)^{-1/\gamma^*}\right)^{1-\gamma^*/\gamma} \le \bar{h}^{1-\gamma^*/\gamma} + \left(\left(\frac{y \inf_{s\le t} M_s}{1-\alpha}\right)^{-1/\gamma^*}\right)^{1-\gamma^*/\gamma} 1_{\left\{\left(\frac{y \inf_{s\le t} M_s}{1-\alpha}\right)^{-1/\gamma^*} \ge \bar{h}\right\}}.$$

To compute the two expectations in (A.13), note first that

$$M_T = \exp\left\{-\left(r + \frac{\mu^2}{2\sigma^2}\right)T - \frac{\mu}{\sigma}W_T\right\} = \exp\left\{-\frac{\mu}{\sigma}W_T^{(\zeta)}\right\},$$

where $W_T^{(\zeta)} = W_T + \zeta T$, $\zeta = \frac{r}{\mu/\sigma} + \frac{\mu/\sigma}{2}$. Therefore, by Corollary A.7 with $b = k = 0$, the first expectation satisfies

$$\mathbb{E}_{x,\bar{h}}\left[\left(M_T\right)^{1-1/\gamma}\right] = \mathbb{E}_{x,\bar{h}}\left[\exp\left\{-\left(1-\frac{1}{\gamma}\right)\frac{\mu}{\sigma}W_T^{(\zeta)}\right\}\right] = \exp\left\{-\left(1-\frac{1}{\gamma}\right)\left(r + \frac{(\mu/\sigma)^2}{2\gamma}\right)T\right\}.$$

As $\gamma > 1$, it follows that

$$\lim_{T\to\infty} \mathbb{E}_{x,\bar{h}}\left[\left(M_T\right)^{1-1/\gamma}\right] = 0. \tag{A.14}$$

For the second expectation, apply Corollary A.7 again with $a = -(1-\frac{1}{\gamma})\frac{\mu}{\sigma}$ and $b = (\frac{1}{\gamma^*} - \frac{1}{\gamma})\frac{\mu}{\sigma}$, $k = \frac{\gamma^* \log h + \log y - \log(1-\alpha)}{\mu/\sigma}$:

$$\begin{aligned}
&\lim_{T\to\infty} \frac{1}{T} \log\left\{\mathbb{E}_{x,\bar{h}}\left[\left(\inf_{s\le t} M_s\right)^{-1/\gamma^*+1/\gamma}\left(M_T\right)^{1-1/\gamma} 1_{\{\inf_{s\le t}(M_s)\le \frac{(1-\alpha)\bar{h}^{-\gamma^*}}{y}\}}\right]\right\} \\
&= \lim_{T\to\infty} \frac{1}{T} \log \mathbb{E}_{x,\bar{h}}\left[\exp\left\{aW_T^{(\zeta)} + b\left(W_T^{(\zeta)}\right)^*\right\} 1_{\{\left(W_T^{(\zeta)}\right)^* \ge k\}}\right] \\
&\le \max\left\{\frac{(a+b)(a+b+2\zeta)}{2}, \frac{a(a+2\zeta)}{2}, -\frac{\zeta^2}{2}\right\} \\
&= \max\left\{-\left(1-\frac{1}{\gamma^*}\right)\left(r + \frac{\mu^2}{2\gamma^*\sigma^2}\right), -\left(1-\frac{1}{\gamma}\right)\left(r + \frac{\mu^2}{2\gamma\sigma^2}\right), -\frac{\zeta^2}{2}\right\} < 0.
\end{aligned}$$

Hence,

$$\lim_{T\to\infty} \mathbb{E}_{x,\bar{h}}\left[\left(\inf_{s\le t} M_s\right)^{-1/\gamma^*+1/\gamma}\left(M_T\right)^{1-1/\gamma} 1_{\{\inf_{s\le t}(M_s)\le \frac{(1-\alpha)\bar{h}^{-\gamma^*}}{y}\}}\right] = 0. \tag{A.15}$$

By (A.12), (A.13), (A.14), and (A.15), it follows that $\mathbb{E}_{x,\bar{h}}[\tilde{V}(yM_t, \hat{h}_T(y))]$ converges to 0 as $T \to \infty$. □

Finally, to complete the proof of the main Theorem 4.1 it remains to verify that the optimal spending and investment policies have the feedback form in Equations (15) and (16).

Proof of Theorem 4.1. Note that if $\tilde{V}(y,h)$ solves the PDE in (26), then $V(x,h) = \inf_{y>0}(\tilde{V}(y,h) + yx)$ solves the HJB Equation (22) with the boundary condition (23). Thus, by Itô's formula,

$$\int_0^T U(c_t, h_t)dt \le V(x,\bar{h}) - V(X_T, h_T) + \int_0^T V_x(X_t, h_t)X_t\pi_t\sigma \mathrm{d}W_t.$$

As $V_x(X_t, h_t)$ is square integrable, it follows that the last term is a martingale, and passing to the expectation:

$$\mathbb{E}_{x,\bar{h}}\left[\int_0^T U(c_t, h_t)dt\right] \leq V(x, \bar{h}) - \mathbb{E}_{x,\bar{h}}\big[V(X_T, h_T)\big]. \tag{A.16}$$

Finally,

$$V(X_T, h_T) = \inf_{y>0}(\tilde{V}(yM_T, h_T) + yM_T X_T) \geq \inf_{y>0} \tilde{V}(yM_T, h_T)$$

implies that for any $\varepsilon > 0$, there exists an $y^* > 0$ such that

$$\mathbb{E}_{x,\bar{h}}\big[V(X_T, h_T)\big] \geq \mathbb{E}_{x,\bar{h}}\left[\inf_{y>0} \tilde{V}(yM_T, h_T)\right] \geq \mathbb{E}_{x,\bar{h}}\big[\tilde{V}(y^* M_T, h_T)\big] - \varepsilon.$$

Take $T \to \infty, \varepsilon \to 0$, then

$$\lim_{T\to\infty} \mathbb{E}_{x,\bar{h}}\big[V(X_T, h_T)\big] \geq \lim_{T\to\infty} \mathbb{E}_{x,\bar{h}}\big[\tilde{V}(y^* M_T, h_T)\big] = 0.$$

Thus, passing to the limit $T \uparrow \infty$ in (A.16) it follows that

$$\mathbb{E}_{x,\bar{h}}\left[\int_0^\infty U(c_t, h_t)dt\right] \leq V(x, \bar{h}). \tag{A.17}$$

On the other hand, as $V(x, h) = \inf_{y>0}(\tilde{V}(y, h) + xy)$ and $\tilde{V}(y, h) = h^{1-\gamma^*} q(yh^{\gamma^*})$, and for the $\hat{\pi}$ and $\hat{c}_t$ in (32) and (41) the inequality in (A.17) holds as an equality, it follows that these expressions define the optimal policies and $V(x, \bar{h})$ is the value function of the problem, completing the proof. □

Proof of Theorem 4.3.

1. By the definition of the gloom ratio in terms of the function q, $g = -q'(1) = 1/m$, where m is the Merton consumption ratio, and is independent of the shortfall aversion α.
2. The bliss ratio is $b = -q'(1 - \alpha)$. Then, $\frac{\mathrm{d}b}{\mathrm{d}\alpha} = q''(1 - \alpha) > 0$, where the inequality holds because of the convexity of the function q.
3. As z does not depend on α by (17), it is obvious that the optimal investment policy is also independent of the shortfall aversion α by (16).
4. When $\alpha = 0$, then (14) implies $g = b = 1/m$. That is, the model degenerates to the Merton model. When $\alpha = 1$, again by (14), it follows that the bliss point $b = \infty$.
5. First, consider the dependence of the gloom ratio on risk aversion γ:

$$\frac{\mathrm{d}g}{\mathrm{d}\gamma} = -\frac{r}{m^2\gamma^3}\left[\gamma\left(1 - \frac{1}{\rho}\right) + \frac{2}{\rho}\right].$$

   If $0 < \rho < 1$, then g decreases for risk aversion γ close to 1, reaching its minimum at $g_{\min} = \frac{4\rho}{r(\rho+1)^2}$ for $\gamma = \frac{2}{1-\rho}$, then increasing asymptotically to $1/r$.

If $\rho \geq 1$, then the gloom ratio g keeps decreasing with respect to risk aversion and approaches asymptotically $1/r$. Consider now the dependence of the bliss ratio on risk aversion γ:

$$\frac{\mathrm{d}b}{\mathrm{d}\gamma} = -\frac{\partial q'(1-\alpha)}{\partial \gamma} = -\frac{\rho\Big((1-\alpha)^{\frac{\rho+1}{2}}(\gamma-1)+(\gamma\rho+1)\Big)\Big((1-\alpha)^{\frac{\rho+1}{2}}(1-\gamma)+(\gamma\rho+1)\Big)}{r(1-\alpha)(1-\gamma)^2(\rho+1)(\gamma\rho+1)^2}$$

$$= -\frac{\rho\Big((1-\alpha)^{\frac{\rho+1}{2}}(\gamma-1)+(\gamma\rho+1)\Big)\Big[(\rho-(1-\alpha)^{\frac{\rho+1}{2}})\gamma+((1-\alpha)^{\frac{\rho+1}{2}}+1)\Big]}{r(1-\alpha)(1-\gamma)^2(\rho+1)(\gamma\rho+1)^2}.$$

With the exception of $[(\rho-(1-\alpha)^{\frac{\rho+1}{2}})\gamma+((1-\alpha)^{\frac{\rho+1}{2}}+1)]$, which is positive for γ near 1, each term in the above equation is nonnegative. Thus, the bliss ratio decreases with respect to γ near 1. Moreover, if $\rho-(1-\alpha)^{\frac{\rho+1}{2}}<0$, then the bliss ratio decreases to its minimum

$$b_{min} = \frac{(1-\alpha)(\rho+1)^2-\Big[(1-\alpha)^{\frac{\rho+1}{2}}-\rho\Big]^2}{r(1-\alpha)(\rho+1)^2}$$

when $\gamma = \frac{(1-\alpha)^{\frac{\rho+1}{2}}+1}{(1-\alpha)^{\frac{\rho+1}{2}}-\rho}$, and then increases asymptotically to $1/r$.

If $\rho-(1-\alpha)^{\frac{\rho+1}{2}}\geq 0$, the bliss ratio keeps decreasing and approaches $1/r$ asymptotically as γ is infinite. In such a case, the bliss ratio goes to infinity when γ is close to 1. □

The extension of the above result to a positive discount rate β requires the following notation:

$$m_\beta = \frac{\beta}{\gamma}+\left(1-\frac{1}{\gamma}\right)\left(r+\frac{\mu^2}{2\gamma\sigma^2}\right),\quad m_\beta^* = \frac{\beta}{\gamma^*}+\left(1-\frac{1}{\gamma^*}\right)\left(r+\frac{\mu^2}{2\gamma^*\sigma^2}\right),$$

$$f = \frac{1}{m_\beta}-\frac{1}{r},\quad g = \frac{1}{1-\gamma}\left(\frac{1}{m_\beta}-\frac{1}{\beta}\right),$$

$$k_\pm = \frac{1}{2}+\frac{\sigma^2(r-\beta)}{\mu^2}\pm\sqrt{\left(\frac{1}{2}+\frac{\sigma^2(r-\beta)}{\mu^2}\right)^2+\frac{2\beta\sigma^2}{\mu^2}}.$$

Note that m_β is the Merton consumption fraction for to risk aversion γ, while m_β^* is the corresponding consumption fraction for risk aversion $\gamma^* = \alpha+(1-\alpha)\gamma$. For brevity, the remainder of this section uses the notation

$$q'(1-\alpha) = \lim_{z\to(1-\alpha)^+} q'(z),\ \text{ and }\ q''(1-\alpha) = \lim_{z\to(1-\alpha)^+} q''(z).$$

Lemma A.9. *If $m_\beta > 0$ and $q''(1-\alpha)\geq 0$, then the C^2 function $q : [1-\alpha,\infty)\to\mathbb{R}$ defined in the following cases, is convex and nonincreasing on $(1-\alpha,\infty)$.*

Case 1: $r\neq 0$.

$$q(z) = \begin{cases} C_1 z^{k_-}+C_2 z^{k_+}-\frac{z}{r}+\frac{1}{\beta(1-\gamma)} & 1-\alpha<z\leq 1,\\ C_3 z^{k_-}+\frac{\gamma}{(1-\gamma)m_\beta}z^{1-1/\gamma} & z>1,\end{cases}$$

where

$$C_1 = -\frac{\gamma^*(k_+ - 1) + 1}{\gamma^*(k_- - 1) + 1}(1-\alpha)^{k_+ - k_-} C_2 + \frac{(1-\alpha)^{1-k_-}}{\gamma^*(k_- - 1) + 1}\left(\frac{1}{r} - \frac{1}{\beta}\right),$$

$$C_2 = \frac{(1-k_-)f + k_- g}{k_- - k_+}, \quad C_3 = C_1 + \frac{(1-k_+)f + k_+ g}{k_- - k_+}.$$

Case 2: $r = 0$.

$$q(z) = \begin{cases} C_1 z^{k_-} + C_2 z + \frac{2z\log z}{2\beta + \mu^2/\sigma^2} + \frac{1}{\beta(1-\gamma)} & 1-\alpha < z \le 1, \\ C_4 z^{k_-} + \frac{\gamma}{(1-\gamma)m_\beta} z^{1-1/\gamma} & z > 1, \end{cases}$$

where

$$C_1 = -\frac{(1-\alpha)^{1-k_-}}{\gamma^*(k_- - 1) + 1}\left[C_2 + \frac{2(\gamma^* + \log(1-\alpha))}{2\beta + \mu^2/\sigma^2} + \frac{1}{\beta}\right],$$

$$C_2 = \frac{(1-k_-)\frac{1}{m_\beta} + k_- g + \frac{2}{2\beta + \mu^2/\sigma^2}}{k_- - 1}, \quad C_3 = C_1 + \frac{g + \frac{2}{2\beta + \mu^2/\sigma^2}}{k_- - 1}.$$

Proof. The form of the function q, as well as the fact that q is C^2, follows from the previous argument in this section. We now show that $\lim_{z\to(1-\alpha)^+} q''(z) \ge 0$ implies that q is convex and nonincreasing. This proof focuses on the case $r \ne 0$, as the case $r = 0$ is analogous. Some important inequalities are listed below, which follow by straightforward calculation:

$$k_- < 0 < 1 < k_+, \qquad (1-k_-)f + k_- g < 0, \qquad k_- - 1 + 1/\gamma < 0.$$

Therefore, $C_2 > 0$ always holds. The proof proceeds in two steps:

1. First show that q is nonincreasing assuming that it is convex. Note that, as $k_- < 0, \gamma > 0$,

$$\lim_{z\to\infty} q'(z) = \lim_{z\to\infty}\left(C_3 k_- z^{k_- - 1} - \frac{1}{m_\beta} z^{-1/\gamma}\right) = 0.$$

   As q is C^2 and convex on $(1-\alpha, \infty)$, it follows that $q'(z) \le 0,\ \forall z \in (1-\alpha, \infty)$. That is, q is nonincreasing on $(1-\alpha, \infty)$.

2. Show the convexity of q on $(1-\alpha, \infty)$ by proving that

$$q''(z) \ge 0,\ \forall z \in (1-\alpha, \infty).$$

   First, consider q on $[1, \infty)$:

$$q''(z) = C_3 k_-(k_- - 1) z^{k_- - 2} + \frac{1}{\gamma m_\beta} z^{-\frac{1}{\gamma} - 1} = z^{k_- - 2}\left[C_3 k_-(k_- - 1) + \frac{1}{\gamma m_\beta} z^{-\frac{1}{\gamma} - k_- + 1}\right].$$

   By (A.18), the term $[C_3 k_-(k_- - 1) + \frac{1}{\gamma m_\beta} z^{-\frac{1}{\gamma} - k_- + 1}]$ is increasing in z for all $z \in [1, \infty)$. Then $C_3 k_-(k_- - 1) + \frac{1}{\gamma m_\beta} z^{-\frac{1}{\gamma} - k_- + 1} \ge 0, \forall z \in [1, \infty)$ if and only if $C_3 k_-(k_- - 1) + \frac{1}{\gamma m_\beta}(1)^{-\frac{1}{\gamma} - k_- + 1} \ge 0$.

That is, $q''(z) \geq 0$ for all $z \in [1, \infty)$ if and only if $q''(1) \geq 0$. Applying a similar argument to q on $(1-\alpha, 1]$, it follows that $q''(z) \geq 0$ for all $z \in (1-\alpha, 1]$ if and only if $q''(1-\alpha) = \lim_{z \to (1-\alpha)^+} q''(z) \geq 0$. Summing up, $q''(1-\alpha) \geq 0$ implies the convexity of q on $(1-\alpha, \infty)$. As q is C^2 on $(1-\alpha, \infty)$, the proof is complete.

□

The proof of Theorem 9.1 is analogous to the one of Theorem 4.1, hence omitted.

A.2 Long-run properties of the optimal policy

This section investigates the dynamics of process h_t/X_t, from which the dynamics of $\hat{c}$, $\hat{\pi}$ follow. First, define the processes $(R_t)_{t\geq 0}, (R_t^*)_{t\geq 0}, (z_t)_{t\geq 0}$ as

$$R_t = \frac{\hat{c}_t}{X_t},\ R_t^* = \frac{\hat{h}_t}{X_t},\ z_t = (\hat{h}_t)^{\gamma^*} V_x(X_t, \hat{h}_t).$$

Then $\pi_t = -\frac{z_t q''(z_t)}{q'(z_t)} \frac{\mu}{\sigma^2}$, and the dynamics of the wealth process X_t is:

$$\frac{\mathrm{d}X_t}{X_t} = \left(r - R_t - \frac{z_t q''(z_t)}{q'(z_t)} \frac{\mu^2}{\sigma^2}\right)\mathrm{d}t - \frac{z_t q''(z_t)}{q'(z_t)} \frac{\mu}{\sigma}\mathrm{d}W_t. \tag{A.18}$$

The dynamics of R^* is

$$\frac{\mathrm{d}R_t^*}{R_t^*} = X_t \mathrm{d}\left(\frac{1}{X_t}\right) + \frac{\mathrm{d}\hat{h}_t}{\hat{h}_t}.$$

Therefore,

$$\frac{\mathrm{d}R_t^*}{R_t^*} = \left(-r + R_t + \frac{z_t q''(z_t)}{q'(z_t)} \frac{\mu^2}{\sigma^2} + \frac{(z_t q''(z_t))^2}{(q'(z_t))^2} \frac{\mu^2}{\sigma^2}\right)\mathrm{d}t + \frac{z_t q''(z_t)}{q'(z_t)} \frac{\mu}{\sigma}\mathrm{d}W_t + \frac{\mathrm{d}\hat{h}_t}{\hat{h}_t}.$$

Now, consider the process z. Recall that $R^* = -\frac{1}{q'(z)}$, or $z = p(-1/R^*)$, with p being the inverse function of q'. First, apply Itô's formula to $-1/R^*$:

$$\mathrm{d}\left(-\frac{1}{R_t^*}\right) = \frac{1}{R_t^*}\left(-r + R_t + \frac{z_t q''(z_t)}{q'(z_t)} \frac{\mu^2}{\sigma^2}\right)\mathrm{d}t + \frac{1}{R_t^*} \frac{z_t q''(z_t)}{q'(z_t)} \frac{\mu}{\sigma}\mathrm{d}W_t + \frac{1}{R_t^*} \frac{\mathrm{d}\hat{h}_t}{\hat{h}_t}.$$

Then, apply Itô's formula again to $z = p(-1/R^*)$. As p is the inverse function of q',

$$p' = \frac{1}{q''},\ p'' = -\frac{q'''}{(q'')^3}.$$

Therefore, (note $q'(z_t) = -1/R_t^*$)

$$\mathrm{d}z_t = \left(\frac{q'(z_t)}{q''(z_t)}(r - R_t) - \frac{\mu^2}{\sigma^2} z_t - \frac{\mu^2}{2\sigma^2} \frac{z_t^2 q'''(z_t)}{q''(z_t)}\right)\mathrm{d}t - \frac{\mu}{\sigma} z_t \mathrm{d}W_t + \gamma^* z_t \frac{\mathrm{d}\hat{h}_t}{\hat{h}_t}. \tag{A.19}$$

Recall the HJB equation implies that, differentiating again,

$$\frac{\mu^2}{2\sigma^2}\cdot\frac{z^2q'''(z)}{q''(z)}+\left(\frac{\mu^2}{\sigma^2}-r\right)z-\frac{rq'(z)}{q''(z)}=\begin{cases}\dfrac{1}{q''(z)} & \text{if } 1-\alpha\le z\le 1,\\ \dfrac{z^{-1/\gamma}}{q''(z)} & \text{if } z>1.\end{cases}\tag{A.20}$$

The drift term of the process z in (A.19) simplifies as follows:

1. For $1-\alpha\le z_t\le 1$, that is, between the bliss and gloom, $R_t=R_t^*=-1/q'(z_t)$. By (A.20), the drift term simplifies to $-rz_t$.
2. For $z_t>1$, $R_t=R_t^*z_t^{-1/\gamma}=-z_t^{-1/\gamma}/q'(z_t)$. Again, by (A.20), the drift term simplifies to $-rz_t$.

Therefore, the drift is always $-rz_t$ in both cases. The dynamics of the process z_t follows:

$$z_t=z_0-\int_0^t\frac{\mu}{\sigma}z_s\mathrm{d}W_s-\int_0^t rz_s\mathrm{d}s+\gamma^*\int_0^t z_s\frac{\mathrm{d}\hat{h}_s}{\hat{h}_s}.$$

Note that this stochastic differential equation of z does not depend on risk aversion γ before it hits the boundary $(1-\alpha)$. To understand the long-run properties of the process z, define the *scale function* $\mathfrak{s}^z$ and *speed measure* m^z of z as follows:

$$\mathfrak{s}^z(z)=\frac{1-\alpha}{1+\rho}\left[\left(\frac{z}{1-\alpha}\right)^{1+\rho}-1\right],\qquad m^z(\mathrm{d}z)=\frac{2\rho(1-\alpha)^\rho}{r}z^{-\rho-2}.$$

Lemma A.10. *The process z is positively recurrent, with the invariant density*

$$\nu(\mathrm{d}z)=1_{\{1-\alpha\le z<\infty\}}\frac{1+\rho}{1-\alpha}\left(\frac{z}{1-\alpha}\right)^{-\rho-2}.\tag{A.21}$$

Proof. The recurrence of z follows by Karatzas and Shreve (1991, Proposition 5.5.22). Moreover, a direct calculation shows that:

$$\int_{1-\alpha}^\infty m(dz)<\infty.$$

By Borodin and Salminen (2002, II.2.12), the process z is positively recurrent, and its invariant distribution is the normalized speed measure:

$$\nu(\mathrm{d}z)=\frac{m(\mathrm{d}z)}{m([1-\alpha,\infty))}=1_{\{1-\alpha\le z<\infty\}}\frac{1+\rho}{1-\alpha}\left(\frac{z}{1-\alpha}\right)^{-\rho-2}.\tag{A.22}$$

□

Proof of Theorem 5.1.

1. Recall that when the spending is in the target region, the process z is between 0 and 1. By the ergodic theorem and the form of $\nu(\cdot)$ in (A.22), a direct calculation yields

$$\lim_{T\to\infty}\frac{\int_0^T 1_{\{1-\alpha<z_s<1\}}\mathrm{d}s}{T}=\int_{1-\alpha}^1\nu(z)\mathrm{d}z=1-(1-\alpha)^{1+\rho}.$$

2. Let $g(z)$ be the solution to the following ODE with boundary conditions

$$\frac{\mu^2}{2\sigma^2} z^2 g''(z) - rzg'(z) = -1, \qquad \text{for } z \in (1-\alpha, 1)$$
$$g(1) = 0, \quad g'(1-\alpha) = 0,$$

which is

$$g(z) = \frac{\rho}{(\rho+1)r}\left(\log(z) - \frac{(1-\alpha)^{-\rho-1}\left(z^{\rho+1}-1\right)}{\rho+1}\right).$$

Applying Itô's formula to $g(z_t)$, and integrating from 0 to τ_{gloom},

$$g(z_{\tau_{gloom}}) - g(z_0) = -\tau_{gloom} - \frac{\mu}{\sigma}\int_0^{\tau_{gloom}} z_s g'(z_s)\mathrm{d}W_s + \gamma^* \int_0^{\tau_{gloom}} \frac{z_s g'(z_s)}{\hat{h}_s}\mathrm{d}\hat{h}_s.$$

Taking expectations, note that:
(a) $g(z_{\tau_{gloom}}) = g(1) = 0$.
(b) The stochastic integral is square-integrable, therefore it is a martingale, and has zero mean.
(c) $\hat{h}_t$ only increases on $\{z_t = 1-\alpha\}$, but $g'(1-\alpha) = 0$, implying that $\int_0^{\tau_{gloom}} \frac{z_s g'(z_s)}{\hat{h}_s}\mathrm{d}\hat{h}_s = 0$.

Therefore, $\mathbb{E}_{x,\bar{h}}[\tau_{gloom}] = g(z_0)$ as desired.

3. Note that the first time to reach admits the equivalent definition $\tau_{bliss} = \inf\{t \in \mathbb{R}_+ \mid z_t = 1-\alpha\}$, and recall that the process z reflects only at the boundary $1-\alpha$. Hence, before τ_{bliss}, the process z is a geometric Brownian motion, that is, $z_t = z_0 \exp\{-\frac{\mu}{\sigma}W_t - (r + \frac{\mu^2}{2\sigma^2})t\}$. Then

$$\{z_t \le 1-\alpha\} = \left\{-\frac{\mu}{\sigma}W_t - \left(r + \frac{\mu^2}{2\sigma^2}\right)t \le -\ln\left(\frac{z_0}{1-\alpha}\right)\right\}.$$

Let B be a 1-dimensional Brownian motion under $\mathbb{P}$. Then, by Rogers and Williams (2000, Equation 9.3), with $a = -\infty$, $b = \frac{1}{\sqrt{(\mu/\sigma)^2}}\ln(\frac{z_0}{1-\alpha})$, $c = \frac{r}{\sqrt{(\mu/\sigma)^2}} + \frac{\sqrt{(\mu/\sigma)^2}}{2}$, $\beta = \sqrt{c^2 + 2\lambda} - c$, it follows that for any $\lambda > 0$:

$$\mathbb{E}[e^{-\lambda\tau_{bliss}}] = e^{-b\beta}.$$

Then, as desired,

$$\mathbb{E}[\tau_{bliss}] = -\frac{\mathrm{d}\mathbb{E}[e^{-\lambda\tau_{bliss}}]}{\mathrm{d}\tau}\Big|_{\tau\downarrow 0} = \frac{b}{c} = \frac{\rho}{r(\rho+1)}\log\left(\frac{z_0}{1-\alpha}\right).$$ □

Lemma A.11.

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\mathbb{E}[\log X_T] =& r - r(1-\alpha)^{\rho+1}\left(\frac{\gamma\rho-\rho-1}{\gamma\rho}\right)\\ &+ (1+\rho)(1-\alpha)^{\rho+1}\int_{1-\alpha}^1 \frac{1}{z^{\rho+1}}\left\{\frac{1}{q'(z)} - \frac{zq''(z)}{q'(z)}\frac{\mu^2}{\sigma^2}\left(1+\frac{zq''(z)}{2q'(z)}\right)\right\}\mathrm{d}z.\end{aligned}$$

Proof. Apply Itô's formula to $\log X_T$, together with the dynamic of the wealth process under our optimal strategy in Equation (A.18):

$$\log X_T = \int_0^T r - R_t - \frac{z_t q''(z_t)}{q'(z_t)} \frac{\mu^2}{\sigma^2}\left(1 + \frac{z_t q''(z_t)}{2q'(z_t)}\right) \mathrm{d}t - \int_0^T \frac{z_t q''(z_t)}{q'(z_t)} \frac{\mu}{\sigma} \mathrm{d}W_t.$$

Note that the stochastic integral is square-integrable, hence a martingale with mean zero. Moreover,

$$R_t = \begin{cases} -1/q'(z_t) & z_t \in [1-\alpha, 1), \\ -z_t^{-1/\gamma}/q'(z_t) & z_t \in [1, \infty). \end{cases}$$

Then, using the ergodic theorem, and recalling the invariant density $\nu(z)$ of z in (A.21),

$$\begin{aligned} \lim_{T\to\infty} \frac{1}{T}\mathbb{E}[\log X_T] &= \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T r - R_t - \frac{z_t q''(z_t)}{q'(z_t)} \frac{\mu^2}{\sigma^2}\left(1 + \frac{z_t q''(z_t)}{2z'(z_t)}\right) \mathrm{d}t\right] \\ &= \int_{1-\alpha}^1 \left\{r + \frac{1}{q'(z)} - \frac{zq''(z)}{q'(z)} \frac{\mu^2}{\sigma^2}\left(1 + \frac{zq''(z)}{2q'(z)}\right)\right\} \nu(z)\mathrm{d}z \\ &\quad + \int_1^\infty \left\{r + \frac{z^{-1/\gamma}}{q'(z)} - \frac{zq''(z)}{q'(z)} \frac{\mu^2}{\sigma^2}\left(1 + \frac{zq''(z)}{2q'(z)}\right)\right\} \nu(z)\mathrm{d}z \\ &= r - r(1-\alpha)^{\rho+1}\left(\frac{\gamma\rho - \rho - 1}{\gamma\rho}\right) + (1+\rho)(1-\alpha)^{\rho+1} \\ &\quad \int_{1-\alpha}^1 \frac{1}{z^{\rho+1}}\left\{\frac{1}{q'(z)} - \frac{zq''(z)}{q'(z)} \frac{\mu^2}{\sigma^2}\left(1 + \frac{zq''(z)}{2q'(z)}\right)\right\} \mathrm{d}z. \end{aligned}$$

□

Proof of Theorem 5.2. The expected return on wealth is the safe interest rate r plus the term

$$\int_{1-\alpha}^\infty \pi(z)\nu(z)\mu \mathrm{d}z,$$

where ν is the invariant distribution density of z as in (A.21). As $\pi(z) = -\frac{zq''(z)}{q'(z)} \frac{\mu}{\sigma^2}$, then

$$\int_{1-\alpha}^\infty \pi(z)\nu(z)\mu \mathrm{d}z = -\frac{2r(1-\alpha)^{\rho+1}(\rho+1)}{\rho} \int_{1-\alpha}^\infty \frac{q''(z)}{z^{\rho+1} q'(z)} \mathrm{d}z$$

and the claim follows. □

Theorem A.12. *The weight on the risky asset π_t in (16) is increasing in the target–wealth ratio x/h.*

Proof. Recall from (41) that the risky weight is $\pi(z) = -\frac{zq''(z)}{q'(z)} \frac{\mu}{\sigma^2}$, where $q'(z) = -\frac{x}{h}$ and hence

$$\frac{d\pi}{dx} = \frac{d\pi}{dz}\frac{dz}{dx} = \left(\frac{zq''(z)}{q'(z)}\right)' \frac{\mu}{\sigma^2} \frac{1}{hq''(z)}.$$

As q is convex, it remains to check that $(\frac{zq''(z)}{q'(z)})' > 0$. Because

$$\left(\frac{zq''(z)}{q'(z)}\right)' = -\frac{(\gamma-1)\rho(\rho+1)(\gamma\rho+1)(-\gamma\rho+z^\rho(\rho((\gamma-1)z+1)+1)-1)}{\left((\gamma-1)z^{\rho+1}-(\gamma\rho+1)(\rho+(\gamma-1)(\rho+1)z)\right)^2},$$

it is in turn enough to check that $-\gamma\rho + z^\rho(\rho((\gamma-1)z+1)+1)-1 \le 0$ for $z \in [1-\alpha, 1]$. (Note that when $z > 1$, the wealth to target ratio is lower than the gloom point, and the weight of the risky asset is the Merton weight.) This condition is clearly satisfied for $\gamma = 1$, as this expression reduces to $(\rho + 1)(z^\rho - 1)$. In addition, the expression is decreasing in γ, because its partial derivative is $\rho(z^{\rho+1}-1) \le 0$ and therefore is less than or equal to zero for $\gamma \ge 1$. □

Proof of Proposition 6.1. The Merton strategy $c_t = mX_t$, $\pi_t = \frac{\mu}{\gamma\sigma^2}$ leads to the corresponding wealth process $X_t = x\exp\{\frac{1}{\gamma}(r+\frac{\mu^2}{2\sigma^2})t + \frac{\mu}{\gamma\sigma}W_t\}$. Let $W_t^{(\zeta)} = \zeta t + W_t$, where $\zeta = \frac{r\sigma}{\mu} + \frac{\mu}{2\sigma}$. Denote the corresponding maximum process by $(W_t^{(\zeta)})^* = \sup_{0\le s\le t} W_s^{(\zeta)}$. Then the resulting target spending equals to $h_t = \bar{h} \vee mx\exp\{\frac{\mu}{\gamma\sigma}(W_t^{(\zeta)})^*\}$ and its utility is

$$\begin{aligned} U_t = \frac{1}{1-\gamma}\left(\frac{c_t}{h_t^\alpha}\right)^{1-\gamma} &= \frac{1}{1-\gamma}\left(\frac{mx\exp\{\frac{\mu}{\gamma\sigma}W_t^{(\zeta)}\}}{\bar{h}^\alpha \vee (mx)^\alpha \exp\left\{\frac{\alpha\mu}{\gamma\sigma}\left(W_t^{(\zeta)}\right)^*\right\}}\right)^{1-\gamma} \\ &= \frac{1}{1-\gamma}\left[\left(\frac{mx}{\bar{h}^\alpha}\right)^{1-\gamma} U_t^{(1)} + (mx)^{1-\gamma^*}U_t^{(2)}\right], \end{aligned}$$

where

$$\begin{aligned} U_t^{(1)} &= \exp\left\{\frac{(1-\gamma)\mu}{\gamma\sigma}W_t^{(\zeta)}1_{\{(W_t^{(\zeta)})^* < \kappa_0\}}\right\} \\ U_t^{(2)} &= \exp\left\{\frac{(1-\gamma)\mu}{\gamma\sigma}\left(W_t^{(\zeta)} - \alpha\left(W_t^{(\zeta)}\right)^*\right)1_{\{(W_t^{(\zeta)})^* \ge \kappa_0\}}\right\}, \qquad \kappa_0 = \frac{\gamma\sigma}{\mu}\ln\left(\frac{\bar{h}}{mx}\right). \end{aligned}$$

Then the expected utility of the Merton strategy is

$$\begin{aligned} U^M(x,\alpha,\gamma,h) &= \mathbb{E}\int_0^\infty \frac{1}{1-\gamma}\left[\left(\frac{mx}{\bar{h}^\alpha}\right)^{1-\gamma} U_t^{(1)} + (mx)^{1-\gamma^*}U_t^{(2)}\right]\mathrm{d}t \\ &= \frac{1}{1-\gamma}\int_0^\infty \left(\frac{mx}{\bar{h}^\alpha}\right)^{1-\gamma}\mathbb{E}_{x,\bar{h}}\left[U_t^{(1)}\right] + (mx)^{1-\gamma^*}\mathbb{E}_{x,\bar{h}}\left[U_t^{(2)}\right]\mathrm{d}t. \end{aligned}$$

Applying Corollary A.7, it follows that

$$\mathbb{E}_{x,\bar{h}}\left[U_t^{(1)}\right] = \mathbb{E}\exp\left\{\frac{(1-\gamma)\mu}{\gamma\sigma}W_t^{(\zeta)}\right\} - \mathbb{E}\exp\left\{\frac{(1-\gamma)\mu}{\gamma\sigma}W_t^{(\zeta)}1_{\left\{\left(W_t^{(\zeta)}\right)^*>\kappa_0\right\}}\right\}$$
$$= e^{-mt}\left[\Phi\left(-\kappa_1\sqrt{t}+\kappa_0/\sqrt{t}\right) - e^{2\kappa_0\kappa_1}\Phi\left(-\kappa_1\sqrt{t}-\kappa_0/\sqrt{t}\right)\right],$$
$$\mathbb{E}_{x,\bar{h}}\left[U_t^{(2)}\right] = \frac{2\kappa_2}{\kappa_1+\kappa_2}\exp\left\{-mt - \frac{(1-\gamma)\alpha\mu}{2\gamma\sigma}(\kappa_1+\kappa_2)t\right\}\Phi\left(\kappa_2\sqrt{t}-\kappa_0/\sqrt{t}\right)$$
$$+\frac{2\kappa_1}{\kappa_1+\kappa_2}\exp\left\{\kappa_0(\kappa_1+\kappa_2)-mt\right\}\Phi\left(-\kappa_1\sqrt{t}-\kappa_0/\sqrt{t}\right),$$

where

$$\kappa_1 = \frac{r\sigma}{\mu} + \frac{(2-\gamma)\mu}{2\gamma\sigma}, \quad \kappa_2 = \frac{r\sigma}{\mu} + \frac{(2-2\gamma^*+\gamma)\mu}{2\gamma\sigma}.$$

To calculate the expected utility in (A.23), it remains to evaluate the following integrals

$$\int_0^\infty \mathbb{E}_{x,\bar{h}}\left[U_t^{(1)}\right]\mathrm{d}t = \frac{1}{m} - \frac{\kappa_0\exp\left\{\kappa_0\left(\kappa_1 - \frac{r\sigma}{\mu} - \frac{\mu}{2\sigma}\right)\right\}}{\sqrt{2\pi}m}\int_0^\infty\left(\frac{1}{\sqrt{t^3}}\right)$$
$$\exp\left\{-\frac{1}{2}\left[\left(\frac{r\sigma}{\mu}+\frac{\mu}{2\sigma}\right)\sqrt{t}-\frac{\kappa_0}{\sqrt{t}}\right]^2\right\}\mathrm{d}t, \tag{A.23}$$

$$\int_0^\infty \mathbb{E}_{x,\bar{h}}\left[U_t^{(2)}\right]\mathrm{d}t = \frac{2\exp\left\{\kappa_0\left(\kappa_2 - \frac{r\sigma}{\mu} - \frac{\mu}{2\sigma}\right)\right\}}{\sqrt{2\pi}(\kappa_1+\kappa_2)}\int_0^\infty\left(\left(\frac{\kappa_2^2}{\tilde{m}} - \frac{\kappa_1^2}{m}\right)\frac{1}{2\sqrt{t}} + \left(\frac{\kappa_2}{\tilde{m}}+\frac{\kappa_1}{m}\right)\frac{\kappa_0}{2\sqrt{t^3}}\right)$$
$$\exp\left\{-\frac{1}{2}\left[\left(\frac{r\sigma}{\mu}+\frac{\mu}{2\sigma}\right)\sqrt{t}-\frac{\kappa_0}{\sqrt{t}}\right]^2\right\}\mathrm{d}t \tag{A.24}$$

with $\tilde{m} = \dfrac{\gamma^*-1}{\gamma}(r+(1-\alpha+\alpha\gamma)\frac{\mu^2}{2\gamma\sigma^2})$. To compute the two integrals on the right-hand side of the above equations, let $u = (\frac{r\sigma}{\mu}+\frac{\mu}{2\sigma})\sqrt{t} - \frac{\kappa_0}{\sqrt{t}}$, then $\sqrt{t} = \dfrac{u+\sqrt{u^2+4\kappa_0(\frac{r\sigma}{\mu}+\frac{\mu}{2\sigma})}}{2(\frac{r\sigma}{\mu}+\frac{\mu}{2\sigma})}$, and

$$\frac{\mathrm{d}t}{\sqrt{t^3}} = \left(\frac{1}{\kappa_0} - \frac{u}{\kappa_0\sqrt{u^2+4\kappa_0\left(\frac{r\sigma}{\mu}+\frac{\mu}{2\sigma}\right)}}\right)\mathrm{d}u \qquad \frac{\mathrm{d}t}{\sqrt{t}} = \frac{1}{\frac{r\sigma}{\mu}+\frac{\mu}{2\sigma}}\left(1+\frac{u}{\sqrt{u^2+4\kappa_0\left(\frac{r\sigma}{\mu}+\frac{\mu}{2\sigma}\right)}}\right)\mathrm{d}u.$$

It follows that

$$\int_0^\infty \left(\frac{1}{\sqrt{t^3}}\right) \exp\left\{-\frac{1}{2}\left[\left(\frac{r\sigma}{\mu}+\frac{\mu}{2\sigma}\right)\sqrt{t}-\frac{\kappa_0}{\sqrt{t}}\right]^2\right\}\mathrm{d}t$$

$$=\int_{-\infty}^\infty \exp\left\{-\frac{u^2}{2}\right\}\left(\frac{1}{\kappa_0}-\frac{u}{\kappa_0\sqrt{u^2+4\kappa_0\left(\frac{r\sigma}{\mu}+\frac{\mu}{2\sigma}\right)}}\right)\mathrm{d}u$$

$$=\frac{\sqrt{2\pi}}{\kappa_0}\int_0^\infty \left(\frac{1}{\sqrt{t}}\right)\exp\left\{-\frac{1}{2}\left[\left(\frac{r\sigma}{\mu}+\frac{\mu}{2\sigma}\right)\sqrt{t}-\frac{\kappa_0}{\sqrt{t}}\right]^2\right\}\mathrm{d}t$$

$$=\frac{1}{\frac{r\sigma}{\mu}+\frac{\mu}{2\sigma}}\int_{-\infty}^\infty \exp\left\{-\frac{u^2}{2}\right\}\left(1+\frac{u}{\sqrt{u^2+4\kappa_0\left(\frac{r\sigma}{\mu}+\frac{\mu}{2\sigma}\right)}}\right)\mathrm{d}u=\frac{\sqrt{2\pi}}{\frac{r\sigma}{\mu}+\frac{\mu}{2\sigma}}.$$

Hence, by (A.23) and (A.24)

$$\int_0^\infty \mathbb{E}_{x,\bar{h}}\left[U_t^{(1)}\right]\mathrm{d}t=\frac{1-\exp\left\{\kappa_0\left(\kappa_1-\frac{r\sigma}{\mu}-\frac{\mu}{2\sigma}\right)\right\}}{m}\qquad \int_0^\infty \mathbb{E}_{x,\bar{h}}\left[U_t^{(2)}\right]\mathrm{d}t=\frac{\exp\left\{\kappa_0\left(\kappa_2-\frac{r\sigma}{\mu}-\frac{\mu}{2\sigma}\right)\right\}}{(1-\alpha)m}.$$

Substituting these expressions into (A.23) leads to an expected utility equal to

$$U^M(x,\alpha,\gamma,h)=\frac{(x\bar{h}^{-\alpha})^{1-\gamma}}{1-\gamma}m^{-\gamma}\left(1+\frac{\alpha}{1-\alpha}\left(\frac{\bar{h}}{mx}\right)^{1-\gamma}\right). \tag{A.25}$$

(For $\bar{h}=mx$, this expression reduces to $\frac{x^{1-\gamma^*}}{1-\gamma^*}m^{-\gamma^*}$.) To obtain the equivalent relative loss, recall that the value function has the expression

$$U^{OP}(x,\alpha,\gamma,h)=V(x,\bar{h})=\bar{h}^{(1-\gamma)(1-\alpha)}\inf_{z>0}(q(z)+\frac{x}{\bar{h}}z),$$

and hence

$$V(x,h)=h^{(1-\gamma)(1-\alpha)}(q(z)-q'(z)z),\ \text{where } z \text{ satisfies } q'(z)=-\frac{x}{h}.$$

Setting the above expression for $V(x,h)$ equal to (A.25), (21) follows, while the expression in (20) results from replacing x with $x(1-L)$ in the first-order condition $q'(z)=-\frac{x}{h}$ and solving for L. □

Solution of the finite-horizon PDE The explicit solution of the PDE (39) with (40) is

$$q(z,t)=\begin{cases} q_0(z,\tau)+\dfrac{(e^{-r\tau}-1)z}{r}+\dfrac{1-e^{-\beta\tau}}{\beta(1-\gamma)}, & 1-\alpha<z<1,\\[2ex] q_0(z,\tau)-2(F_1+F_2+F_3+F_4+F_5)+\dfrac{\gamma(1-e^{-m\tau})}{(1-\gamma)m}z^{1-\frac{1}{\gamma}}, & z\geq 1,\end{cases}$$

and $q_0(z,\tau)$ equals

$$
\begin{aligned}
&F_1\cdot \operatorname{erfc}\left(\frac{\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}}{2}\sqrt{\frac{\tau}{r\rho}}-\frac{x}{2}\sqrt{\frac{\rho}{r\tau}}\right)\\
&\quad+F_2\cdot \operatorname{erfc}\left(-\frac{\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}}{2}\sqrt{\frac{\tau}{r\rho}}-\frac{x}{2}\sqrt{\frac{\rho}{r\tau}}\right)\\
&\quad+F_3\cdot \operatorname{erfc}\left(\frac{r+r\rho-\beta\rho}{2}\sqrt{\frac{\tau}{r\rho}}-\frac{x}{2}\sqrt{\frac{\rho}{r\tau}}\right)+F_4\cdot \operatorname{erfc}\left(-\frac{r-r\rho+\beta\rho}{2}\sqrt{\frac{\tau}{r\rho}}-\frac{x}{2}\sqrt{\frac{\rho}{r\tau}}\right)\\
&\quad+F_5\cdot \operatorname{erfc}\left(\frac{2r-\gamma(r-r\rho+\beta\rho)}{2\gamma}\sqrt{\frac{\tau}{r\rho}}-\frac{x}{2}\sqrt{\frac{\rho}{r\tau}}\right)\\
&\quad+G_1\cdot \operatorname{erfc}\left(-\frac{\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}}{2}\sqrt{\frac{\tau}{r\rho}}+\frac{w}{2}\sqrt{\frac{\rho}{r\tau}}\right)\\
&\quad+G_2\cdot \operatorname{erfc}\left(\frac{\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}}{2}\sqrt{\frac{\tau}{r\rho}}+\frac{w}{2}\sqrt{\frac{\rho}{r\tau}}\right)\\
&\quad+G_3\cdot \operatorname{erfc}\left(-\frac{r+r\rho-\beta\rho}{2}\sqrt{\frac{\tau}{r\rho}}+\frac{w}{2}\sqrt{\frac{\rho}{r\tau}}\right)\\
&\quad+G_4\cdot \operatorname{erfc}\left(\frac{r+r\rho-\beta\rho}{2}\sqrt{\frac{\tau}{r\rho}}+\frac{w}{2}\sqrt{\frac{\rho}{r\tau}}\right)+G_5\cdot \operatorname{erfc}\left(-\frac{r-r\rho+\beta\rho}{2}\sqrt{\frac{\tau}{r\rho}}+\frac{w}{2}\sqrt{\frac{\rho}{r\tau}}\right)\\
&\quad+G_6\cdot \operatorname{erfc}\left(\frac{r-r\rho+\beta\rho}{2}\sqrt{\frac{\tau}{r\rho}}+\frac{w}{2}\sqrt{\frac{\rho}{r\tau}}\right)\\
&\quad+G_7\cdot \operatorname{erfc}\left(-\frac{2r-\gamma^*(r-r\rho+\beta\rho)}{2\gamma^*}\sqrt{\frac{\tau}{r\rho}}+\frac{w}{2}\sqrt{\frac{\rho}{r\tau}}\right)\\
&\quad+H_1\cdot \operatorname{erfc}\left(-\frac{\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}}{2}\sqrt{\frac{\tau}{r\rho}}+\frac{u}{2}\sqrt{\frac{\rho}{r\tau}}\right)\\
&\quad+H_2\cdot \operatorname{erfc}\left(\frac{\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}}{2}\sqrt{\frac{\tau}{r\rho}}+\frac{u}{2}\sqrt{\frac{\rho}{r\tau}}\right)\\
&\quad+H_3\cdot \operatorname{erfc}\left(\frac{r+r\rho-\beta\rho}{2}\sqrt{\frac{\tau}{r\rho}}+\frac{u}{2}\sqrt{\frac{\rho}{r\tau}}\right)+H_4\cdot \operatorname{erfc}\left(-\frac{r-r\rho+\beta\rho}{2}\sqrt{\frac{\tau}{r\rho}}+\frac{u}{2}\sqrt{\frac{\rho}{r\tau}}\right)\\
&\quad+H_5\cdot \operatorname{erfc}\left(-\frac{|2r-\gamma(r-r\rho+\beta\rho)|}{2\gamma}\sqrt{\frac{\tau}{r\rho}}+\frac{u}{2}\sqrt{\frac{\rho}{r\tau}}\right)\\
&\quad+H_6\cdot \operatorname{erfc}\left(\frac{|2r-\gamma(r-r\rho+\beta\rho)|}{2\gamma}\sqrt{\frac{\tau}{r\rho}}+\frac{u}{2}\sqrt{\frac{\rho}{r\tau}}\right)
\end{aligned}
$$

$$+H_7 \cdot \operatorname{erfc}\left(-\frac{|2r-\gamma^*(r-r\rho+\beta\rho)|}{2\gamma^*}\sqrt{\frac{\tau}{r\rho}}+\frac{u}{2}\sqrt{\frac{\rho}{r\tau}}\right)$$

$$+H_8 \cdot \operatorname{erfc}\left(\frac{|2r-\gamma^*(r-r\rho+\beta\rho)|}{2\gamma^*}\sqrt{\frac{\tau}{r\rho}}+\frac{u}{2}\sqrt{\frac{\rho}{r\tau}}\right),$$

where

$$\tau = T-t, \qquad w = x-\log(1-\alpha), \qquad u = x-2\log(1-\alpha),$$

$$\begin{aligned}
F_1 =& e^{\frac{x}{2r}\left(r+r\rho-\beta\rho-\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}\right)} \\
&\times\left\{-\frac{m(r-\beta+\gamma\beta)-\gamma r\beta}{4(1-\gamma)r\beta m}-\frac{r[m(r+\beta-\gamma\beta)-(2-\gamma)r\beta]+\rho(r-\beta)[m(r-\beta+\gamma\beta)-\gamma r\beta]}{4(1-\gamma)r\beta m\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}}\right\},
\end{aligned}$$

$$\begin{aligned}
F_2 =& -e^{\frac{x}{2r}\left(r+r\rho-\beta\rho+\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}\right)} \\
&\times\left\{\frac{m(r-\beta+\gamma\beta)-\gamma r\beta}{4(1-\gamma)r\beta m}-\frac{r[m(r+\beta-\gamma\beta)-(2-\gamma)r\beta]+\rho(r-\beta)[m(r-\beta+\gamma\beta)-\gamma r\beta]}{4(1-\gamma)r\beta m\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}}\right\},
\end{aligned}$$

$$F_3 = \frac{e^{-\beta\tau}}{2(1-\gamma)\beta}, \qquad F_4 = -\frac{e^{-r\tau+x}}{2r}, \qquad F_5 = -\frac{\gamma}{2(1-\gamma)m}e^{-m\tau+\frac{(\gamma-1)x}{\gamma}},$$

$$G_1 = e^{\frac{r+r\rho-\beta\rho-\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}}{2r}w} \times \frac{(1-\alpha)(r-\beta)[2r-\gamma^*(r-r\rho+\beta\rho)+\gamma^*\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}]}{4(\gamma^*)^2 r\beta\rho m^*},$$

$$G_2 = e^{\frac{r+r\rho-\beta\rho+\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}}{2r}w} \times \frac{(1-\alpha)(r-\beta)[2r-\gamma^*(r-r\rho+\beta\rho)-\gamma^*\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}]}{4(\gamma^*)^2 r\beta\rho m^*},$$

$$G_3 = \frac{e^{-\beta\tau}}{2\beta(1-\gamma)}, \qquad G_4 = \frac{(1-\alpha)re^{-\beta\tau+\frac{r+r\rho-\beta\rho}{r}w}}{2\beta(r+\gamma^*\rho(r-\beta))}, \qquad G_5 = \frac{(1-\alpha)e^{-r\tau+\frac{(r-\beta)\rho}{r}w}}{2[\gamma^*(r-r\rho+\beta\rho)-r]},$$

$$G_6 = -\frac{(1-\alpha)e^{-r\tau+w}}{2r},$$

$$G_7 = -\frac{(\gamma^*)^2\rho(r-\beta)[2r-\gamma^*(r-r\rho+\beta\rho)]}{2(1-\gamma)m^*(r+\gamma^*\rho(r-\beta))[r-\gamma^*(r-r\rho+\beta\rho)]}e^{-m^*\tau+(1-\frac{1}{\gamma^*})w},$$

$$\begin{aligned}
H_1 =& \frac{(1-\alpha)^{\frac{r+(r-\beta)\rho}{r}} e^{\frac{r+r\rho-\beta\rho-\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}}{2r}u}}{4\beta[\gamma\rho\beta-r(1-\gamma)(1+\gamma\rho)]} \\
&\times\left\{\frac{\gamma\beta^2\rho+r^2(1-\gamma^*)^2(1+\gamma\rho)-r\beta(1-\gamma)(1+2(1-\alpha)\gamma\rho)}{\gamma^*\beta\rho-r(1-\gamma^*)(1+\gamma^*\rho)}\right. \\
&+\frac{1}{(\gamma^*)^2\rho m^*\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}}\left[\gamma\beta^3\rho^2-(1-\gamma^*)^2r^3(1+\rho)(1+\gamma\rho)\right. \\
&+r^2\beta(\gamma(3-\gamma^*)(1-\gamma^*)\rho^2-(\gamma^*(\gamma^*-2\gamma+2)+\gamma-2)\rho+\gamma-1) \\
&\left.\left.-r\beta^2\rho((3-2\gamma^*)\gamma\rho-2\gamma+1)\right]\right\},
\end{aligned}$$

$$
\begin{aligned}
H_2 =& \frac{(1-\alpha)^{\frac{r+(r-\beta)\rho}{r}} e^{\frac{r+r\rho-\beta\rho+\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}}{2r}u}}{4\beta[\gamma\rho\beta - r(1-\gamma)(1+\gamma\rho)]} \\
&\times \left\{ \frac{\gamma\beta^2\rho + r^2(1-\gamma^*)^2(1+\gamma\rho) - r\beta(1-\gamma)(1+2(1-\alpha)\gamma\rho)}{\gamma^*\beta\rho - r(1-\gamma^*)(1+\gamma^*\rho)} \right. \\
&- \frac{1}{(\gamma^*)^2\rho m^*\sqrt{(r+r\rho-\beta\rho)^2+4r\beta\rho}}\left[\gamma\beta^3\rho^2 - (1-\gamma^*)^2 r^3(1+\rho)(1+\gamma\rho)\right. \\
&+ r^2\beta(\gamma(3-\gamma^*)(1-\gamma^*)\rho^2 - (\gamma^*(\gamma^*-2\gamma+2)+\gamma-2)\rho+\gamma-1) \\
&\left.\left. - r\beta^2\rho((3-2\gamma^*)\gamma\rho - 2\gamma+1)\right]\right\},
\end{aligned}
$$

$$
H_3 = -\frac{(1-\alpha)^{\frac{2r+(r-\beta)\rho}{r}} r e^{-\beta\tau+\frac{r+(r-\beta)\rho}{r}u}}{2\beta(r+\gamma^*\rho(r-\beta))}, \qquad H_4 = \frac{(1-\alpha)^{\frac{r+(r-\beta)\rho}{r}} e^{-r\tau+\frac{(r-\beta)\rho}{r}u}}{2(r-\gamma^*(r-r\rho+\beta\rho))},
$$

$$
\begin{aligned}
H_5 =& (1-\alpha)^{\frac{r+(r-\beta)\rho}{r}} e^{-m\tau+\frac{\gamma(r+r\rho-\beta\rho)-|2r-\gamma(r-r\rho+\beta\rho)|}{2\gamma r}u} \frac{\gamma^3\alpha r\rho}{4(\gamma\beta\rho - r(1-\gamma)(1+\gamma\rho))} \\
&\times \left[\frac{2r-\gamma(r-r\rho+\beta\rho)}{((\gamma+\gamma^*)r - \gamma\gamma^*(r-r\rho+\beta\rho))|2r-\gamma(r-r\rho+\beta\rho)|} - \frac{1}{\alpha(1-\gamma)^2 r + \gamma(2-\gamma)r + \gamma\gamma^*\rho(r-\beta)}\right],
\end{aligned}
$$

$$
\begin{aligned}
H_6 =& -(1-\alpha)^{\frac{r+(r-\beta)\rho}{r}} e^{-m\tau+\frac{\gamma(r+r\rho-\beta\rho)+|2r-\gamma(r-r\rho+\beta\rho)|}{2\gamma r}u} \frac{\gamma^3\alpha r\rho}{4(\gamma\beta\rho - r(1-\gamma)(1+\gamma\rho))} \\
&\times \left[\frac{2r-\gamma(r-r\rho+\beta\rho)}{((\gamma+\gamma^*)r - \gamma\gamma^*(r-r\rho+\beta\rho))|2r-\gamma(r-r\rho+\beta\rho)|} + \frac{1}{\alpha(1-\gamma)^2 r + \gamma(2-\gamma)r + \gamma\gamma^*\rho(r-\beta)}\right],
\end{aligned}
$$

$$
\begin{aligned}
H_7 =& -(1-\alpha)^{\frac{r+(r-\beta)\rho}{r}} e^{-m^*\tau+\frac{\gamma^*(r+r\rho-\beta\rho)-|2r-\gamma^*(r-r\rho+\beta\rho)|}{2\gamma^* r}u} \\
&\times \frac{(\gamma^*)^4\alpha r^2\rho(2r-\gamma^*(r-r\rho+\beta\rho))}{4\alpha(\gamma^*\beta\rho - r(1-\gamma^*)(1+\gamma^*\rho))(r-\gamma^*(r-r\rho+\beta\rho))(r+\gamma^*\rho(r-\beta))} \\
&\times \left[\frac{2r-\gamma^*(r-r\rho+\beta\rho)}{((\gamma+\gamma^*)r - \gamma\gamma^*(r-r\rho+\beta\rho))|2r-\gamma^*(r-r\rho+\beta\rho)|} + \frac{1}{\alpha(1-\gamma)^2 r + \gamma(2-\gamma)r + \gamma\gamma^*\rho(r-\beta)}\right],
\end{aligned}
$$

$$
\begin{aligned}
H_8 =& (1-\alpha)^{\frac{r+(r-\beta)\rho}{r}} e^{-m^*\tau+\frac{\gamma^*(r+r\rho-\beta\rho)-|2r-\gamma^*(r-r\rho+\beta\rho)|}{2\gamma^* r}u} \\
&\times \frac{(\gamma^*)^4\alpha r^2\rho(2r-\gamma^*(r-r\rho+\beta\rho))}{4\alpha(\gamma^*\beta\rho - r(1-\gamma^*)(1+\gamma^*\rho))(r-\gamma^*(r-r\rho+\beta\rho))(r+\gamma^*\rho(r-\beta))} \\
&\times \left[\frac{2r-\gamma^*(r-r\rho+\beta\rho)}{((\gamma+\gamma^*)r - \gamma\gamma^*(r-r\rho+\beta\rho))|2r-\gamma^*(r-r\rho+\beta\rho)|} + \frac{1}{\alpha(1-\gamma)^2 r + \gamma(2-\gamma)r + \gamma\gamma^*\rho(r-\beta)}\right].
\end{aligned}
$$

$\square$