Life-Changing Decisions and Portfolio Choice

Zoltán Rácz *
Stockholm School of Economics

May 30, 2023

Abstract

How do long-term saving targets affect optimal saving and portfolio choice decisions? I analyze a continuous time stochastic optimal control and stopping time model in which the agent may up- or downgrade her utility flow, income or liquidity constraint at a chosen time at the cost of a monetary payment. This general framework covers applications such as home purchase, voluntary retirement, bankruptcy or starting a private business. For general preferences an analytical solution is provided and it is shown that under the natural borrowing constraint, the presence of such options increases risk taking and savings, and this effect is stronger closer to the optimal switching point. The deviation from optimal policies of Merton's benchmark model is characterized as a function of the monetary value of switching states and the expected subjective discount factor at the time of phase transition.

1 Introduction

An extensive literature has examined household decisions regarding savings and portfolio choice over the life cycle, starting with Merton (1969) and Samuelson (1969). A fundamental question behind all models in this field is why households save at all, as this will also determine their willingness to hold risky assets. In most models the primary purpose of saving is smoothing consumption across time and states of nature. However, there exists substantial survey and anecdotal evidence that some part of saving activity is motivated by the aim of fulfilling long-term goals, such as buying a house, starting a private enterprise, early retirement or financing the education of one's offspring. It is an interesting question how in general

*I am grateful to Lars Ljungqvist and Paolo Sodini for their advice, feedback and support. I also thank Jess Benhabib, Jaroslav Borovička, Maximilian Huber, Jörgen Weibull and seminar participants at the Stockholm School of Economics and New York University for useful discussions and comments.

*zoltan.racz@phdstudent.hhs.se

such motives affect one's optimal saving and portfolio allocation decisions. One way of modeling this phenomenon is the following: assume that an individual is always in one of several discrete states, which affects her utility and determines her labor market status and ability to borrow. Each of these aspects has a bearing on saving decisions, as already shown in a sizable literature. However, once it is possible to switch discrete states, it is also the properties of potential alternative states that influence one's current decisions, even if the state switch does not take place immediately. For example, the potential motive of collecting money for the down-payment of a house obviously should be taken into account when trying to understand the saving decisions of renters. How exactly options to switch states shape current consumption-saving and portfolio choice decisions? How does the utility gain from potential switches affect expected lifetime utility? It is not a priori obvious that an interesting answer can be given to these questions in a general framework. Instead, it could be the case that optimal stopping time decisions affect a standard portfolio choice model through ways, which depend crucially on the exact context. In this case for each economic application a separate model would be required to understand the effects of the possibility of some state switch. The main contribution of this paper is showing that this is not so: There are general and intuitive consequences of introducing discrete decisions in an otherwise standard portfolio choice model. In particular, even the exact functional forms of the optimal policies' deviations from the no-switch benchmark solution are independent of the exact nature of alternatives states.

First, I show that utilizing the Principle of Optimality, instead of optimizing over an infinite sequence of stopping times representing all potential future state switches, it is sufficient to solve the optimal saving and portfolio choice problem combined with optimal stopping only until the first state switch. Following several papers in the literature of optimal portfolio choice models, this partial optimization problem is solved by duality methods, marginal utility being the dual continuous state variable. As opposed to wealth, this choice has a substantial advantage: considering the optimal policy, due to the presence of transaction costs, the path of wealth is not necessarily a continuous function. In contrast, state switches optimally take place such that marginal utilities before and after a phase transition are equalized, making the dual variable a continuous function of time in optimum. It is a standard finding that in an interior point, i.e. when no immediate switch is optimal, a Hamilton-Jacobi-Bellman equation needs to be satisfied, which for the dual problem simplifies to a linear differential equation in the infinite horizon case. This is in contrast to the primal problem, which typically leads to a non-linear equation, as the one investigated in Merton (1969). I provide a full solution of the differential equation involved and show how free parameters are pinned down by boundary conditions of the optimal stopping problem in a general setting. Optimal switches are characterized by critical thresholds: a phase transition is optimal if marginal utility falls below a lower limit (or wealth goes above a corresponding limit) or if it surpasses an upper threshold (corresponding to wealth falling under a critical value).

It turns out that apart from mathematical convenience, dual methods are also a source of economic intuition: The dual value function is additively separable into three terms, the first of which stands for expected lifetime utility when no switching option or borrowing constraint exists (or equivalently in our framework, the borrowing limit is the natural one). The other two terms represent the effect of switches taking place when reaching lower and upper boundaries, respectively. This decomposition provides an easy way to analyze welfare effects of potential state switches. Crucially, the first term is unaffected by the presence and features of optimal switch option. Furthermore, the other two terms are independent of the exact functional form of utility functions and instead they are qualitatively determined by the stochastic process followed by marginal utility in optimum. In this paper asset prices, which are the sole source of uncertainty, are assumed to follow a geometric Brownian motion process and the same property is inherited by the dual variable in optimum. In this case it turns out that the additional terms in the utility decomposition are power functions, where the exponents are determined by parameters of the asset price process and the impatience parameter. It is shown that apart from multiplicative factors, these functions represent the expected value of the subjective discount factor when the optimal switches happen. This is how the uncertainty regarding the arrival of phase transitions is taken into account into the value function. One general implication of this separation of the dual value function is the following: If the complete market benchmark model has an analytical solution, and the sources of uncertainty give rise to analytically tractable extra terms as geometric Brownian motion does in our case, then the introduction of switching options or a borrowing constraint does not affect the presence of an analytic solution for a particular problem. An interesting additional insight is that formally there is little difference between analyzing switches which are optimal for poor agents from the effects of borrowing limits. Intuitively, one can think of being borrowing constrained as a separate state, which decreases the value of not being constrained in a way that constrained agents are indifferent between the two states.

In addition to values, optimal policies can also be additively separated into one term driven by the benchmark model with no switches or borrowing limit, and two others representing these additions. In particular, frictionless net wealth, which is the sum of financial wealth and discounted labor income is shared between three parts. The first one finances future consumption in the current state with no borrowing limit assuming no switch in the future. The two other terms represent financing needs induced by switches or a borrowing limit, discounted based on the distance from the relevant boundaries. For the the case of state switches, this financing demand depends on the difference of human capital across the two involved discrete states, the transaction cost and finally the difference in wealth levels needed to maintain the expected optimal utility paths, given the starting value for marginal utility at the phase transition. On the other hand, when the boundary represents a borrowing limit, the corresponding wealth demand equals to the optimal amount of precautionary saving, relative to the benchmark model with the natural borrowing constraint. To sum up, the agent optimally dedicates some fraction of her net wealth to saving towards some goal or for precautionary reasons. The share of these components from total net wealth is a function of how large financing needs appear when reaching the boundaries and how soon the boundary is expected to be achieved.

A similar decomposition for optimal risky investment is also possible. It turns out that the ratio of risky investments dedicated to staying in the current state and the part of savings allocated for the same reason, equals the optimal risky share in the standard models by Merton (1969) and Samuelson (1969). If there is a borrowing constraint, the corresponding term for risky investment is negative and hence the total risky share decreases if the agent gets closer to the borrowing limit and thus the weight of the relevant term is increases. Considering the term belonging to a switching option relevant for low marginal utility levels (and hence high wealth), both the allocated risky investment and total savings for this target are positive. Furthermore, their ratio gives a higher risky share for this high wealth target component than for the staying component, when the utility function is in the constant relative risk aversion class. It is unclear, whether the latter conclusion holds more generally. Finally, for switches which are optimal for a high enough marginal utility (so low enough wealth), the sign of the effects is ambiguous. If the considered switch promises higher utility than staying in the benchmark model with a natural borrowing limit, then its presence increases consumption (through a negative financing demand) and increases risky investment. This makes the agent to reach the relevant threshold faster than otherwise. If however this is not true, and the switch is optimal only when compared to hitting a borrowing limit, its effects are qualitatively similar to that of limited borrowing, even though quantitatively smaller.

There already exists a sizable literature investigating the combination of stopping time problems with portfolio choice. In particular, this literature so far has considered frameworks where an initial and a final discrete state are given, and the optimal stopping time decision concerns an irreversible switch from the former to the latter. The final state typically corresponds to the some variant of the standard model by Merton. The objects of interest, apart from the stopping decision itself, are the optimal consumption and portfolio choice policies leading up to the switch. An early example is Jeanblanc et al. (2004) solving the problem of an indebted household with a borrowing limit, who optimally files for bankruptcy when her wealth drops below a critical value. It turns out that the presence of bankrupcty makes the household consume more and be less averse. However, the agent is still more risk averse than she would be without a borrowing constraint. In addition, several papers (Farhi and Panageas (2007), Dybvig and Liu (2010), Choi and Shim (2006), Choi et al. (2008), Barucci and Marazzina (2012)) analyze how the ability to optimally choose the time of voluntary retirement affects portfolio allocation and saving decisions. They find that retirement is optimal when wealth grows over a threshold and that the option of early retirement increases risk-taking and savings, while the effect on consumption depends on how utility from leisure is modeled. The closest article from this literature to this paper is Jeon and Koo (2021) who solves the early retirement problem for general utility functions and hence is able to distinguish which findings of earlier papers depend on particular assumptions on the utility function and which are generated directly by the optimal stopping decision. The current paper contributes to this literature in three dimensions. First, I examine a more general framework where states can freely differ along their felicity functions, incomes and borrowing limits, and state changes are subject to transaction costs. In particular, this model includes most above papers as special cases, and can explore whether different findings of this literature are particular to the economic problem they investigate. Second, instead of investigating only irreversible transitions and their effect on optimal policies before the change, I allow for arbitrary switches between states and it is determined endogenously, whether some of these states are final in the sense that it is optimal to stay in them forever. In particular, this means that optimization takes place over a sequence of stopping times instead of a single one. Finally, I provide an economically more intuitive characterization of the optimal policies and optimal value than the rest of the literature, relying on a decomposition of the dual value function.

Another related strand of literature is that of portfolio choice in the presence of durable goods, starting with the seminal paper of Grossman and Laroque (1990). In their framework utility is derived from holding a durable good, the owned quantity of which can only be adjusted subject to a transaction cost. Savings are kept either in risky or safe assets as in Merton (1969). Optimal durable choice is characterized by three numbers $\underline{z} < z^* < \bar{z}$. If the ratio of consumption (c) and wealth (a) is in $[\underline{z}, \bar{z}]$, the level of the durable good is not adjusted. Whenever c/a is outside of this range, a new durable level c' is set such that the $c'/a' = z^*$ optimal consumption ratio is reached. Compared to the no transaction cost case, the consumer behaves in a more risk averse manner just after purchasing a new durable, and in a less risk averse manner just before purchasing a new one. This framework of Grossman and Laroque (1990) has been extended by several papers which found similar results: by adding non-durable consumption with several utility functional forms as Beaulieu (1993), Damgaard et al. (2003) and Detemple and Giannikos (1996), considering a slightly more general structure for transaction costs as Cuoco and Liu (2000), or applying it to investigate the effects of costs related to housing transactions as Stokey et al. (1989). A quite different framework is that of Hindy and fu Huang (1993) who study non-additive preferences featuring local substitution (i.e. consumption levels at nearby points of time are substitutes), which similarly to the durable good literature results in step functions as optimal consumption policies. They find a constant risky share, but which is higher than the one in a comparable Merton model. In a sense these models are more general than that of this paper, as they allow for a continuum of 'discrete' states, characterized by the level of durable goods. However, this feature is achieved through assuming a perfect homogeneity in the model: In some sense therefore, all states are rescaled versions of one another and hence it hard to investigate the effects of the presence of phase transitions on utility and optimal policies in a general setting without also redefining the current state, such as the current paper does. Instead these models are most suited to understand the role of a certain class of transaction costs in economic decisions.

From the technical point of view, this paper relies most on Karatzas et al. (1986), Karatzas and Wang (2000) and He and Pagès (1993) who provide mathematically rigorous treatments of the original Merton problem and two variants with optimal stopping and borrowing constraints, respectively.

The rest of the paper is structured as follows: Section 2 describes the model and shows how under some condition the solution of the full sequential model can be build from the solutions of simpler subproblems. These subproblems are the subject of Section 3. First, the dual problem is defined following the literature and its solution is characterized in the current general framework. This is followed by a thorough discussion of the intuition related to the optimal value and policies. Section 4 returns to the sequential problem and fills in some gaps between the full problem and its subproblems, using the results from the previous Section. Next, an application involving a discrete decision between renting or owning housing is investigated in Section 5 to provide further intuition on the main results of the paper. Finally, Section 6 concludes.

2 The Sequential Problem

2.1 Setup

Time is continuous and the agent maximizes expected discounted utility over infinite time horizon. The agent is always in one of several possible states $i \in \mathcal{I}$ which differ in terms of their respective felicity function transforming flow consumption expenditure to contemporaneous utility, income and borrowing constraint. One can freely decide the of switches into other states, but switches take place subject to a transaction cost.

There exist two sources of income: the exogenous flow of labor income and capital income. Labor income is assumed to be risk-less and depend deterministically on the state $y_t = y_i \geq 0$ [1]. There are two means of investment: the risk-free investment provides a constant $r > 0$ rate of return over time, while the risky investment gives a stochastic return $(\mu + r)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$, where W_t is a one-dimensional Wiener-process

[1] This assumption could most likely be relaxed as in the case of the standard Merton model, either by an income process driven by W_t as in He and Pagès (1993) or with one introducing an independent source of uncertainty as in Koo (1998).

over a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. I consider the case where μ and σ are both strictly positive, so the risky assets provides a positive risk premium compensating for the extra risk. The one risky asset framework is a simplification, but as long as markets are complete, is without loss of generality. Since the only source of randomness is W_t, I will often make use of the filtration $\{\mathcal{F}_t\}$, which is the filtration generated by W_t.

The only continuous state variable is the amount of financial assets, denoted by a_t. At each instant the agent decides about the amount of risky investment ξ_t, and consumption expenditures c_t. Furthermore, the agent decides about the stopping times of state switches. Let τ_k denote the time of the kth switch from a state to a different one and i_k the state after the kth switch. Switching states entails a transaction cost, which is allowed to depend on the states between which the switch occurs and is denoted by $P(i, j)$. When being in state i, the amount of assets cannot be less than the state specific borrowing constraint $-b_i$. We are now in position to define the set of admissible controls as follows:

Definition 1. (Admissible Controls) The collection of stochastic processes c and ξ and random variables $\{\tau_k\}_{i=1}^{\infty}$ and $\{i_k\}_{i=1}^{\infty}$ constitute an admissible control if

(i) $c_t \geq 0$ for all t and both c_t and ξ_t are adapted to filtration $\{\mathcal{F}_t\}$

(ii) For all $t > 0$

$$\int_0^t c_t \,\mathrm{d}t + \int_0^t \xi_t^2 \,\mathrm{d}t < \infty$$

holds $\mathbb{P}$ almost surely.

(iii) For all $k > 0$, τ_k is a stopping time on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ and

$$\tau_k \leq \tau_{k+1}$$

holds $\mathbb{P}$ almost surely. τ_ks are allowed to attain value ∞.

(iv) For all $k > 0$, i_k is an $\mathcal{F}_{\tau_k}$-measurable random variable mapping Ω to

$$\{i \in \mathcal{I} \mid i \neq i_{k-1} \text{ and } a_{\tau_k -} - P(i_{k-1}, i) \geq -b_i\}$$

(v) There is an $\{\mathcal{F}_t\}$-adapted wealth process a_t with a_0 given, implicitly defined by the control variables, i.e. which satisfies the budget constraint

$$\int_0^t c_t \,\mathrm{d}t + a_t + \sum_{\tau_k \leq t} P(i_{k-1}, i_k) = a_0 + yt + r\int_0^t a_t \,\mathrm{d}t + \mu \int_0^t \xi_t \,\mathrm{d}t + \sigma \int_0^t \xi_t \,\mathrm{d}W_t \tag{1}$$

and the borrowing constraint

$$a_t \geq -b_{i_k} \quad \text{if } \tau_k \leq t < \tau_{k+1}$$

$\mathbb{P}$ almost surely for all t.

The set of admissible controls $(c, \xi, \{\tau_k\}, \{i_k\})$ will be denoted by $\mathfrak{c}$.

Condition (ii) is a technical requirement assuring that the stochastic integral in (1) exists. The measurability condition on i_k simply means that it should be known at the time of the kth switch, in which state the switch takes place. Furthermore, a switch can only happen to a state differing from the previous one and only if after paying the transaction cost the borrowing constraint is not violated in the new state. Allowing for the possibility of stopping times τ_k being infinite for all $k > k_0$ in practice corresponds to staying in the k_0th state for eternity being a feasible policy.

The agent maximizes her exponentially discounted lifetime utility

$$\mathcal{V}_i(a_0) = \sup_{(c_t, \xi_t, \{\tau_k\}, \{i_k\}) \in \mathfrak{c}} \mathbb{E}_0 \left[\sum_{k=0}^{\infty} \int_{\tau_k}^{\tau_{k+1}} e^{-\rho t} u\big(c_t, i_k\big) \, \mathrm{d}t \right] \tag{2}$$

over $\mathfrak{c}$, taking $i_0 = i$ and $a_0 \geq -b_i$ as given. It is understood that $\tau_0 = 0$. The felicity function $u(c, i)$ is increasing and strictly concave in its first argument for every $i \in \mathcal{I}$. It is also assumed that the felicity function is continuously differentiable in their first argument and the derivatives' range is $(0, \infty)$ for all i.

To make sure that any sort of utility maximization exercise makes sense, we also have to make sure that no arbitrage is possible in this setup. In standard models, the existence of a borrowing constraint is in itself sufficient to rule out arbitrage strategies Cox and fu Huang (1989). However, apart from stochastic returns, in our framework there is another potential arbitrage machine in the from of transaction costs. Indeed, so far we made no restriction over the set of $P(i, j)$s, even though it is easy to create an infinite source of wealth with an appropriate system of transaction costs such as $P(j, i) = -1$ and $P(i, j) = 0$. In this case repeatedly moving across states i and j becomes and infinite source wealth. A drastic way of avoiding this problem would be allowing only non-negative transaction costs. We take a more lenient approach, since we would like to model situations such as house transactions, where selling corresponds to a negative transaction cost: Intuitively, this should work with no danger of arbitrage if due to frictions, the selling price is always lower, and the buying price is higher than the 'true' value of the house. This observation motivates our next definition.

Definition 2. A system of transaction costs $\{P(i, j) \mid i, j \in \mathcal{I}\}$ is called regular if there exists a set of shadow values $\{s_i \in \mathbb{R} \mid i \in \mathcal{I}\}$ corresponding to each state such that

$$P(i, j) > s_j - s_i$$

for all $i, j \in \mathcal{I}$.

In the housing example, s_i would be the frictionless value of the house. This condition would imply

$$\sum_{\tau_k < t} P(i_{k-1}, i_k) > \max_{i \in \mathcal{I}} s_i - s_{i_0}$$

which puts a uniform lower bound on the total cost of switching states, being equivalent to an upper bound on the total income that can be generated merely by switching. Another consequence of Definition 2 is

$$\sum_{\tau_k < t} P(i_{k-1}, i_k) > \max_{i \in \mathcal{I}} s_i - s_{i_0} + \epsilon \cdot \max\{k \mid \tau_k < t\}$$

with $0 < \epsilon < \min_{i,j \in \mathcal{I}}\{P(i,j) - s_j + s_i\}$, which implies that with regular transaction costs all admissible sequence of stopping times diverges almost surely. Since the sequence of stopping times is increasing with probability one, this is equivalent to stopping times not having cluster points, which intuitively means that stopping times have to be spread over time. Indeed, if the stopping times converged to some finite t, the left hand side of the budget constraint (1) would diverge, which cannot be financed almost surely with any admissible policy. This point will be relevant for one of the assumptions of Theorem 1. In addition to regularity, there is another desirable, even though less crucial property of transaction costs which simplifies our discussion. Consider a situation in which the agent moves from state i to j and then instantly from state j to k. Such a move might be strictly preferred if $P(i,k) > P(i,j) + P(j,k)$, yet apart from the involved transaction costs state j has no bearing on the optimal value. In particular, by redefining $P(i,k) \equiv P(i,j) + P(j,k)$ it is possible to avoid this double transition without changing the optimal value and the optimal policies apart from erasing the transition through j. This is of some notational convenience in our later discussion. For the above reasons we make the following assumption.

Assumption 1. The set of transaction costs is regular and satisfies the triangle inequality, i.e.

$$P(i,k) \leq P(i,j) + P(j,k)$$

holds for every $i, j, k \in \mathcal{I}$.

2.2 Principle of Optimality

Since working with a sequence of stopping times is rather complicated, we take advantage of the Principle of Optimality. Informally speaking, the Principle of Optimality states that under some technical conditions, dynamic optimization problems can be separated into two subproblems by any time threshold. In particular, an optimal plan has the property that any initial segment of the plan is optimal given the continuation plan, and that the continuation plan is optimal for the problem started at the time threshold. It of course depends on the nature of the given problem what choice of the initial segment works best to characterize the solution. In discrete time dynamic programming the Bellman equation is derived by separating the first time period from the rest. In optimal control theory, the Hamilton-Jacobi-Bellman equation is obtained by considering an infinitesimally small initial segment. It turns out that in our current framework an insightful approach is to separate the problem into two at the first stopping time. Formally, for state i and a given corresponding continuation value function U_i we can define the

Definition 3. (First Stopping Problem)

$$V_i(a_0) = \sup_{((c_t)_0^\tau, (\xi_t)_0^\tau, \tau) \in \mathfrak{c}_i} \mathbb{E}_0 \left[\int_{t=0}^{\tau} e^{-\rho t} u\big(c_t, i\big)\, \mathrm{d}t + e^{-\rho\tau} U_i(a_\tau) \right]$$

where $a_0 \geq -b_i$ is given and $\mathfrak{c}_i$ denotes the set of admissible controls truncated at τ_1 when starting in state i.

The following theorem states that under some technical assumptions, once we obtain the solutions of a system of First Stopping Problems with conforming continuation value functions such that the suprema are attained, we can obtain a solution to the full sequential problem. This theorem is an analogue of Theorem 9.2 in Stokey et al. (1989) adapted to our setup [2].

Theorem 1. *Suppose that for all $i \in \mathcal{I}$, the First Stopping Problem is solved by V_i with continuation values*

$$U_i(a) = \max_{j \in \mathcal{I} \setminus i} V_j(a - P(i,j)). \tag{3}$$

Furthermore, assume that for all i, the supremum is attained, i.e. for all $a_0 \geq -b_i$ there exists an admissible control $((c_t^i(a_0))_0^{\tau^i}, (\xi_t^i(a_0))_0^{\tau^i}, \tau^i)$ such that

$$V_i(a_0) = \mathbb{E}_0 \left[\int_0^{\tau^i} e^{-\rho t} u\big(c_t^i(a_0), i\big)\, \mathrm{d}t + e^{-\rho\tau^i} U_i(a_{\tau^i}) \right]$$

Finally, assume that

$$\mathbb{E}_0 \left[e^{-\delta\tau_{n+1}} U_{i_n}(a_{\tau_{n+1}}) \right] \to 0 \tag{4}$$

holds as $n \to \infty$ for all admissible policies. Then

$$V_i(a_0) = \mathcal{V}_i(a_0)$$

for all i and an optimal policy of the sequential problem can be built from the given optimal policies of the First Stopping Problems as follows:

[2] It is probable that an analogue of their Theorem 9.4. also exists in this framework, i.e. under some conditions all solutions of the sequential problem solve the system of First Stopping Problems. This question is not pursued in the current version of this paper.

(i)

$$\tau_{k+1} = \tau_k + \tau^{i_k}$$

and

$$i_{k+1} = \arg\max_{j \in \mathcal{I} \setminus i_k} V_j(a_{\tau_{k+1}} - P(i_k, j))$$

for all $k \geq 0$.

(ii)

$$c_t(a_0) = c^{i_k}_{t-\tau_k}(a_{\tau_k})$$
$$\xi_t(a_0) = \xi^{i_k}_{t-\tau_k}(a_{\tau_k})$$

for all t, where k is the one such that $\tau_k \leq t < \tau_{k+1}$ holds.

Condition (3) means that for each i the relevant continuation value corresponds to the highest value that can be reached by switching, taking the transaction costs into account. Here it is understood that $V_j(a) = -\infty$ is $a < -b_j$. Equation (4) is a transversality condition. There is an important difference however from standard setups: As convergence is taken along a sequence of stopping times, it is not enough that the growth of function V is limited along a. It is also necessary that any admissible sequence of switching times is scattered enough. For example, if the switching times converge to a finite value, condition (4) cannot hold, except if V is zero at $a_{\lim \tau_n}$. This is exactly the kind of anomaly prevented by Assumption 1.

Proof. The proof broadly follows that of Theorem 9.2 in Stokey et al. (1989), but since the environments and technicalities are rather different, I include the full proof.

Let $i_0 \in \mathcal{I}$ and $a_0 \geq -b_{i_0}$ be arbitrary. Then

$$
\begin{aligned}
V_{i_0}(a_0) = &\sup_{(c_t)_{t=0}^{\tau},(\xi_t)_{t=0}^{\tau},\tau} \mathbb{E}_0\left[\int_0^{\tau} e^{-\rho t} u\big(c_t, i_0\big)\,\mathrm{d}t + e^{-\rho\tau} U_{i_0}(a_\tau)\right] \\
&\geq \mathbb{E}_0\left[\int_0^{\tau_1} e^{-\rho t} u\big(c_t, i_0\big)\,\mathrm{d}t + e^{-\rho\tau_1} U_{i_0}(a_{\tau_1})\right] \\
&\geq \mathbb{E}_0\left[\int_0^{\tau_1} e^{-\rho t} u\big(c_t, i_0\big)\,\mathrm{d}t + e^{-\rho\tau_1} V_{i_1}(Q(i_0, i_1, a_{\tau_1}))\right] \\
=&\mathbb{E}_0\left[\int_0^{\tau_1} e^{-\rho t} u\big(c_t, i_0\big)\,\mathrm{d}t + e^{-\rho\tau_1} \sup_{(c_t)_{t=0}^{\tau},(\xi_t)_{t=0}^{\tau},\tau} \mathbb{E}_0\left[\int_0^{\tau} e^{-\rho t} u\big(c_t, i_1\big)\,\mathrm{d}t\right.\right. \\
&\left.\left. + e^{-\rho\tau} U_{i_1}(a_\tau)\Big| a_0 = Q(i_0, i_1, a_{\tau_1 -})\right]\right] \\
=&\mathbb{E}_0\left[\int_0^{\tau_1} e^{-\rho t} u\big(c_t, i_0\big)\,\mathrm{d}t + e^{-\rho\tau_1} \sup_{(c_t)_{t=\tau_1}^{\tau_1+\tau},(\xi_t)_{t=\tau_1}^{\tau_1+\tau},\tau} \mathbb{E}_{\tau_1}\left[\int_{\tau_1}^{\tau_1+\tau} e^{-\rho(t-\tau_1)} u\big(c_t, i_1\big)\,\mathrm{d}t\right.\right. \\
&\left.\left. + e^{-\rho(\tau)} U_{i_1}(a_{\tau_1+\tau})\Big| a_{\tau_1} = Q(i_0, i_1, a_{\tau_1 -})\right]\right] \\
\geq&\mathbb{E}_0\left[\int_0^{\tau_1} e^{-\rho t} u\big(c_t, i_0\big)\,\mathrm{d}t + e^{-\rho\tau_1} \mathbb{E}_{\tau_1}\left[\int_{\tau_1}^{\tau_1+\tau} e^{-\rho(t-\tau_1)} u\big(c_t, i_1\big)\,\mathrm{d}t\right.\right. \\
&\left.\left. + e^{-\rho(\tau)} U_{i_1}(a_{\tau_1+\tau})\Big| a_{\tau_1} = Q(i_0, i_1, a_{\tau_1 -})\right]\right] \\
=&\mathbb{E}_0\left[\int_0^{\tau_1} e^{-\rho t} u\big(c_t, i_0\big)\,\mathrm{d}t\right] + \mathbb{E}_0\left[\mathbb{E}_{\tau_1}\left[\int_{\tau_1}^{\tau_2} e^{-\rho t} u\big(c_t, i_1\big)\,\mathrm{d}t\right.\right. \\
&\left.\left. + e^{-\rho\tau_2} U_{i_1}(a_{\tau_2})\Big| a_{\tau_1} = Q(i_0, i_1, a_{\tau_1 -})\right]\right] \\
=&\mathbb{E}_0\left[\int_0^{\tau_1} e^{-\rho t} u\big(c_t, i_0\big)\,\mathrm{d}t + \int_{\tau_1}^{\tau_2} e^{-\rho t} u\big(c_t, i_1\big)\,\mathrm{d}t + e^{-\rho\tau_2} U_{i_1}(a_{\tau_2})\right]
\end{aligned}
$$

where the second line follows from the definition of sup, the third one from condition (3) and the fourth one from the definition of V_{i_1}. The fifth line is by the strong Markov property of Ito processes, the sixth one is obtained again using the properties of sup. Finally, the seventh line is from defining $\tau_2 = \tau_1 + \tau$ and the last line is by the law of iterated expectations. By induction it can similarly be shown that

$$V_{i_0}(a_0) \geq \mathbb{E}_0\left[\sum_{k=0}^{n}\int_{\tau_k}^{\tau_{k+1}} e^{-\rho t}u\big(c_t, i_k\big)\,\mathrm{d}t\right] + e^{-\rho\tau_{n+1}}\mathbb{E}_0\left[U_{i_n}(a_{\tau_{n+1}})\right]$$

By letting $n \to \infty$ using (4) we get

$$V_{i_0}(a_0) \geq \mathcal{V}_{i_0}(a_0). \tag{5}$$

It is easy to verify that the chain of inequalities becomes a chain of equalities by substituting in the proposed policies implying that the inequality (5) is in fact an equality when the transversality condition (4) holds. □

Before turning our attention to characterize the solution of the First Stopping Problem, it is useful to point out that thanks to our assumption of constant income, in a sense the exact value of the borrowing limit is a superfluous parameter. In particular, given all the parameters that characterize a state, there is a way of transforming the rest of the parameters such that we can set the borrowing limit to 0 without affecting the optimal value and policies. The intuition is that for liquidity reasons it is only the difference of wealth and the borrowing limit that matters. Therefore once the difference in terms of interest income is taken into account, we can simply erase borrowing at the cost of increasing wealth and changing transaction costs by the corresponding amount. This results in a slight simplification of the discussions in the coming section.

Proposition 1. *The borrowing constraint b_i can be set equal to* 0 *without loss of generality. In particular, if we define*

$$\hat{P}(j,i) = P(j,i) - b_i \quad \textit{and} \quad \hat{P}(i,j) = P(i,j) + b_i \quad \forall j$$

then

$$V_i(a) \equiv V_{u_i,b_i,y_i}(a) = V_{u_i,0,y_i - rb_i}(a + b_i) \equiv \hat{V}_i(a + b_i)$$

Proof. Suppose c_t and ξ_t are admissible and in particular satisfy the budget constraint before the switching time from state i:

$$\mathrm{d}a_t = a_t r\,\mathrm{d}t + \xi_t\big(\mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t\big) - c_t\,\mathrm{d}t + y_i\,\mathrm{d}t$$

with $a_t \geq -b_i$. We can rewrite the law of motion using $\hat{a}_t = a_t + b_i \geq 0$ as follows

$$\begin{aligned}\mathrm{d}\hat{a}_t = \mathrm{d}a_t &= (\hat{a}_t - b)r\,\mathrm{d}t + \xi_t\big(\mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t\big) - c_t\,\mathrm{d}t + y_i\,\mathrm{d}t \\ &= \hat{a}_t r_f\,\mathrm{d}t + \xi_t\big(\mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t\big) - c_t\,\mathrm{d}t + (y_i - rb)\,\mathrm{d}t\end{aligned}$$

In addition, it is easy to check that the shifts in $\hat{P}$ compensate for redefining the wealth variable as $\hat{a}$. Therefore the same policies can be financed in the two setups, which implies identical values and optima. □

3 The First Stopping Problem

3.1 Duality

We first analyze the First Stopping Problem. Apart from the slightly more general setting, the beginning of this section follows very closely earlier papers such as Farhi and Panageas (2007), Dybvig and Liu (2010) or Jeanblanc et al. (2004) where portfolio choice and optimal stopping problems are combined in the context of early retirement or bankruptcy. On the theoretical side, the solution relies on duality methods developed in Karatzas and Wang (2000) and He and Pagès (1993) for problems featuring stopping times and borrowing constraints, respectively. The following discussion up to Proposition 2 is relatively standard and is reported for the sake of completeness and readability of the subsequent sections. Given an admissible policy (c, ξ, τ) define the implied value as

$$J(a_0, c, \xi, \tau) = \mathbb{E}\left[\int_0^\tau e^{-\rho t} u(c_t)\, \mathrm{d}t + e^{-\rho\tau} U(a_\tau)\right] \tag{6}$$

where U is a given strictly concave, increasing function. Throughout this section it is assumed that the borrowing limit b is 0. Then we have

$$V(a) = \sup_{c,\xi,\tau} J(a, c_t, \xi_t, \tau) \tag{7}$$

For any concave function $f : X \to \mathbb{R}$ we can define its convex conjugate [3] $\widetilde{f}$ by

$$\widetilde{f}(y) = \sup_{x\in X} \Big\{ f(x) - xy \Big\}. \tag{8}$$

The convex conjugate plays an important role in optimization theory and in particular the construction of dual problems as in this paper. For reference, some useful standard properties of the convex conjugate operator are collected in Remark 2 in the Appendix.

Denote by $\widetilde{u}$ and $\widetilde{U}$ the convex conjugate of u and U, respectively. It is perhaps best to clarify at this point that throughout the paper, for a generic function g, $\widetilde{g}$ will denote a dual function to g in some sense, which may or may not be defined directly as the convex conjugate above. This will be made clear in each instant.

As well-known (Cox and fu Huang (1989)), the current Black-Scholes environment is arbitrage-free and complete, and thus a unique equivalent martingale measure exists. The corresponding likelihood ratio process is $\exp\Big\{ -\int_0^t \kappa\, \mathrm{d}W_s - \theta t \Big\}$ where

$$\kappa = \frac{\mu}{\sigma} \quad \text{and} \quad \theta = \frac{\kappa^2}{2}$$

[3] Rigorously speaking, this is the Legendre-Fenchel transform of $-f(-x)$

are defined for the sake of obtaining more compact formulas. In this case the stochastic discount factor process H follows

$$H(t) = e^{-rt} \exp\Big\{ -\int_0^t \kappa \, \mathrm{d}W_s - \theta t \Big\}.$$

Now we can define

$$\widetilde{J}(\{X_t\}, \lambda, \tau) = \mathbb{E}\left[\int_0^\tau \Big[e^{-\rho t} \widetilde{u}(\lambda e^{\rho t} X_t H_t) + \lambda X_t H_t y \Big] \, \mathrm{d}t + e^{-\rho\tau} \widetilde{U}(\lambda e^{\rho\tau} X_\tau H_\tau) \right] \quad (9)$$

where $\{X_t\}$ is a decreasing positive adapted process with $X_0 = 1$, and $\lambda > 0$ is a positive real. Note that X is a diminishing term multiplying the standard stochastic discount factor representing the effect of the borrowing constraint. The following Lemma establishes a duality relation between J and $\widetilde{J}$.

Lemma 1. *If c, ξ and τ constitute an admissible policy, $\lambda > 0$ and X is a decreasing, adaptive, positive process with $X_0 = 1$, then*

$$J(a, c_t, \xi_t, \tau) \leq \widetilde{J}(\{X_t\}, \lambda, \tau) + \lambda a \quad (10)$$

and we have equality if and only if

$$\lambda e^{\rho\tau} X_\tau H_\tau = \arg\max_y \, \{U(a_\tau) - a_\tau y\} \quad (11)$$

$$u'(c_t) = \lambda e^{\rho t} X_t H_t \quad \forall 0 \leq t < \tau \quad (12)$$

$$a_0 = \mathbb{E}\left[\int_0^\tau H_s(c_s - y) \, \mathrm{d}s + H_\tau a_\tau \right] \quad (13)$$

$$0 = \mathbb{E}\left[\int_0^\tau a_t \, \mathrm{d}X_t \right] \quad (14)$$

The conditions for equality are intuitive: (11) and (12) are first order conditions, stating that instantanious marginal utilities must equal the modified stochastic discount factor multiplied with the Lagrange-multiplier of the budget constraint. So far we have not made any assumptions on the differentiability properties of U, but note that if U is continuously differentiable, then (11) is equivalent to $U'(a_\tau) = \lambda e^{\rho\tau} X_\tau H_\tau$. Equation (13) states that the consolidated budget constraint has to be exhausted with equality, while (14) means that X_t has a locally constant path when $a_t > 0$. Therefore future discount factors are depressed by X along a given realization of W to the extent that wealth has been constrained along this path obstructing the transformation of wealth over time. Another source of intuition for the latter condition is given by a deterministic finite horizon example in Section

4 of He and Pagès (1993): in that case X_t proves to be identical to the time integral of the borrowing constraints' Lagrange multipliers from time t up to final time T. Naturally, the Lagrange multiplier at time t is zero when the condition is non-binding at that time, leading to a locally constant X_t.

Define the dual value function as

$$\widetilde{V}(\lambda) = \sup_{\tau} \inf_{\{X_t\}} \widetilde{J}(\{X_t\}, \lambda, \tau) \tag{15}$$

where X and λ are as in the definition of $\widetilde{J}$. Now by Lemma 1 and the properties of inf and sup we have

$$V(a) = \sup_{c,\xi,\tau} J(a, c_t, \xi_t, \tau) \le \sup_{\tau} \inf_{\{X_t\},\lambda} \left\{ \widetilde{J}(\{X_t\}, \lambda, \tau) + \lambda a \right\} \le \inf_{\lambda} \left\{ \widetilde{V}(\lambda) + \lambda a \right\} \tag{16}$$

We can claim that our original problem is successfully solved if we can determine the value of the last term, we establish that both inequalities in (16) are in fact equalities and if we clarify how to produce the optimal policies from the dual value function. Our strategy to do so is through a verification theorem. We can give conditions on $\widetilde{V}$, τ and X such that for all λ we have $\widetilde{J}(\{X_t\}, \lambda, \tau) = \widetilde{V}(\lambda)$. Furthermore, for all a using $\widetilde{V}$ and choosing an appropriate λ, we can build an admissible policy satisfying all conditions of Lemma 1. Let us now turn our attention on the last term.

Take λ and process X_t as given and define

$$Z_t^X = \lambda e^{\rho t} X_t H_t \tag{17}$$

where the X superscript emphasizes the dependence of Z^X on the X process. Now the dual problem can be formulated as

$$\widetilde{V}(\lambda) = \sup_{\tau} \inf_{\{X_t\}} \mathbb{E}\left[\int_0^{\tau} e^{-\rho t} \left(\widetilde{u}(Z_t^X) + y Z_t^X \right) \mathrm{d}t + e^{-\rho\tau} \widetilde{U}(Z_\tau^X) \right] \tag{18}$$

subject to

$$\begin{aligned} \frac{\mathrm{d}Z_t^X}{Z_t^X} &= \mathrm{d}X_t + (\rho - r)\,\mathrm{d}t - \kappa\,\mathrm{d}W_t \\ Z_0^X &= \lambda \\ X_0 &= 1 \\ &X \text{ is decreasing, adapted and positive.} \end{aligned}$$

This is an optimal control problem, but unlike in the case of the primal problem, the control process is extremely well-behaved. Indeed, we already know from Lemma 1 that in the case of an optimal policy we can expect X to be constant everywhere except when the borrowing constraint binds. In addition, the Hamilton-Jacobi-Bellman equation characterizing the optimal value function is linear and hence easy to solve, in contrast to that of the primal problem. Before proceeding, note that $\widehat{V}$ clearly has to be a decreasing function, since increasing λ weakly increases the set of possible values. This is because X can always jump downward at time 0 to compensate for an increase in λ, but not the other way round. Furthermore, $\widehat{V}$ can be shown to be convex, since all functions on the right hand side of (18) are convex.

Before starting to analyze the properties of the solution of the dual problem, it is important to discuss the issue of the solution's existence. Since for the case of dual problems in portfolio theory existence results are standard, and the exact conditions are somewhat orthogonal to the main focus of this paper, the question of existence is not treated explicitly here. Instead, it is implicitly assumed that the felicity function is well-behaved enough to make the dual problem well-defined. The interested reader is directed to He and Pagès (1993) and Farhi and Panageas (2007) who both present existence results relying on integrability conditions on the utility functions u and U, and whose results can be straightforwardly merged to cover the current setup.[4] It should be noted however, that the conditions in both papers allow for a constant risk aversion felicity function, if it satisfies the same parametric conditions which are demanded in this paper as well later on. The solution is characterized by the following proposition, which on the one hand generalizes Theorem 4 in He and Pagès (1993) with controllable stopping times, but on the other hand I specialize their general income and price processes to our simpler setup.

Proposition 2. *Consider the infinitesimal generator of Z_t*

$$\mathcal{A}f(z) = -\rho f(z) + (\rho - r)zf'(z) + \theta z^2 f''(z)$$

Suppose that we can find a convex function $\widehat{V}(Z) : (0, \infty) \to \mathbb{R}$ such that

(i) $\widehat{V} \in C(0, \infty) \cap C^1(0, \infty) \cap C^2\big((0, \infty) \setminus \partial D\big)$ *and the second order derivatives are locally bounded near ∂D, where D is defined as*

$$D := \{Z \in (0, \infty) \mid \widehat{V}(Z) > \widetilde{U}(Z)\}; \tag{19}$$

(ii) $\widehat{V}(Z) \geq \widetilde{U}(Z) \quad \forall Z \in (0, \infty)$ *and*

(iii)

$$\max\{\mathcal{A}\widehat{V}(Z) + \widetilde{u}(Z) + yZ, \widehat{V}'(Z)\} = 0$$

for all $Z \in (0, \infty)$.

[4] It is possible that there is a direct connection between these integrability conditions assuring the existence of the solution of the dual problem and Assumption 3 in this paper. This intriguing question will be investigated in further research.

Assume furthermore, that there exists a decreasing, adapted, positive Ito process with $X_0 = 1$ such that it is almost surely continuous for $t > 0$ and given an arbitrary $\lambda > 0$ the following conditions hold:

(iv) $\mathrm{d}X_t \widehat{V}'(Z_t^X) = 0$ *for all* t.

(v)
$$\mathcal{A}\widehat{V}(Z_t^X) + \widetilde{u}(Z_t^X) + yZ_t^X = 0 \qquad \forall t \text{ with } Z_t^X \in D.$$

(vi) The family $\{\widehat{V}(Z_\tau^X) \mid \tau \le \tau_D\}$ *is uniformly integrable, where* $\tau_D = \inf\{t > 0 \mid Z_t^X \notin D\}$.

Then
$$\widehat{V}(\lambda) = \widetilde{J}(\{X_t, \lambda, \tau\}) = \widetilde{V}(\lambda)$$
is the solution of the dual problem and $\tau = \tau_D$ *is an optimal stopping time. Furthermore, for* $a_0 \ge 0$ *given, and choosing* λ *such that*

$$\widehat{V}'(\lambda) = -a_0 \tag{20}$$

holds, the optimal policies of the First Stopping Problem and the corresponding wealth process are determined as follows:

$$u'(c_t) = Z_t^X \tag{21}$$
$$\xi_t = \frac{\mu}{\sigma^2} Z_t^X \widehat{V}''(Z_t^X) \tag{22}$$
$$a_t = -\widehat{V}'(Z_t^X) \tag{23}$$

and the optimal stopping times coincide.

Condition (i) implicitly contains the so-called smooth-pasting condition: it states that the value function is continuously differentiable even on the optimal stopping boundary. This can be interpreted as a first-order condition with respect to the stopping time, i.e. the marginal utility in the case of staying and stopping has to coincide at the optimal switch. As $\widetilde{V}'$ will be equal to wealth, (iv) is a version of the condition that X is constant when the borrowing constraint does not bind. The Hamilton-Jacobi-Bellman equation in (v) corresponds to

$$0 = -\rho\widetilde{V}(Z) + \widetilde{V}'(Z)Z\Big(\rho - r\Big) + \widetilde{V}''(Z)Z^2\theta + \widetilde{u}(Z) + Zy, \tag{24}$$

which has to hold for all $Z \in D$ unless $\widetilde{V}'(Z) = 0$. There is a parallel between the economic intuition and mathematical role of X_t. One the one hand, X adjusts the stochastic discount factor process such that agents on the borrowing constraint optimally do not borrow. This is mirrored by the fact that X controls Z_t^X so that it stays in the region where equation (24) applies. This might be done with a jump at $t = 0$ and continuously afterwards. In particular, Z_t^X is kept away from the interior of the region where $-\widehat{V}'(Z) = a_t = 0$. This is feasible by a decreasing process, as since $\widehat{V}$ is decreasing and convex, it can only be constant over an interval of the form $(\widehat{Z}, \infty)$.

3.2 Characterizing the Solution of the Dual Problem

After seeing how to produce solutions of the original First Stopping Problem from the solution of the dual problem, it is worth exploring further the differential equation which has to solved by the dual function on D according to (v). In particular, first we provide the general solution to (24) and then show how the boundary conditions implicit in (i) and the integrability condition (vi) pin down the free parameters. Finally, we investigate how the final functional form depends on whether or not optimal stops and the borrowing constraint are relevant for a particular problem and provide some economic intuition for the involved parameters.

Before getting started, we need to discuss the topological properties of set D, the domain of the Hamilton-Jacobi-Bellman equation. (19) implies that D is an open subset of the real line, and as such, it can be characterized as a countable union of disjoint open intervals. Therefore the solution of (24) should be provided by solving the differential equation separately for each subintervals, according to the boundary conditions specific to the given subinterval, and then building the full solution by merging the subsolutions and $\widetilde{U}$ appropriately over the full domain. To avoid these complications in this paper I make the following assumption:

Assumption 2. The First Stopping Problem is such that D is an interval.

One motivation for this assumption is that in most obvious economic applications any state would be optimal over an interval of wealth levels. Since consumption is a monotonic function of wealth, this implies an optimal interval in the marginal utility (Z) space. In addition, extending the analysis below is straightforward to the case when D is a collection of intervals, however little additional insight could be gained at a significant cost of clarity.

Therefore we have to solve (24), which is an inhomogeneous second order linear differential equation in $\widetilde{V}$, over an interval D of the real line. This means that given a particular solution, any other solution of (24) on D can be written as the sum of a particular solution and some solution of the homogeneous equation

$$-\rho\widetilde{V}(Z) + \widetilde{V}'(Z)Z\Big(\rho - r\Big) + \widetilde{V}''(Z)Z^2\theta = 0 \tag{25}$$

which is a Cauchy-Euler equation, for which the general solution is known as:

$$\widetilde{V}^{hom}(Z) = BZ^{\varphi_+} + CZ^{\varphi_-} \tag{26}$$

where B and C are arbitrary reals and φ_+, φ_- are the two solutions of the quadratic equation

$$\theta\varphi^2 + (\rho - r - \theta)\varphi - \rho = 0 \tag{27}$$

so

$$\varphi_+, \varphi_- = \frac{-(\rho - r - \theta) \pm \sqrt{(\rho - r - \theta)^2 + 4\rho\theta}}{2\theta} \tag{28}$$

which implies that $\varphi_- < 0$ and $\varphi_+ > 0$. In fact, we even have $\varphi_+ > 1$, since by substituting $\varphi = 1$ in equation (27) we get $-r < 0$, implying that 1 lies in between the two roots. To proceed and put condition (vi) into use, we need to make sure that there exists a sufficiently integrable particular solution.

Assumption 3. Equation (24) has a convex particular solution $\widetilde{V}_p$ over $(0, \infty)$ such that the set of random variables $\{e^{-\delta\tau}\widetilde{V}_p(Z_\tau) \mid \tau < \infty \text{ is a stopping time}\}$ is uniformly integrable, where Z_t denotes the process controlled by a constant $X_t = 1$ process, i.e.

$$\frac{\mathrm{d}Z_t}{Z_t} = (\rho - r)\,\mathrm{d}t - \kappa\,\mathrm{d}W_t.$$

In addition, if $y = 0$, this particular solution is a decreasing function in Z.

Of course, it is of interest whether we can hope Assumption 3 to hold for any commonly used utility functions. A positive answer regarding constant relative risk aversion utiliy functions is given in the following proposition:

Proposition 3. *Assume that*

$$u(c) = h^\gamma \frac{c^{1-\gamma}}{1-\gamma} + n$$

with risk aversion parameter $\gamma > 0$ satisfying

$$\rho > (1-\gamma)\Big[r + \frac{\theta}{\gamma}\Big] \tag{29}$$

and $h > 0$ and $n \in \mathbb{R}$ are scale parameters. Then its convex conjugate is

$$\widetilde{u}(Z) = \frac{\gamma}{1-\gamma} h Z^{1-1/\gamma} + n \tag{30}$$

and

$$\widetilde{V}_p(Z) = \frac{\gamma}{1-\gamma} h A Z^{1-1/\gamma} + \frac{n}{\rho} + Z\frac{y}{r} \tag{31}$$

with

$$A = \frac{\gamma}{\rho - (1-\gamma)\big[r + \frac{\theta}{\gamma}\big]}$$

is a particular solution to (24) satisfying the conditions of Assumption 3.

To prove this proposition and then to apply the integrability conditions to pin down free parameters B and C, we will rely on the following lemma characterizing the uniform integrability properties of discounted power functions of Z_t.

Lemma 2. *Let α be an arbitrary real number and consider the process $e^{-\delta t} Z_t^{\alpha}$. Then*

(a) $e^{-\delta t} Z_t^{\alpha}$ is a martingale if and only if $\alpha = \varphi_-$ or $\alpha = \varphi_+$ and is a supermartingale iff $\varphi_- \leq \alpha \leq \varphi_+$.

(b) If $\varphi_- < \alpha < \varphi_+$, then $e^{-\delta t} Z_t^{\alpha}$ converges to 0 in probability as $t \to \infty$. In addition, the set $\{e^{-\delta \tau} Z_\tau^{\alpha} \mid \tau < \infty$ is a stopping time$\}$ is uniformly integrable .

(c) The set $\{e^{-\delta \tau} Z_\tau^{\varphi_+} \mid \tau \leq \tau_{\overline{Z}}$ is a stopping time$\}$ is uniformly integrable with $\tau_{\overline{Z}} = \inf\{t \geq 0 \mid Z_t \geq \overline{Z}\}$ where $0 < \overline{Z} < \infty$.

(d) The set $\{e^{-\delta \tau} Z_\tau^{\varphi_-} \mid \tau \leq \tau_{\underline{Z}}$ is a stopping time$\}$ is uniformly integrable with $\tau_{\underline{Z}} = \inf\{t \geq 0 \mid Z_t \leq \underline{Z}\}$ where $0 < \underline{Z} < \infty$.

(e) The sets $\{e^{-\delta \tau} Z_\tau^{\varphi_+} \mid \tau < \infty$ is a stopping time$\}$ and $\{e^{-\delta \tau} Z_\tau^{\varphi_-} \mid \tau < \infty$ is a stopping time$\}$ are not uniformly integrable.

Proof. For any α, Z_t^{α} follows a geometric Brownian motion and

$$Z_t^{\alpha} = Z_0^{\alpha} \exp\{\alpha(\rho - r - \theta)t + \alpha\kappa W_t\}$$

which implies that

$$\mathbb{E}[e^{-\delta t} Z_t^{\alpha}] = Z_0^{\alpha} \exp\{[-\rho + \alpha(\rho - r - \theta) + \alpha^2\theta]t\} \tag{32}$$

(a) Follows directly from (32) and the definition of $\varphi_\pm$.

(b)

$$\mathbb{E}[|e^{-\delta t} Z_t^{\alpha}|] = \mathbb{E}[e^{-\delta t} Z_t^{\alpha}] \to 0$$

when $\varphi_- < \alpha < \varphi_+$, since in this case $-\rho + \alpha(\rho - r - \theta) + \alpha^2\theta < 0$. This means that $\{e^{-\delta t} Z_t^{\alpha}\}$ converges to 0 in L^1 and hence also in probability. By Vitali's convergence theorem this is equivalent to set $\{e^{-\delta t} Z_t^{\alpha} \mid t \geq 0\}$ being uniformly integrable, which property can then be extended to all stopping times, as $e^{-\delta \tau} Z_\tau^{\alpha} = \mathbb{E}[e^{-\delta t} Z_t^{\alpha} \mid \mathcal{F}_\tau]$ and taking conditional expectations with respect to arbitrary sub-σ-algebras preserves uniform integrability.

(c) Let $\tau \leq \tau_{\overline{Z}}$ be arbitrary. Then $Z_\tau \leq \overline{Z}$ and so

$$0 \leq e^{-\delta \tau} Z_\tau^{\varphi_+} \leq e^{-\delta \tau} \overline{Z}^{\varphi_+} \leq \overline{Z}^{\varphi_+}$$

This means that $\{e^{-\delta \tau} Z_\tau^{\alpha} \mid \tau$ is a stopping time$\}$ is a set of uniformly bounded random variables and hence is trivially uniformly integrable.

(d) As (c). Note that $\varphi_- < 0$ so $Z_\tau \geq \underline{Z}$ implies $Z_\tau^{\varphi_-} \leq \underline{Z}^{\varphi_-}$.

(e) Let $K > 0$ be arbitrary and define

$$\tau = \inf\{t \geq 0 \mid e^{-\delta t} Z_t^{\varphi_+} \geq K\}$$

Then

$$\begin{aligned}\mathbb{E}[e^{-\delta\tau} Z_\tau^{\varphi_+} \mathbb{1}_{e^{-\delta\tau} Z_\tau^{\varphi_+} \geq K}] &= \mathbb{E}[e^{-\delta\tau} Z_\tau^{\varphi_+}] - \mathbb{E}[e^{-\delta\tau} Z_\tau^{\varphi_+} \mathbb{1}_{e^{-\delta\tau} Z_\tau^{\varphi_+} < K}] \\ &= Z_0^{\varphi_+} - 0 = Z_0^{\varphi_+}\end{aligned}$$

The first term is obtained by the Optional Sampling Theorem for martingales and the second term is 0 as the event $\mathbb{1}_{e^{-\delta\tau} Z_\tau^{\varphi_+} < K}$ is uniformly zero, as every geometric Brownian motion is a continuous process. This means that for any $\epsilon < Z_0^{\varphi_+}$, for every $K > 0$ we can find a stopping time τ such that $\mathbb{E}[e^{-\delta\tau} Z_\tau^{\varphi_+} \mathbb{1}_{e^{-\delta\tau} Z_\tau^{\varphi_+} \geq K}] < \epsilon$ does not hold. This implies that the set $\{e^{-\delta\tau} Z_\tau^{\varphi_+} \mid \tau \text{ is a stopping time}\}$ is not uniformly integrable. The proof for φ_- can be obtained simply by switching all the φs.

□

It is worth noting that point (e) of the above Lemma implies that no non-zero solution of the homogeneous equation satisfies the integrability requirement for general finite stopping times. This means that if there exists a sufficiently integrable particular solution, as posited by Assumption 3, it has to be unique. Now we can turn to proving Proposition 3.

Proof of Proposition 3. The convex dual in (30) is obtained by point (iii) in Remark 2. Simple substitution shows that $\widetilde{V}_p$ in (31) solves the HJB-equation. Finally, $\widetilde{V}_p$ satisfies the integrability requirement being the sum of three functions satisfying the same condition by (b) in Lemma 2. We already saw that $\varphi_- < 1 < \varphi_+$ is true and $\varphi_- < 1 - \frac{1}{\gamma} < \varphi_+$ can also be shown by substituting $1 - \frac{1}{\gamma}$ into (27) under (29). □

Notice that the particular solution in the CRRA case contains a separable term related to y. It is easy to show that this is true in general.

Remark 1. *Let $\widetilde{V}_p$ be a particular solution of (24) satisfying the integrability condition in Assumption 3. Then for all $Z \in (0, \infty)$*

$$\widetilde{V}_p(Z) = \widetilde{V}_{p_0}(Z) + Z\frac{y}{r}$$

where $\widetilde{V}_{p_0}$ is the integrable particular solution of an otherwise identical equation with $y = 0$.

We are finally in position to show how functional forms and parameters for the dual value function are pinned down by some qualitative properties of the solution.

Theorem 2. *Assume that $\widetilde{V}$ is a solution of the dual problem with τ, X_t and $D = (\underline{Z}, \overline{Z})$ are as in Proposition 2, where $\underline{Z} = 0$ and $\overline{Z} = \infty$ are allowed.*

1. *If $\tau = \infty$ (i.e. $D = (0, \infty)$) and $y = 0$, then X_t must be constant 1 and $B = C = 0$, so*

$$\widetilde{V}(Z) = \widetilde{V}_p(Z) = \widetilde{V}_{p_0}(Z) \qquad \forall Z \in (0, \infty)$$

and

$$V(a_0) = \widetilde{V}_{p_0}(\lambda) + a_0\lambda \qquad \forall a_0 \in (0, \infty)$$

with $\widetilde{V}'_{p,0}(\lambda) = -a_0$.

2. *If $\tau = \infty$ and $y > 0$, then there exists a $\widehat{Z}$ such that*

$$\widetilde{V}(Z) = \begin{cases} \widetilde{V}_{p_0}(Z) + BZ^{\varphi_+} + Z\frac{y}{r} & \text{if } Z \leq \widehat{Z} \\ \widetilde{V}_{p_0}(\widehat{Z}) + B\widehat{Z}^{\varphi_+} + \widehat{Z}\frac{y}{r} & \text{if } Z \geq \widehat{Z} \end{cases}$$

and

$$X_t = \min\left\{1, \inf_{0 \leq s \leq t} \frac{\widehat{Z}}{Z_s}\right\} \qquad \forall t \geq 0. \tag{33}$$

Parameters B and $\widetilde{Z}$ are pinned down such that $\widetilde{V}(Z)$ is twice continuously differentiable at $\widehat{Z}$. Finally,

$$V(a_0) = \widetilde{V}_{p_0}(\lambda) + B\lambda^{\varphi_+} + \left(a_0 + \frac{y}{r}\right)\lambda \qquad \forall a_0 \geq 0$$

with $\widetilde{V}'_{p,0}(\lambda) + \varphi_+ B\lambda^{\varphi_+ - 1} = -\left(a_0 + \frac{y}{r}\right)$.

3. *If $D = (0, \overline{Z})$ with $\overline{Z} < \infty$, then*

$$\widetilde{V}(Z) = \begin{cases} \widetilde{V}_{p_0}(Z) + BZ^{\varphi_+} + Z\frac{y}{r} & \text{if } Z \leq \overline{Z} \\ \widetilde{U}(Z) & \text{if } Z \geq \overline{Z} \end{cases}$$

and X_t is constant 1. Parameters B and $\overline{Z}$ are such that $\widetilde{V}(Z)$ is once continuously differentiable at $\overline{Z}$. Finally,

$$V(a_0) = \widetilde{V}_{p_0}(\lambda) + B\lambda^{\varphi_+} + \left(a_0 + \frac{y}{r}\right)\lambda \qquad \forall a_0 \geq \underline{a}$$

with $\widetilde{V}'_{p,0}(\lambda) + \varphi_+ B\lambda^{\varphi_+ - 1} = -\left(a_0 + \frac{y}{r}\right)$ and

$$V(a_0) = U(a_0) \qquad \forall a_0 \leq \underline{a}$$

where

$$\underline{a} = -\widetilde{V}'_{p,0}(\overline{Z}) - \varphi_+ B\overline{Z}^{\varphi_+ - 1} - \frac{y}{r}.$$

4. *If $D = (\underline{Z}, \infty)$ with $\underline{Z} > 0$ and $y = 0$, then*

$$\widetilde{V}(Z) = \begin{cases} \widetilde{U}(Z) & \text{if } Z \leq \underline{Z} \\ \widetilde{V}_{p_0}(Z) + CZ^{\varphi_-} & \text{if } Z \geq \underline{Z} \end{cases}$$

and X_t is constant 1. Parameters C and $\underline{Z}$ are such that $\widetilde{V}(Z)$ is once continuously differentiable at $\underline{Z}$. Finally,

$$V(a_0) = \widetilde{V}_{p_0}(\lambda) + C\lambda^{\varphi_-} + a_0\lambda \qquad \text{if } \overline{a} \geq a_0 > 0$$

with $\widetilde{V}'_{p,0}(\lambda) + \varphi_- C\lambda^{\varphi_- - 1} = -a_0$ and

$$V(a_0) = U(a_0) \qquad \text{if } a_0 \geq \overline{a}$$

where

$$\overline{a} = -\widetilde{V}'_{p,0}(\underline{Z}) - \varphi_- C\underline{Z}^{\varphi_- - 1}.$$

5. *If $D = (\underline{Z}, \infty)$ with $\underline{Z} > 0$ and $y > 0$, then there exists a $\widehat{Z}$ such that*

$$\widetilde{V}(Z) = \begin{cases} \widetilde{U}(Z) & \text{if } Z \leq \underline{Z} \\ \widetilde{V}_{p_0}(Z) + BZ^{\varphi_+} + CZ^{\varphi_-} + Z\frac{y}{r} & \text{if } \underline{Z} \leq Z \leq \widehat{Z} \\ \widetilde{V}_{p_0}(\widehat{Z}) + B\widehat{Z}^{\varphi_+} + C\widehat{Z}^{\varphi_-} + \widehat{Z}\frac{y}{r} & \text{if } Z \geq \widehat{Z} \end{cases}$$

and

$$X_t = \min\left\{1, \inf_{0 \leq s \leq t} \frac{\widehat{Z}}{Z_s}\right\} \qquad \forall t \geq 0.$$

Parameters B, C, $\underline{Z}$ and $\widetilde{Z}$ are pinned down such that $\widetilde{V}(Z)$ is once continuously differentiable at $\underline{Z}$ and twice continuously differentiable at $\widehat{Z}$. Finally,

$$V(a_0) = \widetilde{V}_{p_0}(\lambda) + B\lambda^{\varphi_+} + C\lambda^{\varphi_-} + \left(a_0 + \frac{y}{r}\right)\lambda \qquad \text{if } \overline{a} \geq a_0 \geq 0$$

with $\widetilde{V}'_{p,0}(\lambda) + \varphi_+ B\lambda^{\varphi_+ - 1} + \varphi_- C\lambda^{\varphi_- - 1} = -\left(a_0 + \frac{y}{r}\right)$ and

$$V(a_0) = U(a_0) \qquad \text{if } a_0 \geq \overline{a}$$

where

$$\overline{a} = -\widetilde{V}'_{p,0}(\underline{Z}) - \varphi_+ B\underline{Z}^{\varphi_+ - 1} - \varphi_- C\underline{Z}^{\varphi_- - 1} - \frac{y}{r}$$

6. *If $D = (\underline{Z}, \overline{Z})$ with $\underline{Z} > 0$ and $\overline{Z} < \infty$,then*

$$\widetilde{V}(Z) = \begin{cases} \widetilde{U}(Z) & \text{if } Z \leq \underline{Z} \\ \widetilde{V}_{p_0}(Z) + BZ^{\varphi_+} + CZ^{\varphi_-} + Z\frac{y}{r} & \text{if } \underline{Z} \leq Z \leq \overline{Z} \\ \widetilde{U}(Z) & \text{if } Z \geq \overline{Z} \end{cases}$$

and X_t is constant 1*. Parameters B, C, $\underline{Z}$ and $\overline{Z}$ are pinned down such that $\widetilde{V}(Z)$ is once continuously differentiable at $\underline{Z}$ and $\overline{Z}$. Finally,*

$$V(a_0) = \widetilde{V}_{p_0}(\lambda) + B\lambda^{\varphi_+} + C\lambda^{\varphi_-} + \Big(a_0 + \frac{y}{r}\Big)\lambda \qquad \text{if } \underline{a} \le a_0 \le \overline{a}$$

with $\widetilde{V}'_{p,0}(\lambda) + \varphi_+ B\lambda^{\varphi_+-1} + \varphi_- C\lambda^{\varphi_--1} = -\Big(a_0 + \frac{y}{r}\Big)$, *and*

$$V(a_0) = U(a_0) \qquad \text{if } a_0 \le \underline{a} \text{ or } a_0 \ge \overline{a}.$$

where

$$\overline{a} = -\widetilde{V}'_{p,0}(\underline{Z}) - \varphi_+ B\underline{Z}^{\varphi_+-1} - \varphi_- C\underline{Z}^{\varphi_--1} - \frac{y}{r}$$
$$\underline{a} = -\widetilde{V}'_{p,0}(\overline{Z}) - \varphi_+ B\overline{Z}^{\varphi_+-1} - \varphi_- C\overline{Z}^{\varphi_--1} - \frac{y}{r}$$

Proof. 1. Assume that $\widetilde{V}'(Z_t^X) = 0$ for some time t. Then by Proposition 2 $a_t = 0$, but $c_t = \big(u'\big)^{-1}(Z_t) > 0$ which is infeasible as $y = 0$. This implies that $\widetilde{V}'(Z) < 0$ for all $Z \in (0, \infty)$ and hence $X_t = 1$ for all t almost surely. In this case Z_t and Z_t^X coincide, (24) has to be solved everywhere, but the integrability condition would be violated unless $B = C = 0$ by (e) in Lemma 2, implying $\widetilde{V} = \widetilde{V}_{p,0}$.

2. Since in this case positive consumption can be financed even at $a = 0$, there must be some finite Z where $\widetilde{V}'$ is zero. Define $\widehat{Z} = \inf\{Z \mid \widetilde{V}'(Z) = 0\}$. Since $\widetilde{V}$ is decreasing and convex, it has to be constant for $Z \ge \widehat{Z}$. Use $\widehat{Z}$ and B to make $\widetilde{V}$ twice continuously differentiable to satisfy (i) in Proposition 2. Note that this is feasible with two free parameters, as the continuity of $\widetilde{V}$ itself is already guaranteed by the functional form. This is in contrast with Case 3., for example. Defining X as in (33) makes Z_t^X and hence $e^{-\delta\tau}(Z_\tau^X)^{\varphi_+}$ bounded from above. Due to this B being non-zero does not affect uniform integrability. On the other hand, the uniform integrability of $\widetilde{V}_{p,0}(Z_\tau)$ was assumed. This property can extended to $\widetilde{V}_{p,0}(Z_\tau^X)$ with X following (33) by an application of the reflection principle of the Wiener-process.

3. This time B and $\overline{Z}$ are set to match values and first derivatives of $\widetilde{V}$ and $\widetilde{U}$. Uniform integrability holds due to (c) in Lemma 2.

4. This time integrability holds due to (d) in Lemma 2., but otherwise this case is symmetric to 3.

5. and 6. can be obtained by combining proofs of previous cases.

□

Since for the optimal value it is irrelevant if there are no alternative states at all or there are, but they all happen to be inferior, Case 1. (i.e. no switch is optimal) corresponds to standard portfolio choice models like Merton (1969) and Samuelson (1969). This means that $\widetilde{V}_{p_0}$ can be thought of as the optimal dual function in a world without switches and borrowing constraint. The latter point holds since when income is zero, the borrowing constraint is optimally avoided by the agent, and hence is irrelevant. Case 2 has been treated in He and Pagès (1993). By Proposition 1 and Remark 1, introducing positive income to Case 1 but allowing for the natural borrowing constraint $b = \frac{y}{r}$ would simply add the positive term $Z\frac{y}{r}$ to the dual value function. This means that B represents the effect of a stricter borrowing constraint relative to the natural one, which intuitively lowers the value function, so we can expect $B < 0$. Since $\varphi_+ > 1$, this deviation from the unconstrained case increases in Z in a convex pattern, i.e. limited borrowing constraint affects utility relatively more when marginal utility is high. The upper limit $\widehat{Z}$ represents the fact that in this case optimal consumption is bounded from below by a positive number and hence marginal utility is bounded from above. From the technical point of view this bound is enforced by process X. In particular, X controls Z_t^X such that it cannot go over $\widehat{Z}$. In case 3. $\overline{Z}$ represents the marginal utility over which it is optimal to switch into another state. Now the sign of B is ambiguous, however B surely has to be larger than in case 2. since the optimal switch implies that changing states provides higher utility than waiting until reaching the borrowing constraint. A model falling into this class was analyzed in Jeanblanc et al. (2004). In case 4. $C > 0$, representing the extra utility from the option of moving to another state when marginal utility is low enough, i.e. when being rich enough. Models in cases 4. and 5. have been analyzed in the sizable early retirement literature. Having understood the functional form of the dual utility function, we can attempt to investigate how having options to switch in the future affect optimal policies, combining Theorems 2 and 2. In particular, we know that in the continuation region initial marginal utility λ is pinned down by

$$\widetilde{V}'_{p,0}(\lambda) + \varphi_+ B\lambda^{\varphi_+ - 1} + \varphi_- C\lambda^{\varphi_- - 1} = -\left(a_0 + \frac{y}{r}\right) \tag{34}$$

and λ determines consumption through u'. To obtain a clean formula on risky investments, we introduce the dual relative risk aversion coefficient as

$$\frac{1}{\tilde{\gamma}(\lambda)} = -\frac{\lambda\widetilde{V}''(\lambda)}{\widetilde{V}'(\lambda)}.$$

It is easily shown that if $\widetilde{V}$ is a convex conjugate, $\tilde{\gamma}(\lambda)$ equals the relative risk aversion of the primal value function evaluated at $\left(V'\right)^{-1}(\lambda)$.

$$\begin{aligned}
\xi_0 &= \frac{\mu}{\sigma^2}\lambda \widetilde{V}''(\lambda) = \frac{\mu}{\sigma^2}\Big[\lambda \widetilde{V}''_{p_0}(\lambda) + \varphi_+(\varphi_+ - 1)B\lambda^{\varphi_+ - 1} + \varphi_-(\varphi_- - 1)C\lambda^{\varphi_- - 1}\Big] \\
&= \frac{\mu}{\sigma^2}\Big[\frac{\lambda \widetilde{V}''_{p_0}(\lambda)}{\widetilde{V}'_{p_0}(\lambda)}\Big(\widetilde{V}'(\lambda) + \big(\widetilde{V}'_{p_0}(\lambda) - \widetilde{V}'(\lambda)\big)\Big) + \varphi_+(\varphi_+ - 1)B\lambda^{\varphi_+ - 1} + \varphi_-(\varphi_- - 1)C\lambda^{\varphi_- - 1}\Big] \\
&= \frac{\mu}{\sigma^2}\Big[\frac{a_0 + \frac{y}{r}}{\tilde{\gamma}(\lambda)} + \varphi_+\Big(\varphi_+ - 1 + \frac{1}{\tilde{\gamma}(\lambda)}\Big)B\lambda^{\varphi_+ - 1} + \varphi_-\Big(\varphi_- - 1 + \frac{1}{\tilde{\gamma}(\lambda)}\Big)C\lambda^{\varphi_- - 1}\Big]
\end{aligned}$$

For the ease of interpretation, assume temporarily that u is a CRRA utility function satifying the conditions of Proposition 3. In this case $\tilde{\gamma} = \gamma$ and both $\varphi_+\Big(\varphi_+ - 1 + \frac{1}{\gamma}\Big)$ and $\varphi_-\Big(\varphi_- - 1 + \frac{1}{\gamma}\Big)$ are positive. Note that when $B = C = 0$, we obtain the constant risky share solution from Merton (1969) and Samuelson (1969). When either B or C is 0, it is straightforward to infer the qualitative effects of switches on optimal policies. In Case 2. with $B < 0$, the presence of the borrowing constraint depresses both consumption and risky investment keeping a_0 and y constant. In Case 3. the sign of these effects is ambiguous, depending on whether or not the option to switch is more valuable than being subject to a borrowing constraint. Finally, in Case 4. the option of optimal switching when rich, decreases consumption, but increases risky saving. All these above statements can be found in the previously mentioned papers investigating the corresponding case. There are some shortcomings of this approach however: First, it is hard to make any general statements on cases 5. and 6. as then B and C are determined jointly and until now, their signs in this case are unclear. Also, the above discussion of optimal risky investment relied on assuming a special functional form for utility. Finally, all the above conclusions are qualitative in nature and hence give no insight on the size of these effects. These problems are in focus next.

3.3 Decomposition of dual value

Having completed the precise characterization of the dual value function, we turn to an alternative (albeit not so rigorous) derivation, providing us valuable intuition on the emergence and role of power functions in the value function. The final goal is linking the effects of optimal stopping on policy functions to economically interpretable fundamentals such as the value and expected time of switching. Considering an arbitrary stopping time τ and a decreasing process X, the candidate dual value function can be reorganized as below:

$$\begin{aligned}
&\mathbb{E}\bigg[\int_0^\tau e^{-\rho t}\Big(\widetilde{u}(Z_t^X)+yZ_t^X\Big)\,\mathrm{d}t+e^{-\rho\tau}\widetilde{U}(Z_\tau^X)\bigg]\\
&=\mathbb{E}\bigg[\int_0^\infty e^{-\rho t}\Big(\widetilde{u}(Z_t^X)+yZ_t^X\Big)\,\mathrm{d}t-\int_\tau^\infty e^{-\rho t}\Big(\widetilde{u}(Z_t^X)+yZ_t^X\Big)\,\mathrm{d}t+e^{-\rho\tau}\widetilde{U}(Z_\tau^X)\bigg]\\
&=\mathbb{E}\bigg[\int_0^\infty e^{-\rho t}\Big(\widetilde{u}(Z_t^X)+yZ_t^X\Big)\,\mathrm{d}t+e^{-\rho\tau}\bigg(\widetilde{U}(Z_\tau^X)-\int_\tau^\infty e^{-\rho(t-\tau)}\Big(\widetilde{u}(Z_t^X)+yZ_t^X\Big)\,\mathrm{d}t\bigg)\bigg]
\end{aligned} \tag{35}$$

This representation of dual utility is a replication of the analysis by Farhi and Panageas (2007) in a somewhat more general setting. The integral on the left is not a function of τ, but represents the value without any switch. On the other hand, the second term is the discounted net gain of switching, where $\widetilde{U}(Z_\tau^X)$ is the gain and $\int_\tau^\infty e^{-\rho(t-\tau)}\Big(\widetilde{u}(Z_t^X)+yZ_t^X\Big)\,\mathrm{d}t$ is the cost of leaving the current state. In particular, when there is no borrowing constraint and hence X can be ignored, there is a perfect analogue with exercising an American option as pointed out and discussed in detail by Farhi and Panageas (2007). I will show below that further insights can be gained by relying on what we already know about the optimal solution. First assume that τ is the first exit time of an interval $(\underline{Z},\overline{Z})$ where the endpoints are allowed to take values 0 and ∞, respectively. In addition, define X by

$$X_t=\min\Big\{1,\inf_{0\le s\le t}\frac{\widehat{Z}}{Z_s}\Big\}\qquad\forall t\ge 0,$$

where $\widehat{Z}$ is a possibly infinite positive number. Denote the first hitting times of the two endpoints by $\underline{\tau}$ and $\overline{\tau}$ and that of $\widehat{Z}$ by $\hat{\tau}$. Investigating such concrete policies has two main advantages. First, in this case expression (35) can be significantly simplified to an easier to interpret format. In addition, since policies correspond to real numbers, optimization over this special subset of policies can be performed simply by elementary calculus instead of relying on stochastic control methods. Since it was already proven that the optimal policies belong to the subset considered here, such an analysis is a valid tool to understand the full problem. Given these candidate policies, with some abuse of notation we can define the corresponding candidate dual value function as

$$\widetilde{V}(Z,\underline{Z},\overline{Z},\widehat{Z})=\mathbb{E}\bigg[\int_0^\tau e^{-\rho t}\Big(\widetilde{u}(Z_t^X)+yZ_t^X\Big)\,\mathrm{d}t+e^{-\rho\tau}\widetilde{U}(Z_\tau^X)\bigg]$$

Let us rearrange as

$$
\begin{aligned}
\widetilde{V}(Z,\underline{Z},\overline{Z},\widehat{Z}) = \mathbb{E}\Bigg[& \int_0^{\tau\wedge\hat{\tau}} e^{-\rho t}\Big(\widetilde{u}(Z_t)+yZ_t\Big)\,\mathrm{d}t + \mathbb{1}_{\tau<\hat{\tau}} e^{-\rho\tau}\widetilde{U}(Z_\tau) \\
& + \mathbb{1}_{\tau>\hat{\tau}}\Bigg(\int_{\hat{\tau}}^{\tau} e^{-\rho t}\Big(\widetilde{u}(Z_t^X)+yZ_t^X\Big)\,\mathrm{d}t + e^{-\rho\tau}\widetilde{U}(Z_\tau^X)\Bigg)\Bigg] \\
= \mathbb{E}\Bigg[& \int_0^{\infty} e^{-\rho t}\Big(\widetilde{u}(Z_t)+yZ_t\Big)\,\mathrm{d}t + \mathbb{1}_{\tau<\hat{\tau}}\Bigg(e^{-\rho\tau}\widetilde{U}(Z_\tau) - \int_{\tau}^{\infty} e^{-\rho t}\Big(\widetilde{u}(Z_t)+yZ_t\Big)\,\mathrm{d}t\Bigg) \\
& + \mathbb{1}_{\tau>\hat{\tau}}\Bigg(\int_{\hat{\tau}}^{\tau} e^{-\rho t}\Big(\widetilde{u}(Z_t^X)+yZ_t^X\Big)\,\mathrm{d}t + e^{-\rho\tau}\widetilde{U}(Z_\tau^X) - \int_{\hat{\tau}}^{\infty} e^{-\rho t}\Big(\widetilde{u}(Z_t)+yZ_t\Big)\,\mathrm{d}t\Bigg)\Bigg] \\
= \mathbb{E}\Bigg[& \int_0^{\infty} e^{-\rho t}\Big(\widetilde{u}(Z_t)+yZ_t\Big)\,\mathrm{d}t + \mathbb{1}_{\tau<\hat{\tau}} e^{-\rho\tau}\Bigg(\widetilde{U}(Z_\tau) - \int_{\tau}^{\infty} e^{-\rho(t-\tau)}\Big(\widetilde{u}(Z_t)+yZ_t\Big)\,\mathrm{d}t\Bigg) \\
& + \mathbb{1}_{\tau>\hat{\tau}} e^{-\rho\hat{\tau}}\Bigg(\int_{\hat{\tau}}^{\tau} e^{-\rho(t-\hat{\tau})}\Big(\widetilde{u}(Z_t^X)+yZ_t^X\Big)\,\mathrm{d}t + e^{-\rho(\tau-\hat{\tau})}\widetilde{U}(Z_\tau^X) \\
& - \int_{\hat{\tau}}^{\infty} e^{-\rho(t-\hat{\tau})}\Big(\widetilde{u}(Z_t)+yZ_t\Big)\,\mathrm{d}t\Bigg)\Bigg] \\
= \mathbb{E}\Bigg[& \widehat{V}_{p,0}(Z) + \frac{y}{r}Z + \mathbb{1}_{\tau\le\hat{\tau}} e^{-\rho\tau}\Bigg(\widetilde{U}(Z_\tau) - \widehat{V}_{p,0}(Z_\tau) - \frac{y}{r}Z_\tau\Bigg) \\
& + \mathbb{1}_{\tau>\hat{\tau}} e^{-\rho\hat{\tau}}\Bigg(\widetilde{V}(\widehat{Z},\underline{Z},\overline{Z},\widehat{Z}) - \widehat{V}_{p,0}(Z_{\hat{\tau}}) - \frac{y}{r}Z_{\hat{\tau}}\Bigg)\Bigg]
\end{aligned}
$$

where I used the solution from Case 1 in Theorem 2 to represent the appropriate integrals with $\widetilde{V}_{p_0}$. This could also be done with integrals starting at stopping times thanks to the strong Markov property of Itô-diffusions. Define

$$G(Z) = \widetilde{U}(Z) - \big(\widetilde{V}_{p_0}(Z) + \frac{y}{r}Z\big)$$

denoting the present net value of switching instantly if current marginal utility is Z. In addition, let

$$H(\widehat{Z},\underline{Z},\overline{Z}) = \widetilde{V}(\widehat{Z},\underline{Z},\overline{Z},\widehat{Z}) - \widehat{V}_{p,0}(Z_{\hat{\tau}}) - \frac{y}{r}Z_{\hat{\tau}}$$

be the utility loss from the presence of the borrowing limit when being constrained. Notice that both quantities are expressed relative a benchmark state in which no jump options or borrowing limit exist. To be able to get Z_τ out of the expectation operator we will make use of $Z_\tau = \Big(\mathbb{1}_{\tau=\overline{\tau}} + \mathbb{1}_{\tau=\underline{\tau}}\Big) Z_\tau = \mathbb{1}_{\tau=\overline{\tau}} Z_{\overline{\tau}} + \mathbb{1}_{\tau=\underline{\tau}} Z_{\underline{\tau}}$ as well. By substituting back,

$$\begin{aligned}\widetilde{V}(Z, \underline{Z}, \overline{Z}, \widehat{Z}) = & \widetilde{V}_{p_0}(Z) + \frac{y}{r} Z + \mathbb{E}\Big[e^{-\rho\overline{\tau}} \mathbb{1}_{\overline{\tau}=\min\{\overline{\tau},\underline{\tau},\hat{\tau}\}}\Big] G(\overline{Z}) \\ & + \mathbb{E}\Big[e^{-\rho\underline{\tau}} \mathbb{1}_{\underline{\tau}=\min\{\overline{\tau},\underline{\tau},\hat{\tau}\}}\Big] G(\underline{Z}) + \mathbb{E}\Big[e^{-\rho\hat{\tau}} \mathbb{1}_{\hat{\tau}<\tau}\Big] H(\widehat{Z}, \underline{Z}, \overline{Z})\end{aligned} \tag{36}$$

is obtained. Clearly, only one of $\overline{Z}$ and $\widehat{Z}$ can actually influence the dual value function at once. Indeed, if for example $\overline{Z} < \widehat{Z}$ holds, then with $\overline{Z}$ will be reached sooner almost surely. This however means that the borrowing limit almost never binds, hence it is intuitive that it should have no effect of the value either. On the other hand, if $\overline{Z} > \widehat{Z}$, then the borrowing limit would make it impossible to reach $\overline{Z}$ and hence the corresponding gain from switching is irrelevant. This also means that $\overline{Z}$ can be deleted from being among the arguments of function H.

Since all the above observations hold for an arbitrary triple $(\underline{Z}, \overline{Z}, \widehat{Z})$, they also have to be true for the optimal one. Therefore we have proved the following proposition:

Proposition 4. *Let $\widetilde{V}$ be the dual value function and $(\underline{Z}, \overline{Z}, \widehat{Z})$ denote the optimal thresholds as in Theorem 2. Then one of*

$$\widetilde{V}(Z) = \widetilde{V}_{p_0}(Z) + \frac{y}{r} Z + \mathbb{E}\Big[e^{-\rho\underline{\tau}} \mathbb{1}_{\underline{\tau}\le\hat{\tau}}\Big] G(\underline{Z}) + \mathbb{E}\Big[e^{-\rho\overline{\tau}} \mathbb{1}_{\overline{\tau}\le\underline{\tau}}\Big] G(\overline{Z}) \tag{37}$$

$$\widetilde{V}(Z) = \widetilde{V}_{p_0}(Z) + \frac{y}{r} Z + \mathbb{E}\Big[e^{-\rho\underline{\tau}} \mathbb{1}_{\underline{\tau}\le\hat{\tau}}\Big] G(\underline{Z}) + \mathbb{E}\Big[e^{-\rho\hat{\tau}} \mathbb{1}_{\hat{\tau}<\underline{\tau}}\Big] H(\widehat{Z}, \underline{Z}) \tag{38}$$

holds.

The above expressions have an intuitive interpretation: dual utility is additively separable for a term representing utility without any borrowing limit or switching options, one term for the expected utility gain when switching downwards and another term representing either the expected utility gain when switching upwards or the utility cost of the borrowing limit, respectively. Moreover, both latter terms are multiplicatively separable into a term representing the gain when switching (G) or cost of being bound by the borrowing limit (H) and another term for the expected subjective discount factor at the time of reaching the relevant threshold. These expectations are formally the Laplace-transforms of first exit times of a geometric Brownian motion from an interval, such that realizations when the other end of the interval is hit first, are given 0 value. Fortunately, analytic formulas for these objects exist and are available in Borodin and Salminen (2003). To emphasize the dependence of these expectations on Z, Z_* and Z^*, where $Z \in (Z_*, Z^*)$ denote

$$\underline{f}(Z, Z_*, Z^*) \equiv \mathbb{E}\Big[e^{-\rho\tau_{Z_*}}\mathbb{1}_{\tau_{Z_*}<\tau_{Z^*}}\Big] = \frac{Z^{*\varphi_+-\varphi_-}Z^{\varphi_-} - Z^{\varphi_+}}{Z^{*\varphi_+-\varphi_-}Z_*^{\varphi_-} - Z_*^{\varphi_+}} \tag{39}$$
$$\overline{f}(Z, Z_*, Z^*) \equiv \mathbb{E}\Big[e^{-\rho\tau_{Z^*}}\mathbb{1}_{\tau_{Z^*}<\tau_{Z_*}}\Big] = \frac{Z_*^{\varphi_+-\varphi_-}Z^{\varphi_-} - Z^{\varphi_+}}{Z_*^{\varphi_+-\varphi_-}Z^{*\varphi_-} - Z^{*\varphi_+}}$$

When one of the endpoints is such that it is never hit, we obtain the following special cases,

$$\underline{f}(Z, Z_*, \infty) = \left(\frac{Z}{Z_*}\right)^{\varphi_-} \tag{40}$$
$$\overline{f}(Z, 0, Z^*) = \left(\frac{Z}{Z^*}\right)^{\varphi_+}$$

Substituting these formulas into our decompositions (37) and (38), we can make two observations. First, when in the optimum the Z interval is bounded from only one side, comparing Theorem 2 with formulas (40) one can find a direct correspondence between coefficients B and C and economic fundamentals. For instance in case 2. $Z_* = 0$ and $Z^* = \widehat{Z}$ and $B = \frac{H(\widehat{Z},0)}{\widehat{Z}^{\varphi_+}}$ where $H(\widehat{Z}, 0)$ denotes the discounted total utility loss caused by the borrowing constraint at the moment when being at the borrowing limit. Thus the dual value describes the effect of the borrowing limit by separating it into two distinct terms: how adverse it is to hit the borrowing constraint, and a measure of when it is expected to happen the first time. This is intuitive, since the borrowing constraint only has a bearing on lifetime utility as long as it actually binds. Similar conclusions hold for cases 3. and 4., but in these cases stopping times and gains are linked to voluntary switching instead of hitting the borrowing limit.

Second, when exit times to both directions are relevant, there is no simple link between up- and downward switches and coefficients B and C, since both formulas in (39) contain both exponents of Z. Instead, to link properties of optimal policies to the values and expected times of switches, one must proceed further with the dual utility function decomposition.

3.4 Decomposition of optimal policies

As we know, marginal utility $\lambda = u'(c_0)$ is determined by $a_0 = -\widetilde{V}'(\lambda)$, therefore for example in the $B = C = 0$ case we have

$$a_0 + \frac{y}{r} = -\widetilde{V}_{p_0}'(\lambda_0). \tag{41}$$

Remember that $\widetilde{V}_{p_0}(\lambda)$ denotes the discounted expected utility from an optimal policy path under the assumption that the path starting with marginal utility

λ can be financed by the starting wealth level. Equation (41) then states that this consumption path can just be supported by a net worth of $-\widetilde{V}'_{p_0}(\lambda)$.Therefore $-\widetilde{V}'_{p_0}(\lambda)$ represents the wealth demanded to maintain a lifetime discounted utility of $\widetilde{V}_{p_0}(\lambda)$. Hence, interpreting the equations pinning down λ for the general cases

$$\begin{aligned} a_0 + \frac{y}{r} &= -\widetilde{V}'_{p_0}(\lambda) - G(\underline{Z})\frac{\partial \underline{f}(\lambda, \underline{Z}, \overline{Z})}{\partial \lambda} - G(\overline{Z})\frac{\partial \overline{f}(\lambda, \underline{Z}, \overline{Z})}{\partial \lambda} \\ a_0 + \frac{y}{r} &= -\widetilde{V}'_{p_0}(\lambda) - G(\underline{Z})\frac{\partial \underline{f}(\lambda, \underline{Z}, \widehat{Z})}{\partial \lambda} - H(\widehat{Z}, \underline{Z})\frac{\partial \overline{f}(\lambda, \underline{Z}, \widehat{Z})}{\partial \lambda} \end{aligned} \tag{42}$$

in their current form is difficult, since they contain both monetary terms (a, y/r and $\widetilde{V}'_{p,0}$) and utility-like terms (G and H). Furthermore, when trying to interpret $\xi_0 = \frac{\mu}{\sigma^2}\lambda\widetilde{V}''(\lambda)$ one would encounter a similar problem. This problem is overcome by the following proposition:

Proposition 5. *Let $\widetilde{V}$ be the dual value function and $(\underline{Z}, \overline{Z}, \widehat{Z})$ denote the optimal thresholds as in Theorem 2. Then optimal marginal utility λ is pinned down by one of*

$$\begin{aligned} a_0 + \frac{y}{r} &= -\widetilde{V}'_{p_0}(\lambda) - \underline{f}(\lambda, \underline{Z}, \overline{Z})\frac{\underline{Z}}{\lambda}G'(\underline{Z}) - \overline{f}(\lambda, \underline{Z}, \overline{Z})\frac{\overline{Z}}{\lambda}G'(\overline{Z}) \\ a_0 + \frac{y}{r} &= -\widetilde{V}'_{p_0}(\lambda) - \underline{f}(\lambda, \underline{Z}, \widehat{Z})\frac{\underline{Z}}{\lambda}G'(\underline{Z}) - \overline{f}(\lambda, \underline{Z}, \widehat{Z})\frac{\widehat{Z}}{\lambda}\frac{\partial H(\widehat{Z}, \underline{Z})}{\partial \widehat{Z}}. \end{aligned} \tag{43}$$

Furthermore, the optimal amount of risky investment is determined by

$$\begin{aligned} \xi_0 &= \frac{\mu}{\sigma^2}\left(\frac{-\widetilde{V}'_{p_0}(\lambda)}{\tilde{\gamma}(\lambda)} + \underline{g}(\lambda, \underline{Z}, \overline{Z})\frac{\underline{Z}}{\lambda}G'(\underline{Z}) + \overline{g}(\lambda, \underline{Z}, \overline{Z})\frac{\overline{Z}}{\lambda}G'(\overline{Z})\right) \\ \xi_0 &= \frac{\mu}{\sigma^2}\left(\frac{-\widetilde{V}'_{p_0}(\lambda)}{\tilde{\gamma}(\lambda)} + \underline{g}(\lambda, \underline{Z}, \widehat{Z})\frac{\underline{Z}}{\lambda}G'(\underline{Z}) + \overline{g}(\lambda, \underline{Z}, \widehat{Z})\frac{\widehat{Z}}{\lambda}\frac{\partial H(\widehat{Z}, \underline{Z})}{\partial \widehat{Z}}\right) \end{aligned} \tag{44}$$

for the two cases, where

$$\begin{aligned} \underline{g}(Z, Z_*, Z^*) &= (\varphi_- - 1)\underline{f}(Z, Z_*, Z^*) - \left(\frac{Z}{Z^*}\right)^{\varphi_+}\frac{\varphi_+ - \varphi_-}{\left(\frac{Z_*}{Z^*}\right)^{\varphi_-} - \left(\frac{Z_*}{Z^*}\right)^{\varphi_+}} \\ \overline{g}(Z, Z_*, Z^*) &= (\varphi_+ - 1)\overline{f}(Z, Z_*, Z^*) + \left(\frac{Z}{Z_*}\right)^{\varphi_-}\frac{\varphi_+ - \varphi_-}{\left(\frac{Z^*}{Z_*}\right)^{\varphi_+} - \left(\frac{Z^*}{Z_*}\right)^{\varphi_-}} \end{aligned}$$

The main message of this Proposition 5 is that both net wealth and risky investment are additively separable into three parts: One to finance consumption while being in the current state assuming no borrowing limits, and two others adjusting for the presence of switching options or the effects of a borrowing limit.

The intuition for the latter terms is straightforward: Following the analogy with $-\widetilde{V}'_{p_0}(\lambda)$, let us think of $-\widetilde{U}(Z)'$ as the necessary amount of wealth for an optimal path starting with marginal utility Z if switching states. This way of interpretation will be justified in Section 4. In this case $-G'(Z_\tau)$ is the additional demand for assets at time τ generated by exchanging the utility path implied by staying with the one associated with switching, keeping the marginal utility at time τ as given. Since this monetary cost is given in terms of time τ assets, we have to discount it to current times by multiplying with the ratios of marginal utilities at time τ and now. In addition, the effect of expected delay is taken into account through multiplying with $\overline{f}$ and $\underline{f}$. Combining these terms, for example $-\frac{\underline{Z}}{\lambda}G'(\underline{Z})\underline{f}$ represents the optimal amount of savings set aside today for the potential downward switch, when current marginal utility is λ. Similarly, $-\overline{f}\frac{\widehat{Z}}{\lambda}\frac{\partial H(\widehat{Z},\underline{Z})}{\partial \widehat{Z}}$ represents the extra savings induced by the lack of borrowing. Supporting this interpretation, it is easy to show that

$$-\frac{\partial H(\widehat{Z},\underline{Z})}{\partial \widehat{Z}} = -\widetilde{V}'(\widehat{Z}) + \widehat{V}'_{p,0}(\widehat{Z}) + \frac{y}{r} = -(-\widehat{V}'_{p,0}(\widehat{Z}) - \frac{y}{r})$$

holds, which expresses the cost of hitting the borrowing constraint expressed in monetary terms. Indeed, the right hand side is the negative of the monetary value of an optimal utility flow started from marginal utility $\widetilde{Z}$ assuming income y when the natural borrowing limit is available. To sum up, the total demand for assets can be separated into a term representing staying in the current state, and two terms adjusting for boundaries where eventual switches take place or a borrowing limit binds. Marginal utility λ and hence current consumption has to adjust such that the necessary level of wealth to finance all three aims of saving amounts to total net wealth $a + y/r$.

Consider next risky investments: the first term shows that the optimal risky investment to support future consumption when staying in the current state is determined as the standard optimal risky share $\frac{\mu}{\sigma^2\widetilde{\gamma}(\lambda)}$ multiplied by the amount of savings allocated for the same purpose. Similarly, we can compute the respective risky shares pertaining to the components of saving set aside for the time of reaching any of the boundaries. When there is no switch upwards, the option of switching at $\underline{Z}$ generates a risky share of $\frac{\mu}{\sigma^2}(1-\varphi_-)$ in the respective segment of wealth. Notice that in the case of CRRA preferences, this is strictly higher than the risky share of the first component, since $\varphi_- < 1 - \frac{1}{\gamma} < \varphi_+$ is implied by (29). In the presence of upward switches an additional term appears, further increasing the risky share in a decreasing manner with respect to the distance from the upper boundary. In particular, due to this extra term the risky investment demand generated by the option of downward switch does not converge to 0 in the neighborhood of the upper boundary, even though $\underline{f}$ does. As for the opposite situation, consider first again the simplest case: The risky share of the third segment of wealth is $-\frac{\mu}{\sigma^2}(\varphi_+ - 1)$ when no downward switch is available or optimal. Therefore the optimal amount of risky investment has the opposite sign as savings dedicated for an identical purpose. Moreover, when a downward switch is also present in the optimal policy, this effect is strengthened in a manner analogue to the opposite case.

In order to give some more concrete illustration to the above discussion and explain the source of asymmetries, it is worth considering the possible signs of G' in simple cases. First take case 2. from Theorem 2, that is the case of a borrowing constraint and no optimal switch to another state. Naturally, being at the borrowing limit the agent is worse off than not being constrained and hence $H(\widehat{Z}, \underline{Z}) < 0$. In addition, $\frac{\partial H(\widehat{Z}, \underline{Z})}{\partial \widehat{Z}}$ is also negative, as financing an optimal consumption path starting with a given marginal utility takes more initial wealth in the presence of borrowing constraints than without. Hence the extra wealth set aside due the borrowing limit is positive relative to the unlimited borrowing case, while optimal risky investment is depressed by the borrowing constraint, which is natural. Next consider case 3, or the case of optimal switch when marginal utility is high, i.e. wealth is low. In this case the sign of $G(\overline{Z})$ is ambiguous: when $G(\overline{Z}) < 0$, then switching provides less utility than staying ignoring the effects of the present borrowing limit. However, switching can still be optimal if it provides more utility than waiting in the current state until hitting the borrowing limit. One important fact however is that in optimum the signs of $G(\overline{Z})$ and $G'(\overline{Z})$ are identical, as follows from (47). In particular, when $G(\overline{Z}) < 0$, and the upward switch is not desirable relative to an unconstrained benchmark, but is still preferred to being constrained, the effects are qualitatively identical to the Case 2, since $G'(\overline{Z})$ is also negative. Of course, keeping the probability of hitting the respective boundaries constant, the effects are smaller in Case 3 than they would be without the switching option as in Case 2., since the switch being optimal implies $\overline{Z} < \widehat{Z}$ and $-G'(\overline{Z}) < -\frac{\partial H(\widehat{Z}, \underline{Z})}{\partial \widehat{Z}}$. In contrast, when $G(\overline{Z}) > 0$ and hence the switch even dominates staying in the unconstrained benchmark, $G(\overline{Z})$ is also positive. This induces negative allocated savings which means increased consumption speeding up the time when the boundary is reached. In addition, risky investment is also increased in optimum. Finally, consider Case 4., in which the optimal switch takes place when marginal utility is low enough. In this case $G(\underline{Z}) > 0$ always holds, otherwise setting the stopping time to ∞ would lead to higher expected utility. Then (48) implies that $G'(\underline{Z})$ is negative. This means that the utility flow after switching is more expensive to finance than the current state, considering marginal utility fixed at the boundaries. There is hence a trade-off: the new state offer higher utility, but switching to it involves some sort of monetary sacrifice. In this case the savings allocated to reaching the boundary is positive, to finance the cost of switching. This also lowers consumption and hence decreases the expected time of reaching the boundary. In addition, the risky investment generated by this saving motive is also positive.

The proof takes advantage of two facts. First, as optimal policies are considered, first order conditions with respect to $\underline{Z}$, $\overline{Z}$ and $\widehat{Z}$ need to hold. Second, partial derivatives of expectations $\underline{f}$ and $\overline{f}$ satisfy a useful set of identities. The latter fact is established in the following Lemma:

Lemma 3. *For all $Z \in (Z_*, Z^*)$, the following equalities hold:*

$$Z\frac{\partial \underline{f}}{\partial Z} = -Z_*\frac{\partial \underline{f}}{\partial Z_*} - Z^*\frac{\partial \underline{f}}{\partial Z^*} \tag{45}$$
$$Z\frac{\partial \overline{f}}{\partial Z} = -Z_*\frac{\partial \overline{f}}{\partial Z_*} - Z^*\frac{\partial \overline{f}}{\partial Z^*}$$

and

$$Z\frac{\partial^2 \underline{f}}{\partial^2 Z} = Z_*\left(\frac{1}{Z}\frac{\partial \underline{f}}{\partial Z_*} - \frac{\partial^2 \underline{f}}{\partial Z_* \partial Z}\right) + Z^*\left(\frac{1}{Z}\frac{\partial \underline{f}}{\partial Z^*} - \frac{\partial^2 \underline{f}}{\partial Z^* \partial Z}\right) \tag{46}$$
$$Z\frac{\partial^2 \overline{f}}{\partial^2 Z} = Z_*\left(\frac{1}{Z}\frac{\partial \overline{f}}{\partial Z_*} - \frac{\partial^2 \overline{f}}{\partial Z_* \partial Z}\right) + Z^*\left(\frac{1}{Z}\frac{\partial \overline{f}}{\partial Z^*} - \frac{\partial^2 \overline{f}}{\partial Z^* \partial Z}\right)$$

where the function arguments (Z, Z_*, Z^*) are omitted for readability.

Proof. Simple calculation delivers:

$$\begin{aligned}
-Z\frac{\partial \underline{f}(Z, Z_*, Z^*)}{\partial Z} =& \frac{\varphi_+ Z^{\varphi_+} - \varphi_- Z^{*\varphi_+ - \varphi_-} Z^{\varphi_-}}{Z^{*\varphi_+ - \varphi_-} Z_*^{\varphi_-} - Z_*^{\varphi_+}} \\
=& \frac{(\varphi_+ - \varphi_-) Z^{*\varphi_+ - \varphi_-} Z^{\varphi_-} - \varphi_+ (Z^{*\varphi_+ - \varphi_-} Z^{\varphi_-} - Z^{\varphi_+})}{Z^{*\varphi_+ - \varphi_-} Z_*^{\varphi_-} - Z_*^{\varphi_+}} \\
=& \frac{-(Z^{*\varphi_+ - \varphi_-} Z^{\varphi_-} - Z^{\varphi_+})(\varphi_- Z^{*\varphi_+ - \varphi_-} Z_*^{\varphi_-} - \varphi_+ Z_*^{\varphi_+})}{(Z^{*\varphi_+ - \varphi_-} Z_*^{\varphi_-} - Z_*^{\varphi_+})^2} \\
&+ \frac{(\varphi_+ - \varphi_-) Z^{*\varphi_+ - \varphi_-} Z^{\varphi_-} (Z^{*\varphi_+ - \varphi_-} Z_*^{\varphi_-} - Z_*^{\varphi_+})}{(Z^{*\varphi_+ - \varphi_-} Z_*^{\varphi_-} - Z_*^{\varphi_+})^2} \\
&- \frac{(Z^{*\varphi_+ - \varphi_-} Z^{\varphi_-} - Z^{\varphi_+})(\varphi_+ - \varphi_-) Z^{*\varphi_+ - \varphi_-} Z_*^{\varphi_-}}{(Z^{*\varphi_+ - \varphi_-} Z_*^{\varphi_-} - Z_*^{\varphi_+})^2} \\
=& Z_*\frac{\partial \underline{f}(Z, Z_*, Z^*)}{\partial Z_*} + Z^*\frac{\partial \underline{f}(Z, Z_*, Z^*)}{\partial Z^*}
\end{aligned}$$

The second equation in (45) follows from exchanging the roles of Z^* and Z_*. Finally, (46) follows from taking the partial derivatives of (45) with respect to Z and then substituting in (45) again into the derived expressions. □

Proof of Proposition 5. Throughout the whole proof, we need to consider the two cases separately when $\overline{Z} \leq \widehat{Z}$ or $\overline{Z} > \widehat{Z}$. Differentiate (36) with respect to $\underline{Z}$ and $\overline{Z}$ in the first case and $\underline{Z}$ and $\widehat{Z}$ in the other one. In case of an interior solution, the following first order conditions hold for optimal policies:

$$0 = G'(\underline{Z})\underline{f} + G(\underline{Z})\frac{\partial \underline{f}}{\partial \underline{Z}} + G(\overline{Z})\frac{\partial \overline{f}}{\partial \underline{Z}} \tag{47}$$

$$0 = G'(\overline{Z})\overline{f} + G(\underline{Z})\frac{\partial \underline{f}}{\partial \overline{Z}} + G(\overline{Z})\frac{\partial \overline{f}}{\partial \overline{Z}} \tag{48}$$

and

$$0 = G'(\underline{Z})\underline{f} + G(\underline{Z})\frac{\partial \underline{f}}{\partial \underline{Z}} + H(\widehat{Z}, \underline{Z})\frac{\partial \overline{f}}{\partial \underline{Z}} + \frac{\partial H(\widehat{Z}, \underline{Z})}{\partial \underline{Z}}\overline{f} \tag{49}$$

$$0 = \frac{\partial H(\widehat{Z}, \underline{Z})}{\partial \widehat{Z}}\overline{f} + G(\underline{Z})\frac{\partial \underline{f}}{\partial \widehat{Z}} + H(\widehat{Z}, \underline{Z})\frac{\partial \overline{f}}{\partial \widehat{Z}} \tag{50}$$

Intuitively, changing any of the boundary values affects the gain at switch and both expectations of discounted hitting times. In optimum these marginal effects have to cancel out. When the optimal boundaries are extremal, that is they take values 0 or ∞, the corresponding terms can be deleted as these boundaries are almost never reached and the rest of the proof goes through without any change. Before proceeding, note that $\frac{H(\widehat{Z},\underline{Z})}{\partial \underline{Z}} = 0$, since $H(\widehat{Z}, \underline{Z})$ depends on $\underline{Z}$ only though $\widetilde{V}(\widehat{Z}, \underline{Z}, \overline{Z}, \widehat{Z})$ and hence $\underline{Z}$ is chosen as the maximizer one, the corresponding partial derivative has to be 0. Of course, the same logic does not go through for $\widehat{Z}$ however, since $\underline{Z}$ influences $H(\widehat{Z}, \underline{Z})$ through two partial derivatives. Now by adding up $\underline{Z}$ times equation (47) (or (49)) and $\overline{Z}$ times equation (48) (or in the second case $\widehat{Z}$ times equation (50)) we can obtain

$$0 = \underline{Z}G'(\underline{Z})\underline{f} + \overline{Z}G'(\overline{Z})\overline{f} + G(\underline{Z})\left(\underline{Z}\frac{\partial \underline{f}}{\partial \underline{Z}} + \overline{Z}\frac{\partial \underline{f}}{\partial \overline{Z}}\right) + G(\overline{Z})\left(\underline{Z}\frac{\partial \overline{f}}{\partial \underline{Z}} + \overline{Z}\frac{\partial \overline{f}}{\partial \overline{Z}}\right)$$

and

$$0 = \underline{Z}G'(\underline{Z})\underline{f} + \widehat{Z}\frac{\partial H(\widehat{Z}, \underline{Z})}{\partial \widehat{Z}}\overline{f} + G(\underline{Z})\left(\underline{Z}\frac{\partial \underline{f}}{\partial \underline{Z}} + \widehat{Z}\frac{\partial \underline{f}}{\partial \widehat{Z}}\right) + H(\widehat{Z}, \underline{Z})\left(\underline{Z}\frac{\partial \overline{f}}{\partial \underline{Z}} + \widehat{Z}\frac{\partial \overline{f}}{\partial \widehat{Z}}\right).$$

Applying (45) and substituting into (42) evaluated at $Z = \lambda$ results in (43). By Theorem 2 optimal risky investment is determined by

$$\xi_0 = \frac{\mu}{\sigma^2}\left(\lambda \widetilde{V}''_{p_0}(\lambda) + \lambda G(\underline{Z})\frac{\partial^2 \underline{f}(\lambda, \underline{Z}, \overline{Z})}{\partial^2 \lambda} + \lambda G(\overline{Z})\frac{\partial^2 \overline{f}(\lambda, \underline{Z}, \overline{Z})}{\partial^2 \lambda}\right) \tag{51}$$

$$\xi_0 = \frac{\mu}{\sigma^2}\left(\lambda \widetilde{V}''_{p_0}(\lambda) + \lambda G(\underline{Z})\frac{\partial^2 \underline{f}(\lambda, \underline{Z}, \widehat{Z})}{\partial^2 \lambda} + \lambda H(\widehat{Z}, \underline{Z})\frac{\partial^2 \overline{f}(\lambda, \underline{Z}, \widehat{Z})}{\partial^2 \lambda}\right)$$

for the two cases. Now observe that the $\overline{Z}, \underline{Z}, \widehat{Z}$ satisfying the first order conditions (47) - (50) are constant in Z. This is because they represent controls affecting boundaries only: an boundary does not change only because the decision maker is somewhere else inside the continuation region. (This observation can be easily checked algebraically as well for (47) and (48), since for example $\underline{f}$, $\frac{\partial \underline{f}}{\partial \underline{Z}}$ and $\frac{\partial \overline{f}}{\partial \underline{Z}}$ contain Z through the same factor, which can therefore be canceled from the equation.) Take now partial derivatives of the first order conditions with respect to Z:

$$0 = G'(\underline{Z})\frac{\partial \underline{f}}{\partial Z} + G(\underline{Z})\frac{\partial^2 \underline{f}}{\partial \underline{Z}\partial Z} + G(\overline{Z})\frac{\partial^2 \overline{f}}{\partial \underline{Z}\partial Z} \tag{52}$$

$$0 = G'(\overline{Z})\frac{\partial \overline{f}}{\partial Z} + G(\underline{Z})\frac{\partial^2 \underline{f}}{\partial \overline{Z}\partial Z} + G(\overline{Z})\frac{\partial^2 \overline{f}}{\partial \overline{Z}\partial Z} \tag{53}$$

and

$$0 = G'(\underline{Z})\frac{\partial \underline{f}}{\partial Z} + G(\underline{Z})\frac{\partial^2 \underline{f}}{\partial \underline{Z}\partial Z} + H(\widehat{Z}, \underline{Z})\frac{\partial^2 \overline{f}}{\partial \underline{Z}\partial Z} \tag{54}$$

$$0 = \frac{\partial H(\widehat{Z}, \underline{Z})}{\partial \widehat{Z}}\frac{\partial \overline{f}}{\partial Z} + G(\underline{Z})\frac{\partial^2 \underline{f}}{\partial \widehat{Z}\partial Z} + H(\widehat{Z}, \underline{Z})\frac{\partial^2 \overline{f}}{\partial \widehat{Z}\partial Z} \tag{55}$$

Somewhat parallel to the case with first derivatives, we add up $\underline{Z}/z$ times equation (47) (or (49)), $-\underline{Z}$ times equation (52) (or (54)), $\overline{z}/z$ times equation (48) (or in the second case $\widehat{z}/z$ times equation (50)) and $-\overline{Z}$ times (53) (or $-\widehat{Z}$ times equation (55)) and hence obtain:

$$\begin{aligned}
0 =& \left(\frac{\underline{Z}}{Z}\underline{f} - \underline{Z}\frac{\partial \underline{f}}{\partial Z}\right)G'(\underline{Z}) + \left(\frac{\overline{Z}}{Z}\overline{f} - \overline{Z}\frac{\partial \overline{f}}{\partial Z}\right)G'(\overline{Z}) \\
&+ \left(\frac{\underline{Z}}{Z}\frac{\partial \underline{f}}{\partial \underline{Z}} - \underline{Z}\frac{\partial^2 \underline{f}}{\partial \underline{Z}\partial Z} + \frac{\overline{Z}}{Z}\frac{\partial \underline{f}}{\partial \overline{Z}} - \overline{Z}\frac{\partial^2 \underline{f}}{\partial \overline{Z}\partial Z}\right)G(\underline{Z}) \\
&+ \left(\frac{\underline{Z}}{Z}\frac{\partial \overline{f}}{\partial \underline{Z}} - \underline{Z}\frac{\partial^2 \overline{f}}{\partial \underline{Z}\partial Z} + \frac{\overline{Z}}{Z}\frac{\partial \overline{f}}{\partial \overline{Z}} - \overline{Z}\frac{\partial^2 \overline{f}}{\partial \overline{Z}\partial Z}\right)G(\overline{Z})
\end{aligned}$$

and

$$\begin{aligned}
0 =& \left(\frac{\underline{Z}}{Z}\underline{f} - \underline{Z}\frac{\partial \underline{f}}{\partial Z}\right)G'(\underline{Z}) + \left(\frac{\widehat{Z}}{Z}\overline{f} - \widehat{Z}\frac{\partial \overline{f}}{\partial Z}\right)\frac{\partial H(\widehat{Z}, \underline{Z})}{\partial \widehat{Z}} \\
&+ \left(\frac{\underline{Z}}{Z}\frac{\partial \underline{f}}{\partial \underline{Z}} - \underline{Z}\frac{\partial^2 \underline{f}}{\partial \underline{Z}\partial Z} + \frac{\widehat{Z}}{Z}\frac{\partial \underline{f}}{\partial \widehat{Z}} - \widehat{Z}\frac{\partial^2 \underline{f}}{\partial \widehat{Z}\partial Z}\right)G(\underline{Z}) \\
&+ \left(\frac{\underline{Z}}{Z}\frac{\partial \overline{f}}{\partial \underline{Z}} - \underline{Z}\frac{\partial^2 \overline{f}}{\partial \underline{Z}\partial Z} + \frac{\widehat{Z}}{Z}\frac{\partial \overline{f}}{\partial \widehat{Z}} - \widehat{Z}\frac{\partial^2 \overline{f}}{\partial \widehat{Z}\partial Z}\right)H(\widehat{Z}, \underline{Z})
\end{aligned}$$

Applying (46) and substituting into (51)

$$\xi_0 = \frac{\mu}{\sigma^2}\left(\lambda \widetilde{V}''_{p_0}(\lambda) + \frac{\underline{Z}}{\lambda}\left(\lambda\frac{\partial \underline{f}}{\partial Z} - \underline{f}\right)G'(\underline{Z}) + \frac{\overline{Z}}{\lambda}\left(\lambda\frac{\partial \overline{f}}{\partial Z} - \overline{f}\right)G'(\overline{Z})\right)$$

$$\xi_0 = \frac{\mu}{\sigma^2}\left(\lambda \widetilde{V}''_{p_0}(\lambda) + \frac{\underline{Z}}{\lambda}\left(\lambda\frac{\partial \underline{f}}{\partial Z} - \underline{f}\right)G'(\underline{Z}) + \frac{\widehat{Z}}{\lambda}\left(\lambda\frac{\partial \overline{f}}{\partial Z} - \overline{f}\right)\frac{\partial H(\widehat{Z},\underline{Z})}{\partial \widehat{Z}}\right)$$

obtained which implies the rest of the Proposition's statements. □

Proposition 5 shows that both the net worth and optimal risky investment of the agent can be decomposed into three terms corresponding to three saving purposes: First, to finance consumption from labor income and savings assuming staying in the current discrete state forever. Second, for saving to reach the upper wealth threshold of switch into an alternative state. The third (possibly negative) term either corresponds to a similar switching option optimal when wealth is low enough, or it stands for the precautionary saving generating by a borrowing limit. However, this decomposition is economically meaningful only if it is stable over time. Namely, thinking of the three components of net worth by saving purpose as three pockets of the agent, it should not be necessary to constantly rebalance the wealth across the pockets as random shocks arrive. Instead, all pockets should be financed separately as time evolves, using only their respective income. This turns out to be the case for the decomposition in Proposition 5, which is formalized below.

Proposition 6. *Define*

$$\hat{a}_{1,t} = -\widetilde{V}'_{p_0}(Z_t)$$

$$\hat{a}_{2,t} = -\underline{f}(Z_t,\underline{Z},\overline{Z})\frac{\underline{Z}}{Z_t}G'(\underline{Z})$$

$$\hat{a}_{3,t} = -\overline{f}(Z_t,\underline{Z},\overline{Z})\frac{\overline{Z}}{Z_t}G'(\overline{Z})$$

or

$$\hat{a}_{3,t} = -\overline{f}(Z_t,\underline{Z},\widehat{Z})\frac{\widehat{Z}}{Z_t}\frac{\partial H(\widehat{Z},\underline{Z})}{\partial \widehat{Z}}$$

and

$$\hat{\xi}_{1,t} = \frac{\mu}{\sigma^2} \frac{-\widetilde{V}'_{p_0}(Z_t)}{\tilde{\gamma}(Z_t)}$$
$$\hat{\xi}_{2,t} = \frac{\mu}{\sigma^2} \underline{g}(Z_t, \underline{Z}, \overline{Z}) \frac{\underline{Z}}{Z_t} G'(\underline{Z})$$
$$\hat{\xi}_{3,t} = \frac{\mu}{\sigma^2} \overline{g}(Z_t, \underline{Z}, \overline{Z}) \frac{\overline{Z}}{Z_t} G'(\overline{Z})$$

or

$$\hat{\xi}_{3,t} = \frac{\mu}{\sigma^2} \overline{g}(Z_t, \underline{Z}, \widehat{Z}) \frac{\widehat{Z}}{Z_t} \frac{\partial H(\widehat{Z}, \underline{Z})}{\partial \widehat{Z}}.$$

By Proposition 5, we have

$$a_t + \frac{y}{r} = \hat{a}_{1,t} + \hat{a}_{2,t} + \hat{a}_{3,t}$$

and

$$\xi_t = \hat{\xi}_{1,t} + \hat{\xi}_{2,t} + \hat{\xi}_{3,t}.$$

All three components of this decomposition are separately self-financing in the sense that

$$\mathrm{d}\hat{a}_{1,t} = \big(y + r\hat{a}_{1,t} - c_t\big)\,\mathrm{d}t + \hat{\xi}_{1,t}\big(\mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t\big) \tag{56}$$
$$\mathrm{d}\hat{a}_{2,t} = r\hat{a}_{2,t}\,\mathrm{d}t + \hat{\xi}_{2,t}\big(\mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t\big) \tag{57}$$
$$\mathrm{d}\hat{a}_{3,t} = r\hat{a}_{3,t}\,\mathrm{d}t + \hat{\xi}_{3,t}\big(\mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t,\big) \tag{58}$$

where c is defined by $u'(c_t) = Z_t$.

Proof. The proof is a straightforward application of Itô's Lemma. I particular, proving equation (56) is along the lines of the proof of Proposition 2, since $\widetilde{V}_{p_0}$ satisfies the HJB equation (24). Equations (57) and (58) hold thanks to the functional forms of f and g and the fact that ψ_+ and ψ_- satisfy equation (27). □

An important consequence of this result is that asking what share of net worth is allocated for a long term saving goal, or what share of net worth is effectively unavailable to the agent due to borrowing constraints are conceptually meaningful questions in this model.

While such a result relating to stopping times or a borrowing constraint is novel, there exists a formal analogue in classical optimal portfolio choice theory. When the agent derives utility from remaining wealth at a given final date T and consumption until time T, it is known that the optimal policies can similarly be decomposed into those of an agent caring only about wealth at the final date and one only concerned about consumption up to T, see Theorem 3.7.10 in Karatzas and Shreve (1998) and the preceding discussion.

4 Returning to the Sequential Problem

We have solved the First Stopping Problem with a general continuation value function U. However, in the full problem continuation values are provided by value functions of other states after taking the transactions costs into account, as in equation (3). In order to conclude that the results developed for the First Stopping Problem in Section 3 apply to the subproblems obtained from a full sequential problem presented in Section 2, one has to check that this U function satisfies the conditions demanded to utilize to apply the duality methods as in this paper. Assuming that for all $i \in \mathcal{I}$, a solution of the First Stopping Problem exists along line of Theorem 2 exists,

$$V_i(a) = \inf_{\lambda \geq 0} \left\{ \widetilde{V}_i(\lambda) + \lambda a \right\}$$

holds for all i, where $\widetilde{V}_i$ is a decreasing, convex function. Then by point (ii) in Remark 2, V_i is an increasing, concave function for all states i. Being the composite of increasing and concave functions, it is then easy to show that for all $j \in \mathcal{I}$, $V_j(a - P(i,j))$ inherits the same properties. Finally, U_i is the upper envelope of these functions, to which the same properties extend. More generally, but sacrificing some analytical tractability, P could be any convex function of wealth such that $\frac{\partial P(i,j,a)}{\partial a} < 1$ holds everywhere. In that case, $U_i(a) = \max_{j \in \mathcal{I} \setminus i} V_j(a - P(i,j,a))$ would still be an increasing concave function and hence the duality approach utilized in this paper could be applied.

When discussing the intuition behind optimal policies in Section 3.3, it was suggested that $-\widetilde{U}'(Z)$ is the monetary equivalent of the utility gained when switching and hence $-G'(Z)$ can be interpreted as the monetary equivalent of the net utility change. This assumption is justified next, after explicitly computing $\widetilde{U}$, when U is derived from the set of First Stopping Problems. Since the convex conjugate function of $V_j(a - P(i,j))$ is

$$\begin{aligned}
\sup_a \left\{ V_j(a - P(i,j)) - Za \right\} =& V_j(a - P(i,j)) - V_j'\big(a - P(i,j)\big)a \Big|_{V_j'\left(a - P(i,j)\right) = Z} \\
=& V_j(a - P(i,j)) - V_j'\big(a - P(i,j)\big)(a - P(i,j)) \\
& - V_j'\big(a - P(i,j)\big)P(i,j)\Big|_{V_j'\left(a - P(i,j)\right) = Z} \\
=& \widetilde{V}_j(Z) - ZP(i,j)
\end{aligned}$$

for all i, j, by point (v) in Remark 2 the dual of the continuation value function is

$$\widetilde{U}_i = \max_{j \in \mathcal{I} \setminus i} \widetilde{V}_j(Z) - ZP(i,j). \tag{59}$$

It is worth mentioning that a nearly as tractable result could also be obtained if P were allowed to be an affine function of current wealth. Indeed, an example for such a framework is Jeanblanc et al. (2004), who assume that when going bankrupt, the household has to pay a fixed penalty and also loses a certain share of her remaining wealth in addition. Now we can investigate what terms $-G'$ consists of. Thanks to the triangle inequality from Assumption 1, we can assume that from state j there is no further transition and hence the decomposition from Proposition 4 can be substituted in for $\widetilde{V}_j$. Therefore,

$$\begin{aligned} -G_i'(Z) = &- \widetilde{U}_i'(Z) - \big(- \widetilde{V}_{p_0,i}'(Z) - \frac{y_i}{r} \big) = -\widetilde{V}_{p_0,j}'(Z) - \big(- \widetilde{V}_{p_0,i}'(Z) \big) \\ &- \frac{Z_{*,j}}{Z} G_j'(Z_{*,j}) \underline{f}(\lambda, Z_{*,j}, Z_j^*) - \frac{Z_j^*}{\lambda} G_j'(Z_j^*) \overline{f}(\lambda, Z_{*,j}, Z_j^*) + \frac{y_i - y_j}{r} + P(i,j) \end{aligned} \tag{60}$$

holds, where with some abuse of notation, $G_j'(Z_j^*)$ denotes either $G_j'(\overline{Z}_j)$ or $\frac{\partial H(\widehat{Z}, \underline{Z})}{\partial \widehat{Z}}$. Equation (60) states that $-G'$ consists of the difference of the wealth levels demanded to maintain the utility flows in states j and i respectively, the saving needs of the switching options available from state j (or the effect of a potential borrowing limit) and finally the net of discounted labor income flows in the two states plus the transaction cost.

So far it was discussed that the methods and intuition in Section 3 apply naturally to such First Stopping Problems, which are subproblems of the sequential optimization problems being the subject of this paper. Next, using the results from Section 3 it is shown that when transaction costs are regular enough and all felicity functions belong to the constant risk aversion class, transversality condition (4) is satisfied automatically and hence by Theorem 1 it is always sufficient to solve the First Stopping Problems to find a solution of the full sequential problem.

Proposition 7. *Suppose that the set of transaction costs satisfies Assumption 1 and for all states $i \in \mathcal{I}$ the felicity function is of the constant risk aversion type*

$$u(c,i) = h_i^{\gamma_i} \frac{c^{1-\gamma_i}}{1-\gamma_i} + n_i$$

with $\gamma_i > 0$ satisfying

$$\rho > (1-\gamma_i)\Big[r + \frac{\theta}{\gamma_i} \Big]$$

and $h_i > 0$ and $n_i \in \mathbb{R}$. Furthermore, assume that the set of First Stopping Problems is solved by conforming U_i functions as posited among the conditions of Theorem 1. Then the transversality condition (4) in Theorem 1 also holds.

Proof. As it was discussed in Section 2, Assumption 1 implies that for any admissible sequence of switching times we have $\tau_k \to \infty$ almost surely. Therefore the tranversality condition holds if

$$\mathbb{E}_0 \Big[e^{-\delta t} U_{i_n}(a_t) \Big] \to 0$$

as $t \to \infty$ for any admissible i_n. By the definition of U and convex conjugates and applying Theorem 2

$$
\begin{aligned}
U_i(a_t) &= \max_{j \in \mathcal{I} \setminus i} V_j(a_t - P(i,j)) \\
&= \max_{j \in \mathcal{I} \setminus i} \inf_{\lambda} \widetilde{V}_j(\lambda) - \lambda P(i,j) + \lambda a_t \\
&= \max_{j \in \mathcal{I} \setminus i} \widetilde{V}_j(Z_t^X) - Z_t^X P(i,j) + Z_t^X \widetilde{V}_j'(Z_t^X)
\end{aligned}
$$

where Z_t^X is the optimally controlled marginal utility process started from Z_0^X such that $\widetilde{V}_{i_0}'(Z_0^X) = a_0$ holds. For now fix j and consider

$$
\widetilde{V}_j(Z_t^X) = \frac{\gamma_j}{1-\gamma_j} h_j A_j (Z_t^X)^{1-1/\gamma_j} + \frac{n_j}{\rho} + B_j (Z_t^X)^{\varphi_+} + C_j (Z_t^X)^{\varphi_-} + \frac{y_j}{r} Z_t^X
$$

where $Z_{*j} \leq Z_t^X \leq Z_j^*$ with $C_j = 0$ whenever $Z_{*j} = 0$ and $B_j = 0$ when $Z_j^* < \infty$. Then

$$
\begin{aligned}
|U_i(a_t)| = \max_{j \in \mathcal{I} \setminus i} \Big\{ & \frac{h_j A_j}{1-\gamma_j} (Z_t^X)^{1-1/\gamma_j} + \frac{n_j}{\rho} + (1-\varphi_+) B_j (Z_t^X)^{\varphi_+} \\
& + (1-\varphi_-) C_j (Z_t^X)^{\varphi_-} - Z_t^X P(i,j) \Big\} \\
\leq & \max_{j \in \mathcal{I} \setminus i} \Big\{ \big| \frac{h_j A_j}{1-\gamma_j} \big| (Z_t^X)^{1-1/\gamma_j} \Big\} + \max_{j \in \mathcal{I} \setminus i} \Big\{ \frac{|n_j|}{\rho} \Big\} + \max_{j \in \mathcal{I} \setminus i} \Big\{ |(1-\varphi_+) B_j| (Z_j^*)^{\varphi_+} \Big\} \\
& + \max_{j \in \mathcal{I} \setminus i} \Big\{ |(1-\varphi_-) C_j| (Z_{*j})^{\varphi_-} \Big\} + Z_t^X \max_{j \in \mathcal{I} \setminus i} \Big\{ |P(i,j)| \Big\}
\end{aligned}
$$

Therefore $e^{-\delta t}|U_i(a_t)|$ and hence $e^{-\delta t} U_i(a_t)$ converge to 0 as $t \to \infty$, if $e^{-\delta t}(Z_t^X)^{1-1/\gamma_j}$ and $e^{-\delta t} Z_t^X$ do so. This however was already proven in Lemma 2 for the case when X is constant 1, while the general case can be shown by applying the reflection principle of Wiener-processes. □

5 Example: Owning or renting

To illustrate the uses and limitations of the results derived in this paper to understand solutions of combined optimal stopping and portfolio choice problems, next we investigate a concrete example. There are two discrete states representing renting and owning a home, both of which corresponds to a fixed housing level $\hat{h}_i$ with $i \in \{R, O\}$. The agent has a Cobb-Douglas utility function over housing and non-durable goods and a CRRA felicity function over the composite good, i.e.

$$
u_i(c) = \frac{\left(\hat{h}_i^{\omega} c^{1-\omega} \right)^{1-\hat{\gamma}}}{1-\hat{\gamma}} \qquad i \in \{R, H\}.
$$

By defining $\gamma = 1 - (1-\omega)(1-\hat{\gamma})$ and $h_i = \left(\frac{1-\gamma}{1-\hat{\gamma}} \hat{h}_i^{\omega(1-\hat{\gamma})}\right)^{1/\gamma}$, we can immediately bring these utility functions in the form of the one in Proposition 3. Assume in addition that γ satisfies the parameter restriction of the same Proposition. The transaction costs to move from one state to the other are as follows: buying a house costs $\hat{P}(R, O) = P$ and selling it results in a cost of $\hat{P}(O, R) = -\alpha P$ where $0 < \alpha < 1$ represents selling costs as a fraction of the house value. Note that these transaction costs are regular, for example with $s_R = 0$ and any $\alpha P < s_O < P$. The net income of renters is the difference of labor income and the rental cost $y_R = y - m$, while home owners simply obtain the same labor income: $\hat{y}_O = y$. Finally, the two discrete states differ across their borrowing limits: In particular, renters cannot borrow $b_R = 0$, while home owners may take a mortgage up to a δ fraction of their home's purchase value $\hat{b}_O = \delta P$. However, by utilizing Proposition 1, we can normalize $b_O = 0$, if we adjust all other quantities related to owning as well: Define therefore

$$P(R, O) = (1-\delta)P, \qquad P(O, R) = (\delta - \alpha)P, \qquad y_O = y - \delta P r. \tag{61}$$

The interpretation of these transformed equations is straightforward. Taking into account changes in borrowing ability, the true cost of buying a house is only $(1-\delta)$ times the house price, i.e. the value of the minimal down-payment. Turning this logic around, when selling the home, one's borrowing capacity drops by δP which has to be subtracted from the nominal income from selling. Finally, income in state O is burdened by interest payments on the mortgage. Of course, this can be offset by holding more liquid wealth than the borrowing capacity.

Assumption 4. Parameters satisfy

(i)
$$h_R = h_O = 1$$
and

(ii)
$$\alpha > \delta.$$

Point (i) is assumed for simplicity and it means that the rented and owned home we consider in this example provide identical utility flows. Setting these values to 1 is a matter of normalization. Point (ii) implies that transaction costs are not too large in the sense that selling a house results in positive additional liquidity even when completely indebted. Since by point (i) the two states solely differ in terms of income and the asymmetry generated by the transactions costs, the problem becomes very tractable. In particular, it can be shown that in this case only one-way transitions are present in the solution, i.e. either renting or owning becomes an endogenous final state from there no transitions are optimal. This is intuitive, since as the transformed problem clarifies, depending in the relation of y_R and y_O one of the states is uniformly superior if considering the two states in isolation. Since transitions back and forth involve a wealth loss due to transaction costs simply staying in the better state of the two remains the best policy even when taking possible transitions into account.[5]

Knowing that only one-way transitions can be optimal, we can write the gains from transitions as follows:

$$G_O(Z) = \frac{Z}{\widehat{Z}_R}\widehat{H}_R\overline{f}(Z,0,\widehat{Z}_R) + \frac{y_R - y_O}{r} - (\delta - \alpha)P = \left(\frac{Z}{\widehat{Z}_R}\right)^{\varphi_+}\widehat{H}_R - \frac{m}{r} + \alpha P$$
$$G_R(Z) = \frac{Z}{\widehat{Z}_O}\widehat{H}_O\overline{f}(Z,0,\widehat{Z}_O) + \frac{y_O - y_R}{r} - (1 - \delta)P = \left(\frac{Z}{\widehat{Z}_O}\right)^{\varphi_+}\widehat{H}_O + \frac{m}{r} - P$$

As the utility effect of hitting the borrowing constraint $\widehat{H}$ is negative, we can determine which transition are optimal depending on the relation of price P and the present value of rental costs $\frac{m}{r}$.

1. $\frac{m}{r} < \alpha P$

   In this case $G_R(Z)$ is always negative, but $G_O(Z)$ is positive when Z is small enough. Therefore it is never optimal to buy a house, but selling is advantageous when Z is small enough. Note that when $\delta P < \frac{m}{r} < \alpha P$, then $y_O > y_R$, therefore without transitions being allowed, it would be better to be born as a homeowner than as a renter. However, selling one's home can still be advantageous, since the income from selling makes up for the lifetime loss of paying a higher rent than the debt costs before selling.

2. $\alpha P \leq \frac{m}{r} \leq P$

   In this case both $G_R(Z)$ and $G_O(Z)$ are negative for all finite Z, making all transitions sub-optimal. This is intuitive, since for buying the house one would need to sacrifice more income at once than the present value of lifetime rental costs. On the other hand, the income from selling would not cover lifetime rental costs for the rest of the time.

3. $P < \frac{m}{r}$

   In this case it is never optimal to sell a house, but buying is advantageous when Z is small enough.

For the sake of illustration let us consider case 3., which perhaps gives rise to the most intuitive setup: It is more advantageous to own a home than renting, but buying is optimal only for sufficiently wealthy agents. We are interested in how the presence of buying affects optimal policies of renters even before the wealth threshold

[5]When $\hat{h}_R < \hat{h}_O$, selling for poor owners and buying for rich renters can be optimal at the same time. Such a more general case will be discussed in a later version of the paper.

for optimal buying is reached. In order to visually represent the decomposition of optimal policies implied by Proposition 5, we first need to compute $\widehat{H}_O$, $\widehat{H}_R$, $\widehat{Z}_O$, $\widehat{Z}_R$ and $\underline{Z}_R$. This is performed in two steps for both owning and renting. First, being a final state, $\widehat{Z}_O$ and parameter B of the dual utility function for home owning can be computed down using to Case 2. in Theorem 2. The corresponding $\widehat{H}_O$ follows from $\widehat{H}_O = B_O \widehat{Z}_O^{\varphi_+}$. Having determined the value of home-owning, we obtain the necessary continuation value when buying as a renter, namely U_R and G_R. Next, the dual value function parameters for renting can be computed using the conditions from Case 5. in Theorem 2. Finally, $\widehat{H}_R$ is pinned down by comparing the obtained B-C representation of $\widetilde{V}_R$ with (38). For the sake of comparison, with a similar strategy I also obtained the solutions of two suitable benchmarks: the renter's problem without an option to buy, and the renter's problem with no borrowing limit (or equivalently, with the natural borrowing limit) and no buying option. Let us begin with the latter case, i.e. when the problem collapses into the one described in Merton (1969). On the left panel of Figure 1, the sum of the two gray areas equals to net worth, i.e. the sum of wealth and human capital, which in this model simply equals y/r.[6]

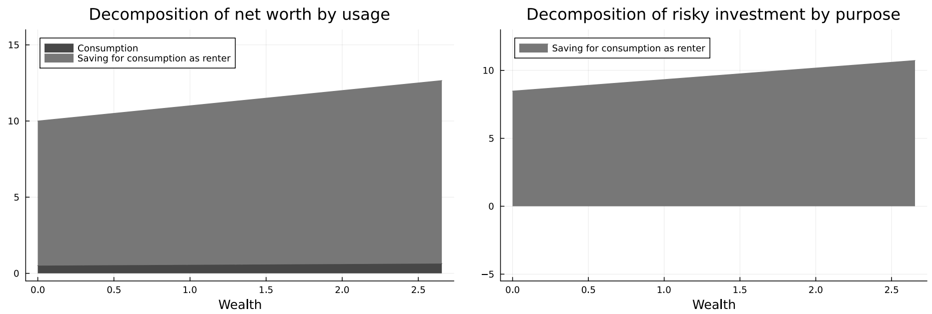

Figure 1: Optimal consumption, saving (defined as not consumed net worth) and risky investment policy functions, when no borrowing limit, and no buying option exists.

Consumption is a linear function of net worth, just as risky investment, which is depicted on the right.

As shown on Figure 2, this simple picture changes significantly when we add the borrowing limit to the model.

[6]The parameters used for this example are presented in Appendix B.

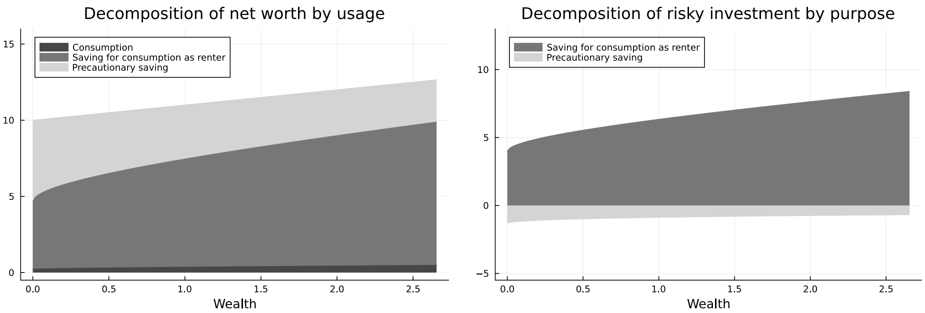

Figure 2: Optimal consumption, saving (defined as not consumed net worth) and risky investment policy functions. Borrowing limit present, but no buying option.

As implied by the decomposition in (43) in Proposition 5, the term representing the borrowing constraint's effect effectively blocks a portion of net worth from being used in consumption-saving problem of the renter. Therefore due to the borrowing limit, the agent considers a part of her net worth absent (the light gray region), and takes only the rest into account when determining optimal consumption and stock holdings which otherwise is done exactly as in the previous case. The fraction of net worth ignored when deciding on consumption is interpreted as precautionary saving. Note the depressing effect due to the constraint is increasing as wealth getting closer to 0. Risky investments are composed of two parts according to equation (44): the first one is a linear function of the resources left after subtracting the borrowing limit's effect. The second is an additional negative term depressing risk taking even further, especially for poorer agents.

Introducing the option to buy a house gives rise one more term from the decompositions in Proposition 5. Now a portion of net worth is reserved to save money for the eventual home purchase. As shown in the left panel of Figure 3, this effect is stronger close to the optimal wealth level to buy (which is the right end of the x-axis.), while the borrowing limit is more relevant for the poor.

However, due to the presence of the extra terms in the g functions from Proposition 5, the decomposition of optimal stock holdings is somewhat difficult to interpret. It is clear that the possibility of buying encourages, while the the borrowing constraint discourages stock holdings. However, the appearance of the purchase option changed the strength of the borrowing limit's effect, due to $\underline{Z}_R$ being an argument of $\overline{g}$. In particular, the depressing effect from the borrowing limit is not decreasing in wealth anymore. All this makes it somewhat difficult to disentangle the effects of the option to buy a house from the borrowing limit.

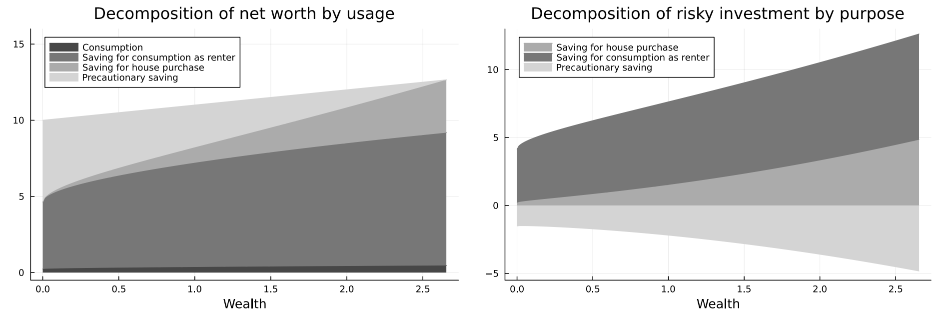

Figure 3: Optimal consumption, saving (defined as not consumed net worth) and risky investment policy functions. Borrowing limit and buying option present.

To make the difference in terms of policies more visible, Figure 4 depicts the percentage changes in consumption and optimal stock holding policies due to the presence of the house buying option, relative to the policies when no such possibility was available, but the borrowing limit existed.

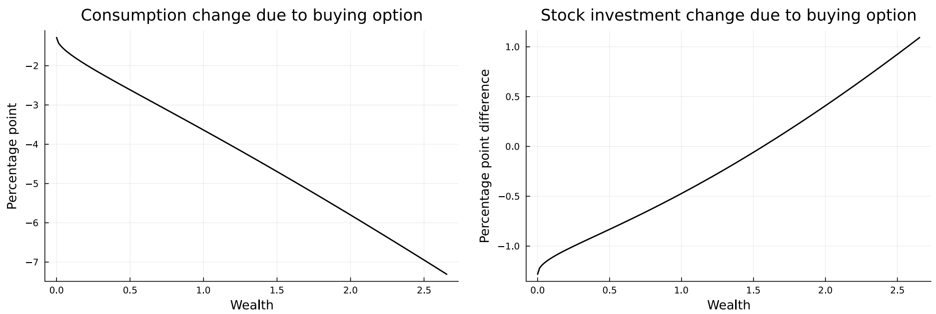

Figure 4: Percentage point deviation of consumption and risky investment policies when buying option is available relative to the case when buying is not allowed.

The saving motive to buy a house decreases consumption over all wealth levels, and the size of the drop is monotonic increasing in wealth, so as the buying threshold is getting closer. On the other hand, the effect on stock holdings is not uniform: while around the borrowing limit the agent becomes more risk averse, closer to the optimal purchase threshold the agent is more risk seeking. Intuitively, it is optimal to avoid being stuck on a low wealth level by being more cautious, but later more risk taking is better in order to accelerate growth.

6 Conclusion

In this paper I investigate the effect of discrete choices on saving and portfolio choice decisions in an extension of the classical model of Merton (1969). These discrete decisions are interpreted as phase transitions and are allowed to alter one's felicity function, borrowing limit or income. As in this model the only state variable is wealth, the optimal timing of discrete transitions is linked to wealth thresholds. The main contribution is showing that the value and optimal policies can be decoupled into a part representing a benchmark model without any switching thresholds, and a part standing for exactly the effect of those. The first term is pinned down by the functional form of the felicity function corresponding to the current state. On the other hand, the terms corresponding to boundaries only depend on the value of reaching the boundaries and some measure of the time it takes to reach them. In addition, the functional form of these additional terms in the optimal value and policies is something inherent to the problem and not related directly on the alternative state into which the transition occurs. The felicity function of other states only plays a role through determining the ideal switch time and the value of transition.

In addition, I show that when only one switching boundary is present, the effects of the presence of the corresponding transition option are tractable and intuitive. Even when two boundaries are active, the decomposition discussed before provides some basis for interpretation. Unfortunately, as illustrated in Section 5, for some important economic problems a full understanding of the involved mechanisms is far from straightforward, which brings up to the directions where this paper could be further developed. The decomposition in section 3.3 and later on relies on comparing switching to staying in a benchmark setup with no borrowing constraint or transition option. However, for some applications it would be key to evaluate the effect of an additional option of transition in isolation. Providing an alternative decomposition, that allows for such an exercise would be valuable addition to this paper. Another important space of improvement is conceptual. In this version I do not pursue the question of existence of solutions and instead rely on existence results provided in slightly different frameworks, which cover the CRRA case, the functional form used for all examples in this paper. Instead, I would attempt to build an existence result within my framework. Apart from completeness, this may be interesting if a condition based on Assumption 3 could be build provided, since it would represent a 'deeper' property relative to the ones in literature, relying on estimating the utility function from above by a CRRA function.

Finally, on a more general level, a somewhat surprising feature of the model being subject of this paper is the existence of an analytical solution in spite of the presence of uncertainty, a borrowing constraint and discrete decisions. Apart from enabling the more complete understanding of the economic intuitions behind the problem, this fact is also of interest from a different , more practical aspect. Since Kaplan et al. (2018), a rapidly growing literature in macroeconomics builds heterogeneous agent models in continuous time. Given the incomplete market nature of these models, the solution of the household's optimization problem is obtained by solving the Hamilton-Jacobi-Bellman numerically, typically with finite difference methods. If there is way of introducing labor income risk in the framework investigated in this paper while retaining the property of having a closed-form solution, that could provide a basis for a heterogeneous agent economy without the usual high computational cost of solving such models. This question I am planning to address in future research.

References

Emilio Barucci and Daniele Marazzina. Optimal investment, stochastic labor income and retirement. *Applied Mathematics and Computation*, 218(9):5588–5604, 2012. ISSN 0096-3003. doi: https://doi.org/10.1016/j.amc.2011.11.052. URL https://www.sciencedirect.com/science/article/pii/S0096300311013956.

J. Joseph Beaulieu. Optimal durable and nondurable consumption with transactions costs. Technical report, 1993.

Andrei N. Borodin and Paavo Salminen. Handbook of brownian motion - facts and formulae (second edition). 2003.

Kyoung Jin Choi and Gyoocheol Shim. Disutility, Optimal Retirement, And Portfolio Selection. *Mathematical Finance*, 16(2):443–467, April 2006. doi: 10.1111/j.1467-9965.2006. URL https://ideas.repec.org/a/bla/mathfi/v16y2006i2p443-467.html.

Kyoung Jin Choi, Gyoocheol Shim, and Yong Hyun Shin. Optimal Portfolio, Consumption-Leisure And Retirement Choice Problem With Ces Utility. *Mathematical Finance*, 18(3):445–472, July 2008. doi: 10.1111/j.1467-9965.2008. URL https://ideas.repec.org/a/bla/mathfi/v18y2008i3p445-472.html.

John C Cox and Chi fu Huang. Optimal consumption and portfolio policies when asset prices follow a diffusion process. *Journal of Economic Theory*, 49(1):33–83, 1989. ISSN 0022-0531. doi: https://doi.org/10.1016/0022-0531(89)90067-7. URL https://www.sciencedirect.com/science/article/pii/0022053189900677.

Domenico Cuoco and Hong Liu. Optimal consumption of a divisible durable good. *Journal of Economic Dynamics and Control*, 24(4):561–613, 2000. URL https://EconPapers.repec.org/RePEc:eee:dyncon:v:24:y:2000:i:4:p:561-613.

Anders Damgaard, Brian Fuglsbjerg, and Claus Munk. Optimal consumption and investment strategies with a perishable and an indivisible durable consumption good. *Journal of Economic Dynamics and Control*, 28(2):209–253, November 2003. URL https://ideas.repec.org/a/eee/dyncon/v28y2003i2p209-253.html.

Jerome B. Detemple and Christos I. Giannikos. Asset and commodity prices with multi-attribute durable goods. *Journal of Economic Dynamics and Control*, 20(8):1451 – 1504, 1996. ISSN 0165-1889. doi: https://doi.org/10.1016/0165-1889(95)00907-8. URL http://www.sciencedirect.com/science/article/pii/0165188995009078.

Philip Dybvig and Hong Liu. Lifetime consumption and investment: Retirement and constrained borrowing. *Journal of Economic Theory*, 145(3):885–907, 2010.

URL https://EconPapers.repec.org/RePEc:eee:jetheo:v:145:y:2010:i:3:p:885-907.

Emmanuel Farhi and Stavros Panageas. Saving and investing for early retirement: A theoretical analysis. *Journal of Financial Economics*, 83(1):87–121, January 2007. URL https://ideas.repec.org/a/eee/jfinec/v83y2007i1p87-121.html.

Sanford J. Grossman and Guy Laroque. Asset pricing and optimal portfolio choice in the presence of illiquid durable consumption goods. *Econometrica*, 58(1):25–51, 1990. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/2938333.

Hua He and Henri F. Pagès. Labor income, borrowing constraints, and equilibrium asset prices. *Economic Theory*, 3(4):663–696, 1993. ISSN 09382259, 14320479. URL http://www.jstor.org/stable/25054733.

Ayman Hindy and Chi fu Huang. Optimal consumption and portfolio rules with durability and local substitution. *Econometrica*, 61(1):85–121, 1993. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/2951779.

Monique Jeanblanc, Peter Lakner, and Ashay Kadam. Optimal bankruptcy time and consumption/investment policies on an infinite horizon with a continuous debt repayment until bankruptcy. *Mathematics of Operations Research*, 29(3):649–671, 2004. ISSN 0364765X, 15265471. URL http://www.jstor.org/stable/30035669.

Junkee Jeon and Hyeng Keun Koo. A unified approach to retirement and consumption-portfolio choice, 2021.

Greg Kaplan, Benjamin Moll, and Giovanni L. Violante. Monetary policy according to hank. *American Economic Review*, 108(3):697–743, March 2018. doi: 10.1257/aer.20160042. URL https://www.aeaweb.org/articles?id=10.1257/aer.20160042.

I. Karatzas and S.E. Shreve. *Methods of Mathematical Finance*. Applications of Mathematics : Stochastic Modelling and Applied Probability. Springer, 1998. ISBN 9780387948393. URL https://books.google.se/books?id=Srpq9Hoj4eMC.

Ioannis Karatzas and Hui Wang. Utility maximization with discretionary stopping. *SIAM J. Control and Optimization*, 39:306–329, 08 2000. doi: 10.1137/S0363012998346323.

Ioannis Karatzas, John P. Lehoczky, Suresh P. Sethi, and Steven E. Shreve. Explicit solution of a general consumption/investment problem. *Mathematics of Operations Research*, 11(2):261–294, 1986. ISSN 0364765X, 15265471. URL http://www.jstor.org/stable/3689808.

Hyeng Keun Koo. Consumption and portfolio selection with labor income: A continuous time approach. *Mathematical Finance*, 8(1):49–65, 1998. doi: https://doi.org/10.1111/1467-9965.00044. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9965.00044`.

Robert C. Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 51(3):247–257, 1969. ISSN 00346535, 15309142. URL `http://www.jstor.org/stable/1926560`.

Paul A. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economics and Statistics*, 51(3):239–246, 1969. ISSN 00346535, 15309142. URL `http://www.jstor.org/stable/1926559`.

Nancy L. Stokey, Robert E. Lucas, and Edward C. Prescott. *Recursive Methods in Economic Dynamics*. Harvard University Press, 1989. ISBN 9780674750968. URL `http://www.jstor.org/stable/j.ctvjnrt76`.

Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*, volume 82. 01 2000. doi: 10.1007/978-3-662-03185-8.

A Proofs and auxiliary results

Remark 2. *Let f be a strictly concave, increasing function over $(0, \infty)$. Then its convex conjugate $\widetilde{f}$*

(i) is a strictly convex, decreasing function, and

(ii) the dual relation

$$f(x) = \inf_y \{\widetilde{f}(y) + xy\}$$

holds. In addition, for a given strictly convex, decreasing function $\widetilde{f}$, the f defined by this expression would be strictly concave and increasing.

In addition, if f is continuously differentiable and the range of its derivative f' is $(0, \infty)$, then

(iii)

$$\widetilde{f}(y) = f\big(I(y)\big) - yI(y)$$

where I denotes the inverse function of f'.

(iv) Furthermore, $\widetilde{f}$ is continuously differentiable and

$$f'(x) = \widetilde{I}'(-x)$$

and

$$\widetilde{f}'(y) = -I'(y)$$

holds for all x and y, where $\widetilde{I}$ denotes the inverse function of $\widetilde{f}'$.

Finally, assume that $\{f_i\}$ is a finite set of functions satisfying the same assumptions as f so far. Define their pointwise maximum F with

$$F(x) = \max_i f_i(x).$$

Then

(v)

$$\widetilde{F}(y) = \max_i \widetilde{f}_i(y)$$

and

(vi) *$\widetilde{F}$ is differentiable at every y such that $\arg\max_i \widetilde{f}_i(y)$ has only one element. In particular,*

$$\widetilde{F}'(y) = \widetilde{f}'_{\arg\max \widetilde{f}_i(y)}(y)$$

for all such y.

Proof of Lemma 1. This proof is essentially a slightly more detailed replication of a section in the Appendix of Farhi and Panageas (2007). By the definition of convex conjugates

$$\begin{aligned}
J(a_0, c_t, \xi_t, \tau) =& \mathbb{E}\left[\int_0^\tau e^{-\rho t} u(c_t)\,\mathrm{d}t + e^{-\rho\tau} U(a_\tau)\right] \\
\leq& \mathbb{E}\left[\int_0^\tau e^{-\rho t} \widetilde{u}(\lambda e^{\rho t} X_t H_t)\,\mathrm{d}t + e^{-\rho\tau} \widetilde{U}(\lambda e^{\rho\tau} X_\tau H_\tau)\right. \\
&\left. + \lambda\Big(\int_0^\tau c_t X_t H_t\,\mathrm{d}t + a_\tau X_\tau H_\tau\Big)\right]
\end{aligned}$$

After a slight reorganization apply integration by parts on the product of X_t and $\mathbb{E}_t[\int_t^\tau H_s(c_s - y)\,\mathrm{d}s + H_\tau a_\tau]$. We use that $X_0 = 1$ and that the quadratic covariation term is 0, since X is a process with bounded total variance.

$$\begin{aligned}
\int_0^\tau c_t X_t H_t\,\mathrm{d}t + a_\tau X_\tau H_\tau &= \int_0^\tau y X_t H_t\,\mathrm{d}t + a_\tau X_\tau H_\tau + \int_0^\tau (c_t - y) X_t H_t\,\mathrm{d}t \\
&= \int_0^\tau y X_t H_t\,\mathrm{d}t + \mathbb{E}\big[\int_0^\tau H_t(c_t - y)\,\mathrm{d}t + H_\tau a_\tau\big] \\
&\quad + \int_0^\tau \mathbb{E}_t\big[\int_t^\tau H_s(c_s - y)\,\mathrm{d}s + H_\tau a_\tau\big]\,\mathrm{d}X_t
\end{aligned}$$

To proceed we need two inequalities involving the discounted value of future net spending. The consolidated budget constraint evaluated between time t and stopping time τ is

$$\int_t^\tau H_s(c_s - y)\,\mathrm{d}s + H_\tau a_\tau = H_t a_t + \int_t^\tau H_s\big[\sigma\xi_s - \kappa a_s\big]\,\mathrm{d}a_s \tag{62}$$

the last term is a supermartingale (being a local martingale with a lower bound) so with $t = 0$

$$\mathbb{E}\left[\int_0^\tau H_s(c_s - y)\,\mathrm{d}s + H_\tau a_\tau\right] \leq a_0 \tag{63}$$

i.e. the present value of future net spending has to be lower than current wealth. On the other hand from

$$0 \leq H_t a_t = \int_t^\tau H_s(c_s - y)\,\mathrm{d}s + H_\tau a_\tau - \int_t^\tau H_s\big[\sigma\xi_s - \kappa a_s\big]\,\mathrm{d}a_s$$

we get

$$0 \leq \mathbb{E}\left[\int_t^\tau H_s(c_s - y)\,\mathrm{d}s + H_\tau a_\tau\right]$$

Putting things together:

$$\begin{aligned}
J(a_0, c_t, \xi_t, \tau) =& \mathbb{E}\left[\int_0^\tau e^{-\rho t} u(c_t)\,\mathrm{d}t + e^{-\rho\tau} U(a_\tau)\right] \leq \\
& \mathbb{E}\Bigg[\int_0^\tau e^{-\rho t}\widetilde{u}(\lambda e^{\rho t} X_t H_t)\,\mathrm{d}t + e^{-\rho\tau}\widetilde{U}(\lambda e^{\rho\tau} X_\tau H_\tau) + \lambda\int_0^\tau y X_t H_t\,\mathrm{d}t \\
& + \lambda\int_0^\tau H_t(c_t - y)\,\mathrm{d}t + H_\tau a_\tau \\
& + \lambda\int_0^\tau \mathbb{E}_t\Big[\int_t^\tau H_s(c_s - y)\,\mathrm{d}s + H_\tau a_\tau\Big]\,\mathrm{d}X_t\Bigg] \leq \\
& \mathbb{E}\Bigg[\int_0^\tau e^{-\rho t}\widetilde{u}(\lambda e^{\rho t} X_t H_t)\,\mathrm{d}t + e^{-\rho\tau}\widetilde{U}(\lambda e^{\rho\tau} X_\tau H_\tau) + \lambda\int_0^\tau y X_t H_t\,\mathrm{d}t \\
& + \lambda a_0 + 0\Bigg]
\end{aligned}$$

and it is apparent that equality follows under conditions (11)-(14), where for (12) point (iii) or Remark 2 was used. In particular, Equation (14) is sufficient for equality in the last line since if the budget constraint is exhausted until τ it has to be exhausted in all subintervals as well. $\square$

Proof of Proposition 2. The first task is to show that given process X, τ_D solves the optimal stopping problem for all λ and that the corresponding $\tilde{J}$ coincides with $\widehat{V}$. The proof of this part is omitted, since it is a straightforward modification of standard optimal stopping and stochastic control results (see for example Theorems 10.4.1 and 11.2.2 in Øksendal (2000) for the cases when $\tau_D < \infty$ and $\tau_D = \infty$, respectively). The only difference is the presence of the additional object X_t and the related condition (iii) making sure that the only term containing X_t drops out. The rest of the proof broadly follows Theorem 4. in He and Pagès (1993), apart from the stopping time and is reported below.

First assume that $\lambda \notin D$. In that case it is optimal to stop immediately in the dual problem, and therefore $\widehat{V}(\lambda) = \widetilde{U}(\lambda)$ and also $\widehat{V}'(\lambda) = \widetilde{U}'(\lambda)$. Let us consider stopping immediately in the primal problem as well, which gives us value $U(a_0)$. By the definition of $\widetilde{U}$ and (20) the chain of inequalities in (16) holds with equality, hence $\tau = 0$ must indeed optimal in the primal problem as well.

Now assume that $\lambda \in D$ and therefore Z^X starts in the continuation region. By the assumptions of Proposition 2, $\widehat{V}$ is continuously differentiable and hence the second inequality holds as an equality. To show the same for the first inequality we rely on Lemma 1. In particular, we have to show that the proposed policies are admissible and equations (11)-(14) are satisfied.

First note that the smoothness conditions on $\widehat{V}$ imply the integrability conditions on c_t and ξ_t and both processes are adapted. Second, since the right hand side of (24) is constant over D, its derivative by Z has to be 0, implying

$$\begin{aligned}0 =& -\rho\widehat{V}'(Z) + \widehat{V}'(Z)\Big(\rho - r\Big) + \widehat{V}''(Z)Z\Big(\rho - r\Big) + \widehat{V}'''(Z)Z^2\theta \\ &+ 2\widehat{V}''(Z)Z\theta + \widetilde{u}'(Z) + y \\ =& -r\widehat{V}'(Z) + \widehat{V}''(Z)Z\Big(\rho - r + 2\theta\Big) + \widehat{V}'''(Z)Z^2\theta - c_t + y \end{aligned} \tag{64}$$

Applying Ito's Lemma, and substituting in equation (64) we get

$$\begin{aligned}&c_t\,\mathrm{d}t - y\,\mathrm{d}t - \mathrm{d}\widehat{V}'(Z_t^X) \\ &= c_t\,\mathrm{d}t - y\,\mathrm{d}t - \Big((\rho - r - \psi_t)Z_t^X\widehat{V}''(Z_t^X) + \theta(Z_t^X)^2\widehat{V}'''(Z_t^X)\Big)\,\mathrm{d}t + \kappa Z_t^X\widehat{V}''(Z_t^X)\,\mathrm{d}W_t \\ &= \Big(-r\widehat{V}'(Z_t^X) + 2\theta Z_t^X\widehat{V}''(Z_t^X)\Big)\,\mathrm{d}t + \kappa Z_t^X\widehat{V}''(Z_t^X)\,\mathrm{d}W_t \\ &= \Big(ra_t + \mu\xi_t\Big)\,\mathrm{d}t + \sigma\xi_t\,\mathrm{d}W_t\end{aligned}$$

where we used $\mathrm{d}X_t\widehat{V}''(Z_t^X) = 0$, which follows from $\mathrm{d}X_t\widehat{V}'(Z_t^X) = 0$ and the fact that $\widehat{V}$ is decreasing and convex. This implies the budget constraint

$$\int_0^t c_s\,\mathrm{d}s + a_t = a_0 + \int_0^t \Big(y + ra_s + \mu\xi_s\Big)\,\mathrm{d}s + \sigma\int_0^t \xi_s\,\mathrm{d}s$$

for all $t < \tau_D$ and thus admissibility. It is standard to show that the latter result also implies the consolidated budget constraint (13).

Next note that (11) follows from (iv) in Remark 2 by $a_\tau = -\widehat{V}'(Z_\tau) = -\widetilde{U}'(Z_\tau)$. Equation (12) was assumed. Finally, (14) follows from (iv) in Proposition 2 and the budget constraint $a_t = -\widehat{V}'(Z_t^X) \geq 0$ holds as $\widehat{V}$ is a decreasing function.

$\square$

B Parameters used in Section 5

γ	r	μ	σ	ρ	y	R	P	β
3	0.03	0.05	0.14	0.05	1	0.7	17.5	0.85

Table 1: Parameters used for the example in Section 5. R and P are both chosen approximately twice as large as empirically plausible values, to make effects more visible.