# Cautious Stochastic Choice, Optimal Stopping and Deliberate Randomization*

Vicky Henderson David Hobson Matthew Zeng
University of Warwick

January 23, 2018

We study Cautious Stochastic Choice (CSC) agents facing optimal timing decisions in a dynamic setting. In an expected utility setting, the optimal strategy is always a threshold strategy - to stop/sell the first time the price process exits an interval. In contrast, we show that in the CSC setting, where the agent has a family of utility functions and is concerned with the worst case certainty equivalent, the optimal strategy may be of non-threshold form and may involve randomization. Our model is consistent with recent experimental evidence in dynamic setups whereby individuals do not play cut-off or threshold strategies.

**Keywords:** Stochastic choice, cautious stochastic choice, randomization, optimal stopping.

**JEL Classification:** D81, G19, G39.

## 1 Introduction

It is well recognized that individual decision making is not fully captured by expected utility theory and many non-expected utility theories have been developed with the aim of providing a better fit to observed behavior. Many of these alternative theories have been well studied in a static setting, but recently there has been much interest in studying non-expected utility preferences in dynamic settings which describe timing problems arising in real world decisions. Examples of theoretical work in this vein include Ebert and Strack (2015) and in experimental

*University of Warwick, Coventry, CV4 7AL. UK. Email: vicky.henderson@warwick.ac.uk, d.hobson@warwick.ac.uk, m.zeng@warwick.ac.uk. We would like to thank participants at the 10th Oxford-Princeton workshop (May 25-26, 2017) and at the Stochastics of Financial Markets Seminar, HU/TU Berlin (9th November, 2017) for helpful comments. Matthew Zeng is supported by a Chancellor's International Scholarship at the University of Warwick.

settings, Oprea et al (2009). Our paper considers agents who face optimal timing decisions in a dynamic setting and who exhibit cautious stochastic choice (CSC). The CSC agent is unsure which utility function to use from a family of possibilities and applies caution to choose the worst-case certainty equivalent. Our main result is that in this optimal stopping setup, a CSC agent may have an optimal strategy which is not of threshold form and may involve randomization. The dynamic CSC model gives predictions which are consistent with recent experimental evidence in dynamic setups whereby individuals do not play cut-off or threshold strategies (Strack and Viefers (2017), Fischbacher, Hoffmann and Schudy (2015)).

Cerreia-Vioglio et al (2015, 2017) (see also Maccheroni (2002)) develop a theory of cautious stochastic choice in a static decision making setting. The agent aims to select a best lottery from a given set. Under CSC the agent has a family of possible utility functions in mind. For a given lottery, and for each utility, the agent computes the certainty equivalent. The agent then values the lottery via the worst-case certainty equivalent. Finally the agent chooses the best lottery which maximizes this value. Since CSC does not satisfy the quasi-convexity property, agents may benefit from mixing (see Cerreia-Vioglio et al (2017)). Our goal in this paper is to ask if these results from a static setting also apply in an optimal stopping problem. We want to understand if CSC agents also seek to randomize in a dynamic setting.

Stopping problems are used in finance, economics and statistics to model dynamic decision making tasks including asset purchase and sales, option pricing, market entry and exit, irreversible investment and job search. In this paper we consider a continuous time optimal stopping model for the sale of an asset in which the price process is given by a one-dimensional time-homogeneous diffusion. If the agent were an expected utility maximizer, it is well known that the optimal stopping rule is given by the first exit time of the price process from an interval, ie. a pure threshold strategy. We formulate an optimal stopping problem with CSC as follows. The agent has a family of utility functions and for a given stopping rule (in an appropriate class of admissible strategies), for each utility, computes the cerainty equivalent. The worst-case is then taken over utilities. The goal is to find the stopping rule which maximizes the worst-case certainty equivalent value.

By considering a stylized but tractable example, we can show the optimal strategy is not necessarily of threshold form. For this example we can calculate the optimal stopping rule and show that it is a non-trivial mixture of threshold strategies. We then consider two realistic models where the asset price follows exponential Brownian motion. In the first model the family of utilities are $S$-shaped and reference level dependent. The second model uses a family of concave utility functions. These examples show CSC agents do randomize in realistic, dynamic optimal stopping settings.

The paper is organised as follows. In the next section we briefly review some of the relevant literature on stochastic choice. Section 3 presents the optimal stopping models - both the classical EU model and our CSC optimal stopping model. Our stylized example is given in Section 4. Section 5 describes and solves two models with $S$-shaped reference dependent or concave families of utilities. We defer supplementary material and proofs to the Appendices. Appendix A outlines the CSC model in its original static setup (Cerreia-Vioglio et al (2015, 2017)) and demonstrates mixing may be beneficial. Further results and proofs on optimal stopping under EU are in Appendix B. Proofs for the stylized and generalized example are in Appendices C and D.

## 2 Literature

A consistent finding in experimental studies of individual decision making is the phenomenon of stochastic or random choice. When subjects are asked to choose from the same set of options many times, they are inconsistent in their choices. Patterns of stochastic choice were first recorded by Tversky (1969) and many studies have replicated, explored and extended his results (see Agranov and Ortoleva (2017) for recent findings and a comprehensive overview, and amongst others Dwenger al al (2013), Hey and Orme (1994)). In particular, recent studies of Agranov and Ortoleva (2017) and Dwenger er al (2013) interpret their experimental results as suggesting the main force is a *deliberate* desire of participants to randomize. Much of this evidence is gathered in static settings. Recently, researchers have studied *dynamic* settings which can better reflect the real decision making situations individuals face in economics and finance (eg. Oprea et al (2009)). Strack and Viefers (2017) conduct an experiment in a sophisticated asset selling task whereby subjects played sixty-five rounds during which they could sell their stock. In each round they observe a path of the market price which follows a random walk with positive expected return. Strack and Viefers (2017) present evidence that players do not play cut-off or threshold strategies over gains - they do not behave time-consistently within rounds 75% of the time, and visit the same price level three times on average before stopping at it. In their study of the impact of automatic selling devices on experimental trading behavior, Fischbacher, Hoffman and Schudy (2015) find that participants tend to set any upper limit further away from the current price than any lower limits and use the upper limit less frequently. In contrast to the behavior of an EU agent, our CSC agent does not only use pure threshold strategies and instead prefers mixed or randomized strategies, consistent with this body of recent evidence. The CSC agent is deliberately randomizing, which is again, in line with the recent experimental findings as described above.

There is a large body of theoretical models which were developed to capture the phenomenon of stochastic or random choice. CSC falls into the class of stochastic models postulating that stochasticity is a *deliberate choice* of the agent.[1] Deliberate randomization (Machina (1985)) emerges in non-EU settings such as prospect theory (see Wakker (1994) in a static setting, and Henderson, Hobson and Tse (2017) and He et al (2017) in dynamic setups). There are fewer models capturing the phenomenon of stochastic choice in the dynamic setting of a stopping problem. Strack and Viefers (2017) combine random utility with regret preferences in a stopping context. Henderson, Hobson and Tse (2017) and He et al (2017) show randomized strategies are optimal in a stopping model with prospect theory preferences. The largest class of stopping models are the bounded rationality Drift Diffusion models (DDM) of which the work of Fudenberg, Strack and Strzalecki (2017) is a recent example. Our CSC model contributes a new dynamic optimal stopping model to this class of stochastic choice models in the literature.

## 3 The Optimal stopping models

We first establish notation and review the theory for the optimal liquidation of an asset in the classical setting of a maximizer of expected utility. For $J$ an interval subset of $\mathbb{R}$, let $F_{\uparrow}^{J}$ be the set of increasing functions $F_{\uparrow}^{J} = \{f : J \mapsto \mathbb{R}; f \text{ increasing}\}$. For $f \in F_{\uparrow}^{J}$ we can define the left-continuous inverse $f^{-1}$. For $K$ a subset of $\mathbb{R}^d$ let $\mathcal{P}(K)$ be the set of Borel probability measures on $K$. Let $\mathcal{L}(Y)$ denote the law of a random variable $Y$. If $Z = (Z_t)_{t \geq 0}$ is a stochastic process and $\mathcal{S}$ is a class of stopping times then let $Q^Z(\mathcal{S}) = \{\mathcal{L}(Z_\tau); \tau \in \mathcal{S}\}$. Let $\delta_z$ be the point mass at $z$.

We work on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$. Let $Y = (Y_t)_{t \geq 0}$ be a $(\mathbb{F}, \mathbb{P})$-stochastic process on this probability space. Let $I^Y$ be the state space of $Y$ and let $\bar{I}^Y$ be the

[1] There are two other main classes of models of stochastic choice. In *random utility* models, subjects maximize a well defined utility function but this changes stochastically over time (eg. Gul and Pesendorfer (2006)). In models of *bounded rationality*, agents have well defined and stable preferences but may not make the best choice because of bounded rationality (see Johnson and Ratcliff (2013) for a review).

closure of $I^Y$. We suppose that $Y$ is a regular, time-homogeneous diffusion with initial value $Y_0 = y$ which lies in the interior of $I^Y$. Further we suppose that $\lim_{t\uparrow\infty} Y_t$ exists. A sufficient condition for this is Assumption 1 below. $Y$ represents the price process of the asset.

### 3.1 Optimal stopping under expected utility

Let $U$ be an increasing utility function, $U \in F_{\uparrow}^{\bar{I}^Y}$. For a maximizer of expected utility the objective is to find the certainty equivalent

$$C^{EU}(\mathcal{S}) = \sup_{\tau\in\mathcal{S}} U^{-1}(\mathbb{E}^y[U(Y_\tau)]) = U^{-1}\left(\sup_{\tau\in\mathcal{S}} \mathbb{E}^y[U(Y_\tau)]\right) \tag{1}$$

over a suitable class $\mathcal{S}$ of stopping times. We introduce three classes of stopping times

- $\mathcal{T}$, the class of all stopping times;
- $\mathcal{T}_T$, the class of (pure) threshold stopping times;
- $\mathcal{T}_R$, the class of randomized threshold stopping times.

Note that $\mathcal{T}_T \subset \mathcal{T}_R \subset \mathcal{T}$. The set of pure threshold stopping times includes stopping immediately and can be written as

$$\mathcal{T}_T = \cup_{(\beta,\gamma)\in\mathcal{D}}\{\tau^Y_{\beta,\gamma}\}, \tag{2}$$

where $\tau^Y_{\beta,\gamma} = \inf_{u\geq 0}\{u : Y_u \notin (\beta,\gamma)\}$ and the union is taken over $(\beta,\gamma)$ in an appropriate set $\mathcal{D} \subseteq ([-\infty, y] \cap \bar{I}^Y) \times ([y,\infty] \cap \bar{I}^Y)$ which we describe below.

In order to able to define the set of randomized threshold stopping times $\mathcal{T}_R$ we suppose that $\mathcal{F}_0$ is rich enough as to support any probability measure $\eta$ on $\mathcal{D}$, and that the dynamics of $Y$ are independent of $\eta$. Then we define a randomized stopping time $\tau_\eta$ by

$$\tau_\eta = \inf_{u\geq 0}\{u : Y_u \notin (\Theta_\beta, \Theta_\gamma) \text{ where } \Theta = (\Theta_\beta, \Theta_\gamma) \text{ is } \mathcal{F}_0 \text{ measurable and has law } \eta\}$$

and set

$$\mathcal{T}_R = \{\tau_\eta; \eta \in \mathcal{P}(\mathcal{D})\}. \tag{3}$$

Often, the best way to solve (1) is via a change of scale. Let $s$ be a strictly increasing function such that $X = s(Y)$ is a local martingale. (Such a function $s$ exists under very mild conditions on $Y$, and is called a scale function. For example, if $Y$ solves the SDE $dY_t = \sigma(Y_t)dB_t + \mu(Y_t)dt$ then $s = s(z)$ is a solution to $\frac{1}{2}\sigma(z)^2 s'' + \mu(z)s' = 0$. Note that if $s$ is a scale function then so is any affine transformation of $s$ and so we may chose any convenient normalization for $s$.) Then $U(Y_\tau) = g(X_\tau)$ where $g = U \circ s^{-1}$ and (1) can be rewritten as

$$C^{EU}(\mathcal{S}) = \sup_{\tau \in \mathcal{S}} U^{-1}(\mathbb{E}^x[g(X_\tau)]) = s^{-1}\left(\sup_{\tau \in \mathcal{S}} g^{-1}(\mathbb{E}^x[g(X_\tau)])\right) \tag{4}$$

where $x = s(y)$. Since the scale function $s$ is fixed, in finding the optimal stopping rule it is sufficient to consider $\sup_{\tau \in \mathcal{S}} g^{-1}(\mathbb{E}^x[g(X_\tau)])$.

We do not make a concavity assumption on $U$. Monotonicity is preserved under the transformation $U \mapsto g$, but in general concavity is not. Indeed, if $g$ is concave then typically stopping immediately ($\tau = 0$) is optimal.

The state space of $X$ is $I^X = s(I^Y)$. Then $\bar{I}^X = s(\bar{I}^Y)$. If $I^X$ is not bounded below then for any level $\gamma$ in the interior of $I^X$ with $\gamma \geq x$ the first hitting time $H^X_\gamma = \inf_{u \geq 0}\{u : X_u = \gamma\}$ is finite almost surely and $C^{EU}(\mathcal{T}) = \sup_{\gamma \in I^X} U^{-1} g(\gamma) = \sup\{\gamma : \gamma \in I^Y\} = \max\{\gamma : \gamma \in \bar{I}^Y\}$. We want to exclude this degenerate case. Hence we make the following assumption:

**Assumption 1.** *$I^X = s(I^Y)$ is bounded below. Then, without loss of generality we may assume that the lower limit of $I^X$ is zero. Any accessible boundary point for $X$ is absorbing.*

The upper limit of $I^X$ may be finite or infinite. Note that since $X$ is a non-negative local martingale $\lim_{t\uparrow\infty} X_t$ exists and hence $\lim_{t\uparrow\infty} Y_t$ exists. We do not exclude $\tau$ such that $\mathbb{P}(\tau = \infty) > 0$ and on the set $\tau = \infty$ we define $X_\tau = \lim_{t\uparrow\infty} X_t$. This is why we want to consider $\bar{I}^X$ as well as $I^X$. Then $\mathcal{T}$ is the set of all stopping times, and not just finite stopping times.

**Example 1.** *Suppose $Y$ is geometric Brownian motion: $dY_t = \sigma Y_t dB_t + \mu Y_t dt$. Let $\psi = 1 - \frac{2\mu}{\sigma^2}$. $Y$ has state space $I^Y = (0, \infty)$. Provided $\psi \neq 0$ we have $s(z) = sgn(\psi) z^\psi$. (If $\psi = 0$ then $s(z) = \ln z$ is the scale function.) If $\psi \leq 0$ then $s(0) = -\infty$. This is equivalent to $2\mu \geq \sigma^2$, in which case $Y$ hits arbitrarily high price levels with probability one and the optimal stopping problem is degenerate. If $\psi > 0$ then $I^X = (0, \infty)$. For $\psi > 0$, $\lim_{t\to\infty} X_t = 0 \in \bar{I}^X \setminus I^X$.*

Note that $\tau^Y_{\beta,\gamma} = \inf_{u \geq 0}\{u : Y_u \notin (\beta, \gamma)\} = \inf_{u\geq 0}\{u : X_u \notin (s(\beta), s(\gamma))\} =: \tau^X_{s(\beta),s(\gamma)}$. Hence $\mathcal{T}_T$ has the alternative representation

$$\mathcal{T}_T = \cup_{(a,b)\in\mathcal{D}^X}\{\tau^X_{a,b}\},$$

for an appropriate set $\mathcal{D}^X$. The right space to choose is $\mathcal{D}^X = [0, x) \times ([x, \infty] \cap \bar{I}^X)$. Then $\mathcal{D}$ in (2) and (3) is given by

$$\mathcal{D} = [s^{-1}(0), y) \times [y, s^{-1}(\infty)]$$

$\mathcal{T}_R$ can also be rewritten as $\mathcal{T}_R = \{\tau^X_\eta : \eta \in \mathcal{P}(\mathcal{D}^X)\}$ where

$$\tau^X_\eta = \inf_{u\geq 0}\{u : X_u \notin (\Theta_\beta, \Theta_\gamma)\text{where } \Theta = (\Theta_\beta, \Theta_\gamma) \text{ has law } \eta.\}$$

Note that the certainty equivalent depends only on the law of $X_\tau$. The following result is classical. (In discrete time see Karni and Safra (1990), in mathematical finance see Dayanik and Karatzas (2003) and recently in discrete time in Strack and Viefers (2017). For a textbook treatment, see Chapter 4, Peskir and Shiryaev (2006).)

**Proposition 1.** *1. $C^{EU}(\mathcal{T}_T) = C^{EU}(\mathcal{T}_R) = C^{EU}(\mathcal{T})$.*

*2. $C^{EU}(\mathcal{T}) = U^{-1}(g^{cv}(s(y)))$ where $g^{cv}$ is the smallest concave majorant of $g = U \circ s^{-1}$.*

**Corollary 1.** *In trying to find the optimal stopping rule in the classical (single utility) case it is sufficient to restrict attention to pure threshold strategies of the form $\tau = \tau^Y_{a,b}$.*

One approach to proving Proposition 1 is to show first that the problem can be recast as one involving the process in natural scale $X$, and then that the problem of maximizing over stopping times can be recast as a maximization over distributions. In particular, we see from (1) or (4) that the certainty equivalent depends on $\tau$ only through the law of the stopped process. Hence, instead of searching over stopping rules we can search over laws of the stopped process instead. In terms of maximizing expected utility of the stopped process, it can be shown that the optimal law places mass on at most two points. Such a distribution can be achieved using a pure threshold rule. This explains why $C^{EU}(\mathcal{T}_T) = C^{EU}(\mathcal{T})$ and the more general result of the first part of the Proposition follows since clearly $C^{EU}(\mathcal{T}_T) \leq C^{EU}(\mathcal{T}_R) \leq C^{EU}(\mathcal{T})$.

### 3.2 Optimal stopping under Cautious Stochastic Choice

Our goal in this section is to develop an optimal stopping model with CSC. Let $Y$ be a time-homogeneous diffusion with state space $I^Y$. Let $\mathcal{W}^Y \subseteq F^{\bar{I}^Y}_{\uparrow}$ be a set of increasing utility functions. The goal is to find $\sup_{\tau\in\mathcal{S}} \inf_{u\in\mathcal{W}^Y} u^{-1}(\mathbb{E}[u(Y_\tau)])$, where $\tau$ is chosen from a suitable set of stopping times $\mathcal{S}$. We define $\Delta^Y = \mathcal{P}(\bar{I}^Y)$. Recall that $Q^Y(\mathcal{S}) = \{\nu : \nu = \mathcal{L}(Y_\tau); \tau \in \mathcal{S}\}$. Clearly we have $Q^Y(\mathcal{T}_T) \subset Q^Y(\mathcal{T}_R) \subseteq Q^Y(\mathcal{T}) \subseteq \Delta^Y$.

As in the classical, single-utility setting, it is often convenient to work with the process $X$ in natural scale rather than $Y$. We set $\mathcal{W}^X = \{g = u \circ s^{-1}; u \in \mathcal{W}^Y\}$. Define $\Delta^X = \mathcal{P}(\bar{I}^X)$. Again we have $Q^X(\mathcal{T}_T) \subset Q^X(\mathcal{T}_R) \subseteq Q^X(\mathcal{T}) \subseteq \Delta^X$.

For a fixed stopping time $\tau$ and a fixed utility $u \in \mathcal{W}$ we define the certainty equivalent

$$C^u_\tau = u^{-1}(\mathbb{E}[u(Y_\tau)]) = u^{-1}(\mathbb{E}[g(X_\tau)]) = s^{-1}(g^{-1}(\mathbb{E}[g(X_\tau)])). \tag{5}$$

Once we have minimized over utilities the value function for a single stopping time is $V_\tau = \inf_{u \in \mathcal{W}^Y} C^u_\tau$. Under CSC the optimal stopping problem is to find $V(\mathcal{S}) = \sup_{\tau \in \mathcal{S}} V_\tau$ where $\mathcal{S}$ is a set of stopping times. Since $V_\tau$ depends on the stopping time only through the law of the stopped process we have

$$V(\mathcal{S}) = \sup_{\nu \in Q^Y(\mathcal{S})} \inf_{u \in \mathcal{W}^Y} u^{-1}\left(\int u(z)\nu(dz)\right) = s^{-1}\left(\sup_{\nu \in Q^X(\mathcal{S})} \inf_{g \in \mathcal{W}^X} g^{-1}\left(\int g(z)\nu(dz)\right)\right) \tag{6}$$

and $\tau^* \in \mathrm{argmax}_{\tau \in \mathcal{S}} V_\tau$. In particular, we want to consider $\mathcal{S} = \mathcal{T}$, $\mathcal{S} = \mathcal{T}_R$ and $\mathcal{S} = \mathcal{T}_T$.

The following result is key to solving (6):

**Proposition 2** (Henderson, Hobson and Zeng (2017))**.**

*1.* $Q^Y(\mathcal{T}_T) \subset Q^Y(\mathcal{T}_R) = Q^Y(\mathcal{T})$

*2.* $V(\mathcal{T}_T) \leq V(\mathcal{T}_R) = V(\mathcal{T})$*.*

The idea behind the proof of the Proposition in Henderson, Hobson and Zeng (2017) is to characterize $Q^X(\mathcal{T})$ and show that any element of $Q^X(\mathcal{T})$ can be realized as the law of $X_\tau$ using a stopping rule which is a mixture of threshold rules. The result for the process in natural scale translates to a result for $Y$.

The second part of the proposition follows immediately from the first. From our perspective, the content of Proposition 2 is that $V(\mathcal{T}_R) = V(\mathcal{T})$. The first main result of this paper, is to show that, unlike in the classical case (see Proposition 1(1)), we may have $V(\mathcal{T}_T) < V(\mathcal{T}_R)$.

# 4 A stylized example

Our goal in this section is to give an example for which we can prove that the optimal stopping rule is not a pure threshold strategy. Instead there is an optimal stopping rule which is a non-trivial mixture of threshold stopping rules. The example is highly stylized, and deliberately simple, and this allows us to give a full and complete solution, ie. we are able to solve for the optimal mixed threshold rule. Crucially, as we will see in the next section, the characteristic features are shared with some realistic, non-stylized examples.

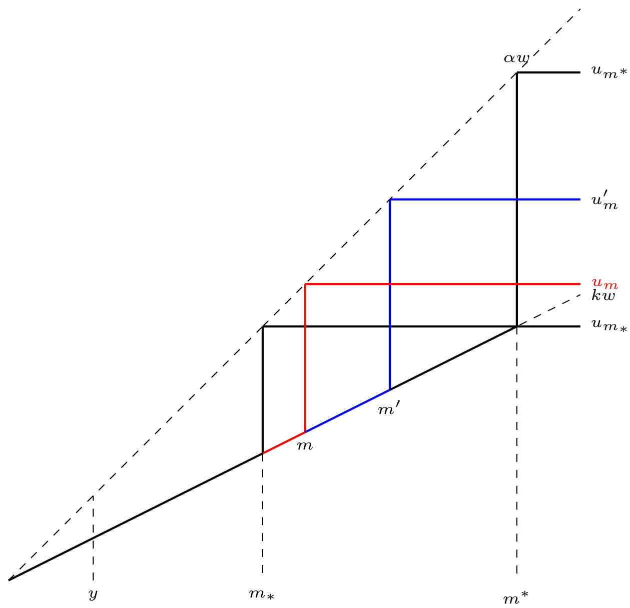



Figure 1: The family of utility functions $\mathcal{W} = \{u_m; m \in \mathcal{M}\}$, where $u_m$ is defined for $m \in \mathcal{M}$ by $u_m(w) = kw, 0 \leq w < m$ and $u_m(w) = \alpha m, w \geq m$ for constants $\alpha > k > 0$. Let $\mathcal{M} = [m_*, m^*]$ with $m_* > \frac{\alpha y}{k}$ and $m^* = \frac{\alpha m_*}{k}$.

We work with a process $Y$ which is already in natural scale, and a family of payoff functions $\{u_m\}_{m \in \mathcal{M}}$. The process $Y$ is assumed to be bounded below (without loss of generality by zero) and unbounded above, to be a local martingale and to have initial value $y > 0$. Then $Y$ is a supermartingale. The canonical example is if $Y$ is a Brownian motion started at $y$ and absorbed at zero. Alternatively, we may consider $Y$ to be geometric Brownian motion with zero drift. The goal in this section is to give an example for which

$$V(\mathcal{T}_T) < V(\mathcal{T}_R) = V(\mathcal{T}).$$

Hence, there is no pure threshold strategy which is optimal within class of all stopping rules.

Fix constants $\alpha > k > 0$, together with $m_* > \frac{\alpha y}{k}$. Set $m^* = \frac{\alpha m_*}{k}$ and let $\mathcal{M}$ be the interval $\mathcal{M} = [m_*, m^*]$. For $m \in \mathcal{M}$ define $u_m : I \equiv [0, \infty) \mapsto [0, \infty)$ by

$$u_m(w) = \begin{cases} kw, & 0 \leq w < m; \\ \alpha m, & w \geq m; \end{cases} \tag{7}$$

and set $\mathcal{W} = \{u_m; m \in \mathcal{M}\}$. Figure 1 illustrates the family of utilities described here.

Note that the results generalize to utility functions which replace $u_m(w) = \alpha m$ with $u_m(w) =$

$J(m)$ for $w \geq m$ in (7), where $J$ is a strictly increasing function with $J(m) > km$. We will consider this more general case in Section 4.4.

### 4.1 Pure threshold strategies

Our first result is that in the stylized example there is no pure threshold strategy which outperforms the trivial strategy of stopping immediately. Note that since $Y$ is a supermartingale and since $u_m(z) \leq \alpha z$, we have for all $\tau \in \mathcal{T}$

$$\mathbb{E}[u_m(Y_\tau)] \leq \alpha \mathbb{E}[Y_\tau] \leq \alpha y < k m_*.$$

Since $u_m^{-1}(w) = \frac{w}{k}$ for $w < km_*$ we have for any $\tau \in \mathcal{T}$

$$u_m^{-1}(\mathbb{E}[u_m(Y_\tau)]) = \frac{1}{k}\mathbb{E}[u_m(Y_\tau)]. \tag{8}$$

Recall $\tau_{\beta,\gamma} = \inf\{s : Y_s \notin (\beta, \gamma)\}$. For $m \in \mathcal{M}$ and $0 \leq \beta \leq y \leq \gamma$ let $\mathcal{G}^m_{\beta,\gamma}$ be the expected utility associated with the stopping time $\tau_{\beta,\gamma}$ and the utility function $u_m$ and let $C^m_{\beta,\gamma}$ be the certainty equivalent: we have $\mathcal{G}^m_{\beta,\gamma} = \mathbb{E}[u_m(Y_{\tau_{\beta,\gamma}})]$ and $C^m_{\beta,\gamma} = (u_m)^{-1}(\mathbb{E}[u_m(Y_{\tau_{\beta,\gamma}})]) = \frac{1}{k}\mathcal{G}^m_{\beta,\gamma}$. Then

$$\mathcal{G}^m_{\beta,\gamma} = \begin{cases} ky & \gamma \in [y,m) \\ \left(\alpha m \frac{y-\beta}{\gamma-\beta} + k\beta\frac{\gamma-y}{\gamma-\beta}\right) & \gamma \geq m \end{cases} \qquad C^m_{\beta,\gamma} = \begin{cases} y & \gamma \in [y,m) \\ \frac{1}{k}\left(\alpha m \frac{y-\beta}{\gamma-\beta} + k\beta\frac{\gamma-y}{\gamma-\beta}\right) & \gamma \geq m. \end{cases} \tag{9}$$

Note that for each $m \in \mathcal{M}$, $\mathcal{G}^m_{\beta,\gamma}$ and $C^m_{\beta,\gamma}$ are non-increasing in $\gamma$ for $\gamma \geq m^*$. Also, for each $m \in \mathcal{M}$, and $\gamma \leq m^*$, $\mathcal{G}^m_{\beta,\gamma}$ and $C^m_{\beta,\gamma}$ are non-increasing in $\beta$ for $0 \leq \beta \leq y$.

The following theorem follows from the fact that in our stylized example $V(\mathcal{T}_T) = y$. This result is proved in Appendix C. From the perspective of the worst agent, any pure threshold strategy can only generate at best a certainty equivalent which is the same as the certainty equivalent from selling the asset immediately.

**Theorem 1.** *In our stylized example no pure threshold strategies outperforms stopping immediately.*

### 4.2 Improvement with Randomization between Two Upper Thresholds

The goal of this section is to show that there are mixtures of threshold strategies which outperform the best pure threshold strategies. In addition we will develop some intuition which we can use to motivate the derivation of the optimal randomized strategy.

The remarks after (9) suggest that it is not sensible to use upper thresholds above $m^*$, and that it is sufficient to only consider lower thresholds which are set to zero. (This result is proved in Lemma 4 in Appendix C.) In this section we consider using stopping rules which are a mixture of $\tau_{0,\gamma}$ and $\tau_{0,\epsilon}$ for $m_* \leq \gamma < \epsilon \leq m^*$. If $\tau$ is this mixed stopping rule and $\epsilon < m^*$ then $u_{m^*}(Y_\tau) = kY_\tau$ and the certainty equivalent is equal to $y$. So, if we hope to outperform pure threshold rules we must take $\epsilon = m^*$.

Let $\mathcal{T}_2^0$ be the set of stopping rules obtained from mixing two pure threshold strategies, both with lower threshold 0, and one with upper threshold at $m^*$, and the other with upper threshold in $[m_*, m^*)$. Then $\mathcal{T}_2^0 = \{\tau_\gamma^\theta : \theta \in [0,1], \gamma \in [m_*, m^*)\}$ where $\tau_\gamma^\theta = \tau_{0,\gamma}$ with probability $\theta$ and $\tau_\gamma^\theta = \tau_{0,m^*}$ otherwise. The randomization over $\tau_{0,\gamma}$ and $\tau_{0,m^*}$ takes place at $t = 0$. Set $\mathcal{C}_\gamma^{m,\theta} = u_m^{-1}\left(\mathbb{E}[u_m(Y_{\tau_\gamma^\theta})]\right)$. Then, by the linearity of $u_m^{-1}$ over the relevant domain (recall (8))

$$\begin{aligned}
\mathcal{C}_\gamma^{m,\theta} &= u_m^{-1}\left(\theta\mathbb{E}[u_m(Y_{\tau_{0,\gamma}})] + (1-\theta)\mathbb{E}[u_m(Y_{\tau_{0,m^*}})]\right) \\
&= \frac{\theta}{k}\mathbb{E}[u_m(Y_{\tau_{0,\gamma}})] + \frac{1-\theta}{k}\mathbb{E}[u_m(Y_{\tau_{0,m^*}})] \\
&= \theta C_{0,\gamma}^m + (1-\theta)C_{0,m^*}^m.
\end{aligned}$$

It follows that

$$\mathcal{C}_\gamma^{m,\theta} = \begin{cases} y\left[\theta + (1-\theta)\frac{\alpha m}{km^*}\right] & m_* \leq \gamma < m \leq m^* \\ \frac{y\alpha m}{k}\left[\frac{\theta}{\gamma} + \frac{(1-\theta)}{m^*}\right] & m_* \leq m \leq \gamma \leq m^*. \end{cases}$$

Fix $\gamma \in [m_*, m^*)$ and $\theta$. As a function of $m$, $H_\gamma^\theta(m) = \mathcal{C}_\gamma^{m,\theta}$ is increasing in $m$ on both $[m_*, \gamma]$ and $(\gamma, m^*]$ with a jump down at $\gamma$. It follows that (with $\mathcal{C}_\gamma^{\gamma+,\theta} = \lim_{m\downarrow\gamma}\mathcal{C}_\gamma^{m,\theta}$)

$$\begin{aligned}
\inf_m \mathcal{C}_\gamma^{m,\theta} &= \min\{\mathcal{C}_\gamma^{m_*,\theta}, \mathcal{C}_\gamma^{\gamma+,\theta}\} \\
&= y\min\left\{\frac{\theta m^*}{\gamma} + (1-\theta); \theta + (1-\theta)\frac{\alpha\gamma}{km^*}\right\}. \qquad (10)
\end{aligned}$$

Continuing to fix $\gamma$ but allowing the mixture parameter $\theta$ to vary, the first term in the minimum in (10) is increasing in $\theta$ whereas the second is decreasing in $\theta$. Also, at $\theta = 0$, $\mathcal{C}_\gamma^{m_*,0} = y < \frac{\alpha\gamma}{km^*}y = \mathcal{C}_\gamma^{\gamma+,0}$ and at $\theta = 1$, $\mathcal{C}_\gamma^{m_*,1} = \frac{m^*}{\gamma}y > y = \mathcal{C}_\gamma^{\gamma+,1}$. It follows that $\inf_m \mathcal{C}_\gamma^{m,\theta}$ is maximized over $\theta$ at the value of $\theta$ for which $\mathcal{C}_\gamma^{m_*,\theta} = \mathcal{C}_\gamma^{\gamma+,\theta}$, namely $\theta = \theta^*(\gamma)$ where

$$\theta^*(\gamma) = \frac{\gamma - m_*}{\frac{m_* m^*}{\gamma} + \gamma - 2m_*} \in (0,1).$$

Then

$$\max_{\theta\in[0,1]} \inf_m \mathcal{C}_\gamma^{m,\theta} = y\frac{m^* - m_*}{\frac{m_* m^*}{\gamma} + \gamma - 2m_*}.$$

Finally, we find the maximizer over $\gamma$ is $\gamma = \gamma^*$ where $\gamma^* = \sqrt{m_* m^*}$, and then $\theta^*(\gamma^*) = \frac{1}{2}$ and

$$\max_{\gamma \in \mathcal{M}} \max_{\theta \in [0,1]} \inf_m \mathcal{C}_\gamma^{m,\theta} = \mathcal{C}_{\gamma^*}^{m_*,\theta^*(\gamma^*)} = y \left[ \frac{1 + \sqrt{\frac{m^*}{m_*}}}{2} \right] > y.$$

Hence, $V(\mathcal{T}_2^0) > V(\mathcal{T}_T)$ and *a fortiori* $V(\mathcal{T}) > V(\mathcal{T}_T)$.

**Theorem 2.** *In our stylized example the best strategy outperforms the best pure threshold strategy.*

In addition to the above result, we can learn something from our analysis about the optimal mixture of thresholds. First we expect that there must be a positive probability that we take an upper threshold of $m^*$, else the certainty equivalent associated with $u_{m^*}$ is $y$. Second, by considering the problem for finite mixtures of upper thresholds, we expect that the certainty equivalent associated with $u_m$ should be constant over $m$.

### 4.3 The Optimal Solution

Let $\mathcal{T}_R^0$ be the subset of $\mathcal{T}_R$ such that the lower threshold in the randomization mixture is always at zero, and the upper threshold is in $\mathcal{M}$. Then

$$\mathcal{T}_R^0 = \{\tau_\eta : \eta \in \mathcal{P}(\{0\} \times \mathcal{M})\}. \tag{11}$$

**Proposition 3.** $V(\mathcal{T}_R^0) = V(\mathcal{T}_R) = V(\mathcal{T})$.

Thus, in the stylized example and when considering optimal mixtures of threshold strategies it is sufficient to restrict attention to mixtures in which the lower threshold is always zero and the upper threshold is contained in $\mathcal{M}$. We can calculate the optimal mixed threshold rule. The proof of the proposition and theorem are given in Appendix C.

**Theorem 3.** *Suppose $\tilde{\eta} \in \mathcal{P}(\mathcal{M})$ is a mixture of a point mass at $m^*$ of size $\theta^*$ and an absolutely continuous measure $\rho$ on $(m_*, m^*)$ with density $C^* \gamma^{-\frac{\alpha}{\alpha-k}}$ where*

$$\theta^* = \frac{1}{(\frac{\alpha}{k})^{\frac{\alpha}{\alpha-k}} - \frac{\alpha-k}{k}} \qquad\qquad C^* = \frac{\frac{\alpha}{\alpha-k}(m^*)^{\frac{k}{\alpha-k}}}{(\frac{\alpha}{k})^{\frac{\alpha}{\alpha-k}} - \frac{\alpha-k}{k}}$$

*Then an optimal strategy is to take a randomized strategy with mixture distribution $\hat{\eta}$ where $\hat{\eta}(\{0\}, d\gamma) = \tilde{\eta}(d\gamma)$. The corresponding value function is*

$$V = y \frac{(\frac{\alpha}{k})^{\frac{k}{\alpha-k}}}{(\frac{\alpha}{k})^{\frac{k}{\alpha-k}} - \frac{\alpha-k}{\alpha}}.$$

It is worth highlighting here that the optimal stopping rule is not unique and although in Theorem 3 we find the optimal mixed threshold rule, there are other stopping times which are also optimal. In other words, suppose $\tau \in \mathcal{T}_R$ is a randomized threshold rule (which is not a pure threshold rule): then there exist other stopping times $\tau' \in \mathcal{T}$ for which $\mathcal{L}(X_\tau) = \mathcal{L}(X_{\tau'})$ or equivalently $\mathcal{L}(Y_\tau) = \mathcal{L}(Y_{\tau'})$.

### 4.4 A Generalized Example

Fix $L > 0$ and suppose $R \in (L, \infty)$. Let $J : [L, R] \mapsto \mathbb{R}$ be a continuously differentiable function with $J(z) > z$ and such that $\sup_{z\in[L,R]} \frac{J(z)}{z} < \kappa < \infty$. Let $K : [L, R] \mapsto \mathbb{R}$ be the largest increasing function such that $K \leq J$. Suppose $J(L) = K(L) \geq R$ and that the set $\{x : K(x) = J(x)\}$ is the union of a finite set of intervals. We write $\{x : K(x) = J(x)\} = \cup_{i=1}^N [\ell_i, r_i]$. Then $\ell_1 = L$ and $r_N = R$.

Let $\mathcal{A} = [L, R]$ and for $\alpha \in \mathcal{A}$ define $u_\alpha : [0, \infty) \mapsto \mathbb{R}$ by

$$u_\alpha(z) = \begin{cases} z & 0 \leq z < \alpha; \\ J(\alpha) & z \geq \alpha. \end{cases} \tag{12}$$

Let $\mathcal{W} = \{u_\alpha : \alpha \in [L, R]\}$.

Let $Y$ be Brownian motion started at $y \in (0, \frac{L}{\kappa})$, and absorbed at 0. Consider the problem of finding

$$\sup_\tau \inf_{u\in\mathcal{W}} u^{-1}(\mathbb{E}[u(Y_\tau)]).$$

Note that for any $\alpha \in [L, R]$ and any stopping time $\tau$, $\mathbb{E}[u_\alpha(Y_\tau)] \leq \kappa\mathbb{E}[Y_\tau] \leq \kappa y < L$. But $u^{-1}(x) = x$ over this range. Hence the $u^{-1}$ may be omitted in the definition of the Cautious Stochastic Utility in this example.

**Theorem 4.** *Let $\theta$ be given by*

$$\theta = \left[1 + \frac{1}{R}\int_L^R d\alpha \frac{\alpha K'(\alpha)}{K(\alpha) - \alpha} \exp\left(\int_\alpha^R d\beta \frac{K'(\beta)}{K(\beta) - \beta}\right)\right]^{-1}$$

*and let $\rho : [L, R] \mapsto \mathbb{R}_+$ be given by*

$$\rho(\alpha) = \frac{\theta}{R}\left\{\frac{\alpha K'(\alpha)}{K(\alpha) - \alpha} \exp\left(\int_\alpha^R d\beta \frac{K'(\beta)}{K(\beta) - \beta}\right)\right\}. \tag{13}$$

*Let $\tilde{\eta} \in \mathcal{P}([L, R])$ be the probability measure with density $\rho$ on $[L, R]$ and a point mass of size $\theta$ at $R$.*

*Then an optimal strategy is to take a randomized threshold strategy with mixture distribution $\hat{\eta}$ where $\hat{\eta}(\{0\}, d\gamma) = \tilde{\eta}(d\gamma)$ and $\hat{\eta}$ does not charge $(0, x) \times [x, \infty]$.*

We prove the theorem in Appendix D. Note that if $J$ is not strictly increasing then we have that $K$ is constant over intervals and $\tilde{\mu}$ does not charge such intervals. The reason for this is that the corresponding $u_\alpha$ strictly dominate other $u_\beta$ and are never the worst case utilities. For this reason they are not relevant in the CSC formulation. In the proof in the appendix, the utility functions are divided into two classes. For elements of the first class, the certainty equivalent is never smallest, and these utilities do not affect the CSC value. However, all elements of the second class are important, and we find the optimal strategy by making sure that the certainty equivalent is constant across utilities in this class, at least for the optimal mixed threshold stopping rule.

# 5 Two realistic models

In the previous section we studied a stylized liquidation problem and showed that in the CSC paradigm it is possible that the optimal strategy is not of threshold form. In this section we give two examples of more realistic models which are based on either $S$-shaped reference dependent utilities or on concave utilities.

## 5.1 An example with $S$-shaped reference dependent utilities

The first model is based on combining CSC with $S$-shaped reference-dependent preferences (Tversky and Kahneman (1992), see Kyle et al (2006), Henderson (2012) Ingersoll and Jin (2013), and Magnani (2017) in context of optimal stopping). This example can be shown to reduce to a form which is very closely related to the stylized example.

Suppose $Y$ follows geometric Brownian motion and solves $dY_t = \sigma Y_t dB_t + \mu Y_t dt$ subject to $Y_0 = y$. We assume $0 < \mu < \frac{1}{2}\sigma^2$. Let $\mathcal{W}^Y = \{u_i : 1 \leq i \leq N\}$ be a family of $S$-shaped reference dependent utility functions with

$$u_i(z) = \begin{cases} (z - R_i)^{\delta_i} & z \geq R_i \\ -\kappa_i (R_i - z)^{\delta_i} & z < R_i \end{cases} \tag{14}$$

where $\{(\delta_i, R_i, \kappa_i)\}_{1 \leq i \leq N}$ is a family of parameters. Here, for each $i$, $\delta_i \in (0,1)$ is a risk aversion/risk seeking parameter, $R_i > 0$ is the reference level and $\kappa_i \geq 1$ is the loss aversion parameter. Our problem is to find the CSC value

$$\sup_\tau \min_i u_i^{-1}(\mathbb{E}[u_i(Y_\tau)]). \tag{15}$$

Define $\psi = 1 - \frac{2\mu}{\sigma^2} \in (0,1)$ and set $s(z) = z^\psi$. Set $X = s(Y)$ and $x = s(y)$. Then $X$ solves $dX_t = \psi\sigma X_t dB_t$ subject to $X_0 = x := y^\psi > 0$. We have $X$ is a non-negative martingale. Set $g_i = u_i \circ s^{-1}$ so that

$$g_i(w) = \begin{cases} (w^{1/\psi} - R_i)^{\delta_i} & w \geq R_i^\psi \\ -\kappa_i (R_i - w^{1/\psi})^{\delta_i} & w < R_i^\psi \end{cases} \tag{16}$$

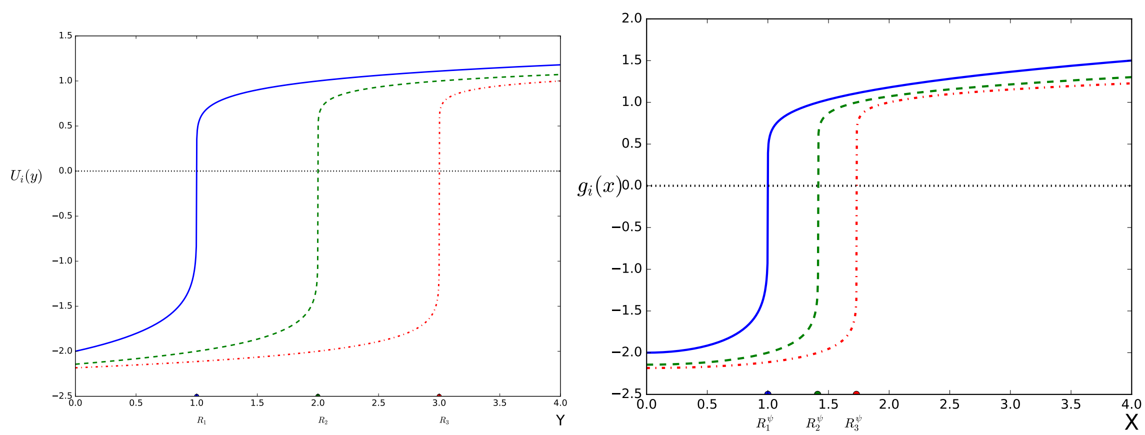
and set $\mathcal{W}^X = \{g_i : 1 \leq i \leq N\}$. By an immediate extension of the arguments leading to (4) we have

$$\sup_\tau \min_i u_i^{-1}(\mathbb{E}[u_i(Y_\tau)]) = s^{-1}\left(\sup_\tau \min_i g_i^{-1}(\mathbb{E}[g_i(X_\tau)])\right)$$

and hence in the search for the optimal stopping rule it is sufficient to consider the problem in natural scale for $X$ and $\mathcal{W}^X$.

Families of functions $\mathcal{W}^Y$ and $\mathcal{W}^X$ are given in Figure 2 for the parameters:

$$\psi = 0.5,\ N = 3 \text{ and } \{(\delta_i, R_i, \kappa_i)\} = \{(0.15, 1, 2), (0.1, 2, 2), (0.08, 3, 2)\}. \tag{17}$$



(a) $u_i \in \mathcal{W}^Y$

(b) $g_i \in \mathcal{W}^X$

Figure 2: The families of $S$-shaped reference dependent utility functions $\mathcal{W}^Y = \{u_i : 1 \leq i \leq N\}$ with $u_i$ defined in (14) and in natural scale $\mathcal{W}^X = \{g_i : 1 \leq i \leq N\}$ with $g_i$ given in (16). Parameters used are $\psi = 1/2$ for the price process, $N = 3$ and $\{(\delta_i, R_i, \kappa_i)\} = \{(0.15, 1, 2), (0.1, 2, 2), (0.08, 3, 2)\}$ for the utility functions where for each $i$, $\delta_i \in (0,1)$ is a risk aversion/risk seeking parameter, $R_i > 0$ is the reference level and $\kappa_i \geq 1$ is the loss aversion parameter.

Note that certainty equivalents are invariant under affine transformations of the objective function: if $h_{a,b}(w) = ah(w) + b$ with $a > 0$ then $h_{a,b}^{-1}(\mathbb{E}[h_{a,b}(Z)]) = h^{-1}(\mathbb{E}[h(Z)])$. Hence,

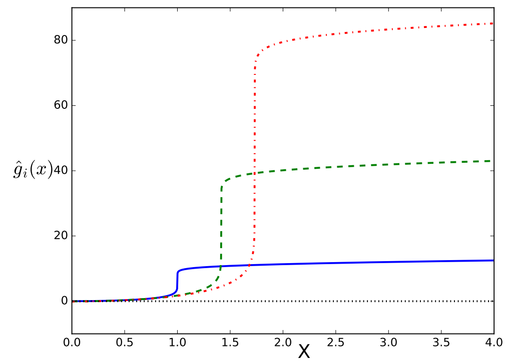



Figure 3: The family of transformed utility functions $\widetilde{\mathcal{W}}^X = \{\tilde{g}_i : 1 \leq i \leq N\}$ where $\tilde{g}_i$ is given by (18) with $\hat{x} = 0.8$. Parameters used are $\psi = 1/2$ for the price process, $N = 3$ and $\{(\delta_i, R_i, \kappa_i)\} = \{(0.15, 1, 2), (0.1, 2, 2), (0.08, 3, 2)\}$ for the utility functions where for each $i$, $\delta_i \in (0, 1)$ is a risk aversion/risk seeking parameter, $R_i > 0$ is the reference level and $\kappa_i \geq 1$ is the loss aversion parameter.

without loss of generality we may replace $\mathcal{W}^X = \{g_i : 1 \leq i \leq N\}$ with $\widetilde{\mathcal{W}}^X = \{\tilde{g}_i : 1 \leq i \leq N\}$ where for fixed $\hat{x} > 0$

$$\tilde{g}_i(w) = \frac{g_i(w) - g_i(0)}{g_i(\hat{x}) - g_i(0)} \tag{18}$$

These linear transformations have been designed so that $\tilde{g}_i(0) = 0$ and $\tilde{g}_i(\hat{x}) = 1$ for all $i$. Then, the functions $\tilde{g}_i$ are of comparable sizes over the region $[0, \hat{x}]$ and we expect that over the relevant range $g_i^{-1}$ does not depend greatly on $i$. The transformed family of functions $\widetilde{\mathcal{W}}^X$ are plotted in Figure 3. The key observation is that the resulting objective functions are similar to those studied in the stylized example in Section 4. Hence we expect a similar conclusion: it is not optimal to use a pure threshold rule, and instead there is an optimal stopping rule which is a non-trivial mixture of threshold rules.

Consider first the certainty equivalent from using a pure threshold strategy $\tau^X_{0,\gamma} = \inf\{t : X_t \notin (0, \gamma)\}$ for $\gamma > x = 0.2$. The certainty equivalents associated with the utilities $(u_i)_{i=1,2,3}$ as a function of the upper threshold are plotted in Figure 4. We see from the figure that the best pure threshold strategy uses an upper threshold of approximately 2.75 and yields a CSC certainty equivalent of 0.7263.

Now suppose we are allowed to search for the best mixed threshold strategy based on two upper thresholds (with the lower threshold set to zero). Figure 5 shows the highest CSC value

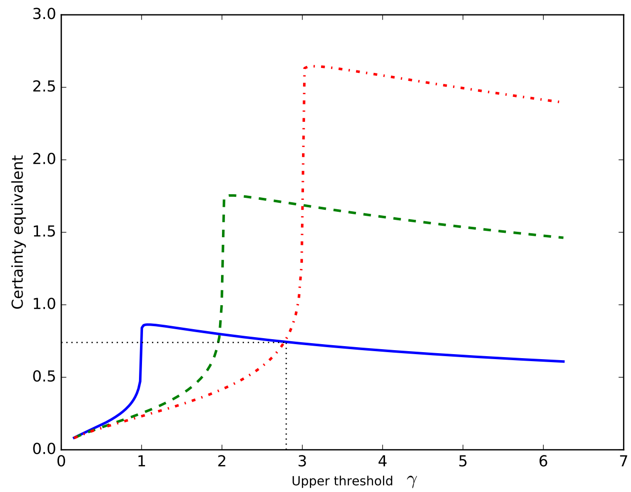



Figure 4: The certainty equivalent value under a pure threshold strategy $\tau^X_{0,\gamma} = \inf\{t : X_t \notin (0, \gamma)\}$ as a function of upper threshold $\gamma$ for $\gamma > X_0 = x = 0.2$. The family of $S$-shaped utility functions $u_i$ as defined in (14) are used. The best pure threshold strategy uses an upper threshold of about 2.75 and gives a CSC certainty equivalent of 0.7263, as marked on the figure. Parameters used are $\psi = 1/2$ for the price process, $N = 3$ and $\{(\delta_i, R_i, \kappa_i)\} = \{(0.15, 1, 2), (0.1, 2, 2), (0.08, 3, 2)\}$ for the utility functions where for each $i$, $\delta_i \in (0, 1)$ is a risk aversion/risk seeking parameter, $R_i > 0$ is the reference level and $\kappa_i \geq 1$ is the loss aversion parameter.

(as the mixture parameter varies) for a given pair of upper thresholds. Figure 6 shows how much probability mass is assigned to the smaller of the two upper thresholds.

The best strategy to to assign probability mass 0.75,0.25 to thresholds 1.1,3.1 respectively, giving a CSC value of 0.8368. From Figure 6 we see that for other pairs of thresholds, it is optimal to place all the weight on a single threshold, but for the optimal pair of thresholds the optimal strategy is a proper mixture. It follows that the best randomized strategy is strictly better than any pure threshold strategy.

We can also consider a mixture which involves at most three upper thresholds. We find that in this restricted class, the optimal randomized strategy assigns probability mass 0.76, 0.11, 0.13 to thresholds 1.1, 2.1, 3.1 respectively and gives a CSC value of 0.8425. Again, we see an improvement as we allow for mixtures over a larger number of thresholds. However, the benefit from adding more upper thresholds is diminishing, and the improvement in the CSC value from allowing mixed strategies which randomize over 4 upper thresholds is negligible.

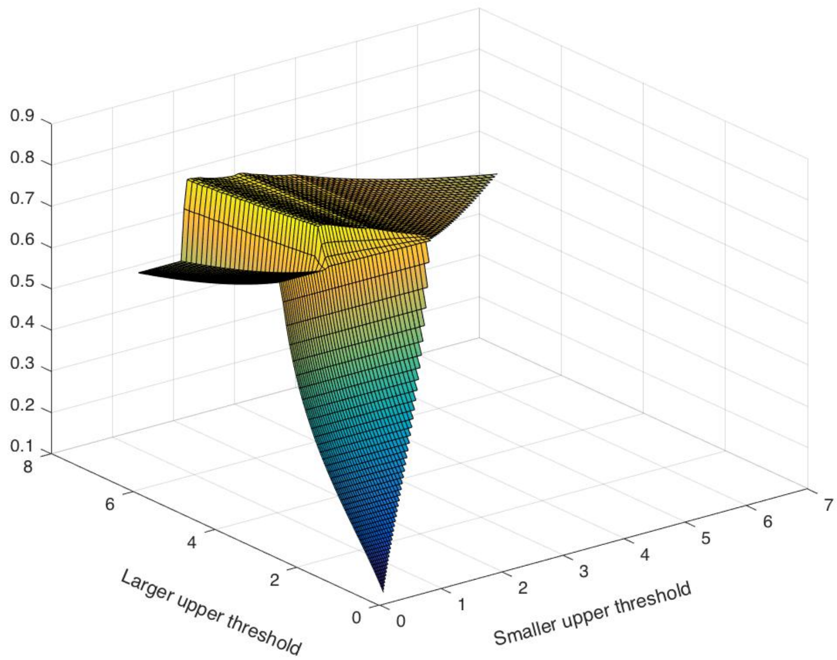



Figure 5: CSC value using the optimal mixture for a given pair of upper threshold levels where $X_0 = x = 0.2$. The family of $S$-shaped utility functions $u_i$ as defined in (14) are used. The best pair of upper thresholds is 1.1, 3.1 giving a CSC certainty equivalent of 0.8368. Parameters used are $\psi = 1/2$ for the price process, $N = 3$ and $\{(\delta_i, R_i, \kappa_i)\} = \{(0.15, 1, 2), (0.1, 2, 2), (0.08, 3, 2)\}$ for the utility functions where for each $i$, $\delta_i \in (0, 1)$ is a risk aversion/risk seeking parameter, $R_i > 0$ is the reference level and $\kappa_i \geq 1$ is the loss aversion parameter.

The results of randomization among upper thresholds for the family of S-shaped utility functions (in Figure 2) are summarized in Table 1.

## 5.2 An example based on concave utilities

In the previous example we used a family of $S$-shaped reference dependent utility functions. We saw that the problem could be reduced to a problem which shared many features with the stylized problem of Section 4. In this section we build a model using concave utility functions. We build our example from the sum of a power utility function and an exponential utility.[2]

[2]Our experience is that it is quite difficult to build examples based on families of concave utilities for which randomization is beneficial, especially if we restrict attention to standard one-parameter families (eg. CRRA or

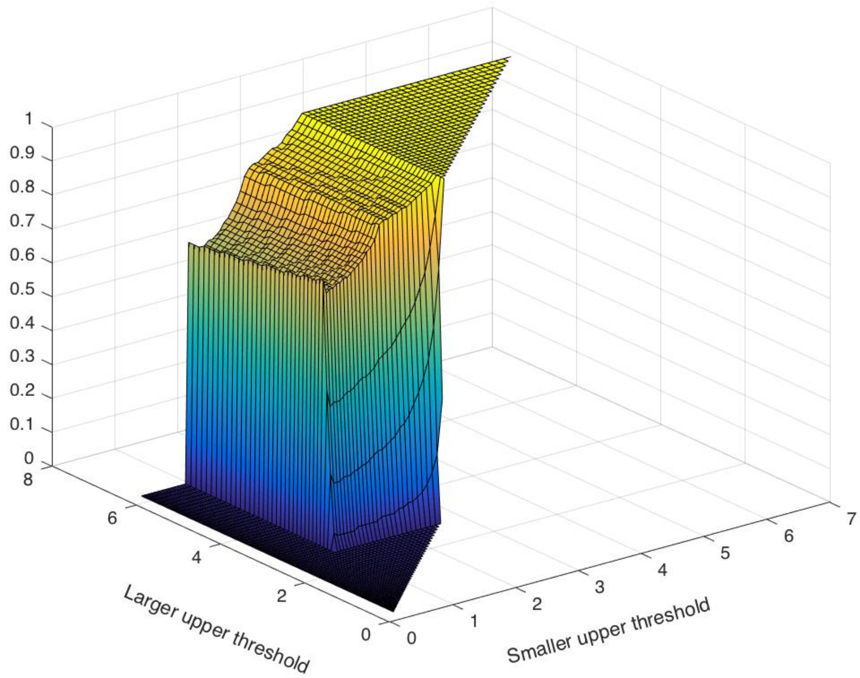



Figure 6: Optimal mixture distribution: the weight placed on the smaller of the upper thresholds for a given pair of upper thresholds. When both upper thresholds are large, it is optimal to not use a mixture, and only stop at the smaller of the upper thresholds; when both upper thresholds are small, it is again optimal not to use a mixture, and only stop at the larger of the upper thresholds. When the smaller upper threshold is in the range $1-3$, it is optimal to use a mixed strategy, with most of the mixture distribution on the smaller of the two upper thresholds. Again, $X_0 = x = 0.2$. The optimal mixture is to place probability mass 0.75 on threshold 1.1 and weight 0.25 on threshold 3.1. The family of $S$-shaped utility functions $u_i$ as defined in (14) are used. Parameters used are $\psi = 1/2$ for the price process, $N = 3$ and $\{(\delta_i, R_i, \kappa_i)\} = \{(0.15, 1, 2), (0.1, 2, 2), (0.08, 3, 2)\}$ for the utility functions where for each $i$, $\delta_i \in (0, 1)$ is a risk aversion/risk seeking parameter, $R_i > 0$ is the reference level and $\kappa_i \geq 1$ is the loss aversion parameter.

As in Section 5.1, suppose $Y$ is geometric Brownian motion with scale function $s(z) = z^{\psi}$ for $\psi \in (0, 1)$. For $\gamma, \kappa, \phi$ non-negative constants, define $f = f_{\gamma,\kappa,\phi} : \mathbb{R}^+ \mapsto \mathbb{R}^+$ by

$$f(z) = z^{\psi} + \frac{1}{\gamma} - \frac{1}{\gamma} e^{-\gamma[(z+\kappa)^{\phi} - \kappa^{\phi}]} \tag{19}$$

Then $f(0) = 0$ and provided $\phi < 1$, $f$ is concave. Set $g(w) = f \circ s^{-1}$ so that $g_{\gamma,\kappa,\phi}(w) =$

CARA). However, this example shows that it is possible to build examples of families of concave utilities for which randomization over thresholds is beneficial.

| Number of Thresholds | Best Thresholds | Best Mass Distribution | Best CSC |
|---|---|---|---|
| 1 | 2.75 | 1 | 0.7263 |
| 2 | (1.1, 3.1) | (0.75, 0.25) | 0.8368 |
| 3 | (1.1, 2.1, 3.1) | (0.76, 0.11, 0.13) | 0.8425 |
| 4 | Negligible improvement over 3 thresholds case | | |

Table 1: Summary of results of randomization among upper thresholds for the family of S-shaped utility functions in Figure 2.

| Number of Thresholds | Best Thresholds | Best Mass Distribution | Best CSC |
|---|---|---|---|
| 1 | 22..68 | 1 | 0.6215 |
| 2 | (3.84, 187.42) | (0.56, 0.44) | 0.6373 |
| 3 | Negligible improvement over 2 thresholds case | | |

Table 2: Summary of results of randomization among upper thresholds for the family of concave utility functions in Figure 7.

$f_{\gamma,\kappa,\phi}(w^{1/\psi})$. Then

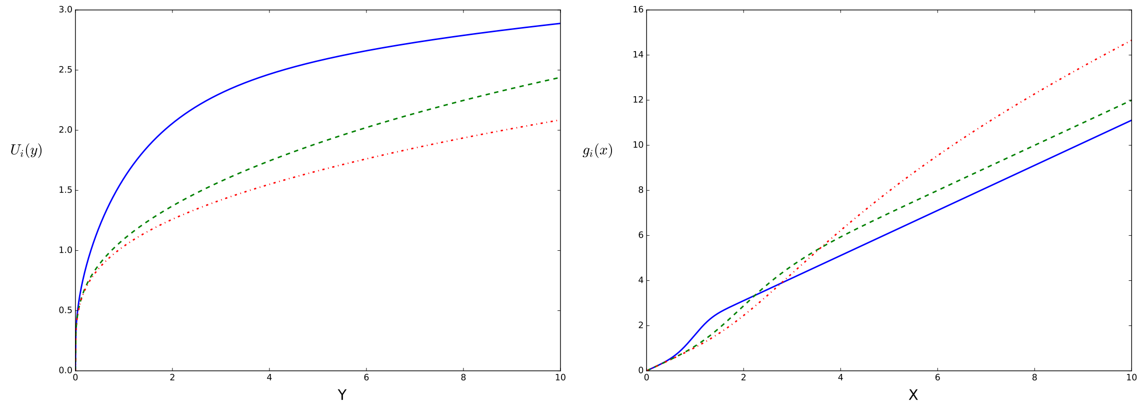
$$g(w) = w + \frac{1}{\gamma} - \frac{1}{\gamma} e^{-\gamma[(w^{1/\psi}+\kappa)^{\phi}-\kappa^{\phi}]}. \tag{20}$$

Provided $\psi < \phi$ we have that $g$ is convex for small values of $w$ and concave for larger values. We will thus assume $\psi < \phi < 1$.

Let $\mathcal{W}^Y = \{u_i : 1 \leq i \leq N\}$ where $u_i(z) = f_{\gamma_i,\kappa_i,\phi_i}(z)$. Then $\mathcal{W}^X = \{g_i : 1 \leq i \leq N\}$ where $g_i(w) = g_{\gamma_i,\kappa_i,\phi_i}(w)$. Families of functions $\mathcal{W}^Y$ and $\mathcal{W}^X$ are given in Figure 7 for the parameters:

$$\psi = 1/4,\ N = 3 \text{ and } \{(\gamma_i, \kappa_i, \phi_i)\} = \{(0.9, 1, 0.9), (0.5, 10, 0.4), (0.2, 20, 0.3)\}. \tag{21}$$

Consider first pure threshold strategies, $\tau^X_{0,\gamma}$ for different upper thresholds $\gamma$. The certainty equivalents associated with the utilities $\{u_i\}_{i=1,2,3}$ as a function of the upper threshold are plotted in Figure 8 with an initial value of $X_0 = x = 0.5$. We see from the figure that the best pure threshold strategy uses an upper threshold of approximately 22.68 and yields a CSC certainty equivalent of 0.6215.




(a) $u_i \in \mathcal{W}^Y$ (b) $g_i \in \mathcal{W}^X$

Figure 7: The families of concave utility functions $\mathcal{W}^Y = \{u_i : 1 \leq i \leq N\}$ with $u_i(z) = f_{\gamma_i,\kappa_i,\phi_i}(z)$ and $f$ defined in (19). In natural scale, $\mathcal{W}^X = \{g_i : 1 \leq i \leq N\}$ with $g_i(w) = g_{\gamma_i,\kappa_i,\phi_i}(w)$ where $g$ is given in (20). Parameters used are $\psi = 1/4$ for the price process, $N = 3$ and $\{(\gamma_i, \kappa_i, \phi_i)\} = \{(0.9, 1, 0.9), (0.5, 10, 0.4), (0.2, 20, 0.3)\}$.

If we now search for the best randomization over two upper thresholds we find that the best strategy is to assign probability mass 0.56, 0.44 to thresholds 3.84, 187.42 respectively and that this gives a CSC value of 0.6373. Again the best randomized strategy is strictly better than any pure threshold strategy. However, allowing randomization over three upper thresholds brings only negligible further benefits. The results of randomization among upper thresholds for the family of concave utilities (in Figure 7) are summarized in Table 2.

# 6 Concluding Remarks

This paper considers agents who exhibit cautious stochastic choice (CSC) and who face optimal timing or stopping decisions in a dynamic setting. We build on the seminal work on CSC in a static setting by Cerreia-Vioglio et al (2015, 2017) and provide a continuous-time optimal stopping model under CSC. In our dynamic setup, the value associated with a stopping rule is not quasi-convex and hence we cannot necessarily expect there to be a pure threshold rule which is optimal. Despite this observation, it is quite a challenge to find examples where it can be clearly demonstrated that the optimal stopping rule is a non-trivial mixture of threshold strategies. This paper has taken up this challenge and provides first, a stylized, tractable example whereby

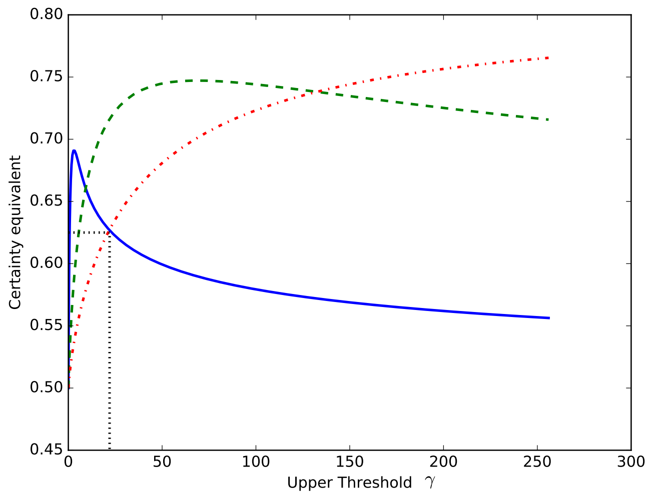



Figure 8: The certainty equivalent value under a pure threshold strategy $\tau^X_{0,\gamma} = \inf\{t : X_t \notin (0, \gamma)\}$ as a function of upper threshold $\gamma$ for $\gamma > X_0 = x = 0.5$. The family of concave utility functions $u_i(z) = f_{\gamma_i,\kappa_i,\phi_i}(z)$ with $f$ defined in (19) are used. The best pure threshold strategy uses an upper threshold of approximately 22.68 and gives a CSC certainty equivalent of 0.6215, as marked on the figure. Parameters used are $\psi = 1/4$ for the price process, $N = 3$ and $\{(\gamma_i, \kappa_i, \phi_i)\} = \{(0.9, 1, 0.9), (0.5, 10, 0.4), (0.2, 20, 0.3)\}$.

the optimal stopping rule and value can be constructed explicitly, and second, two realistic example models under reference-dependent or concave families of utility functions under which pure threshold strategies are not optimal. Our predictions are in line with recent experimental evidence in dynamic settings whereby individuals do not play cut-off or threshold strategies (Strack and Viefers (2017), Fischbacher, Hoffmann and Schudy (2015)).

# 7 References

Agranov, M. And P. Ortoleva, 2017, Stochastic choice and preferences for randomization, *Journal of Political Economy*, 125(1), 40-68.

Cerreia-Vioglio, S., D.Dillenberger, and P. Ortoleva, 2015, Cautious Expected Utility and the Certainty Effect, *Econometrica*, 83, 693-728

Cerreia-Vioglio, S., D.Dillenberger, P. Ortoleva, and G. Riella, 2017, Deliberately Stochastic, Working paper, Columbia University.

Cox, A.M.G. and D.G. Hobson, 2004, An optimal Skorokhod embedding for diffusions, *Stochastic Processes and Applications*, 111, 17-39.

Dayanik, S. and I. Karatzas, 2003, On the optimal stopping problem for one-dimensional diffusions, *Stochastic Processes and their Applications*. 107:2, 173-212.

Dwenger N., Kübler D., and G. Weizsäcker, 2013, Flipping a Coin: Evidence from Laboratory and Field, *SSRN Working paper, id2353282*.

Ebert S. and P. Strack, 2015, Until the Bitter End: On Prospect Theory in a Dynamic Context, *American Economic Review*, 105(4), 1618-1633.

Fudenberg, D., R. Iijima and T. Strzalecki, 2015, Stochastic choice and revealed perturbed utility, *Econometrica*. 83 (6): 2371-2409.

Fudenberg, D., P. Strack and T. Strzalecki, 2017, Speed, Accuracy and the Optimal Timing of Choices, Working paper, Harvard University.

Gul, F. and W. Pesendorfer, 2006, Random Expected Utility, *Econometrica*, 74,121-146.

He X., S. Hu, J. Obloj and X.Y. Zhou, 2017, Path dependent and randomized strategies in Barberis' Casino Gambling model, *Operations Research*, 65, 1, 97-103.

Henderson V., 2012, Prospect Theory, Liquidation and the Disposition Effect, *Management Science*, 58, 2, February 2012, 445-460.

Henderson V., Hobson D. and A.S.L.Tse, 2017, Randomized Strategies and Prospect Theory in a Dynamic Context, *Journal of Economic Theory*, 168, 287-300.

Henderson V., D. Hobson and M. Zeng, 2017, Optimal Stopping and the Sufficiency of Randomized Thresholds, Working paper.

Hey J.D. and Orme C., 1994, Investigating generalizations of expected utility theory using experimental data, *Econometrica*, 1291-1326.

Ingersoll J.E. and L. Jin, 2013, Realization Utility with Reference-Dependent Preferences, *Review of Financial Studies*, 26 (3), 723-767.

Johnson E.J. and R. Ratcliff (2013), Computational and Process Models of Decision Making in Psychology and Behavioral Economics, in *Neutroeconomics: Decision Making and the Brain.* ed. by P.W. Glimcher and E. Fehr, Academic Press, 35-48.

Kahneman, D. and A. Tversky, 1979, Prospect Theory: An Analysis of Decision Under Risk, *Econometrica*, 46, 171-185.

Karni, E. and Safra Z., 1990, Behaviorally Consistent Optimal Stopping Rules, *Journal of*

*Economic Theory*, 51(2), 391-402.

Kyle, A.S., Ou-Yang, H. and W. Xiong, 2006, Prospect Theory and Liquidation Decisions, *Journal of Economic Theory*, 129, 273-288.

Luce, R.D., 1959, *Individual Choice Behavior: A Theoretical Analysis.* Wiley, New York.

Maccheroni F., (2002), Maxmin under Risk, *Economic Theory*, 19, 823-831.

Machina M.J., 1985, Stochastic Choice Functions Generated from Deterministic Preferences over Lotteries, *Economic Journal*, 95, 379, 575-594.

Magnani J., 2016, Reference-Dependent Preferences and Optimal Stopping Decisions: A Laboratory Investigation, Working paper.

Oprea R., Friedman D., and S.T. Anderson, 2009, Learning to wait: A laboratory investigation, *The Review of Economic Studies*, 76(3), 1103-1124.

Pedersen, J. and Pekir, G., 2001, The Azema-Yor embedding in non-singular diffusions, *Stochastic Process. Appl.*, 96, 305-312.

Peskir, G. and A. Shiryaev, 2006, *Optimal Stopping and Free-Boundary Problems.* Birkhauser Verlag, Basel.

Strack P. and P. Viefers, 2017, Too Proud to Stop: Regret in Dynamic Decisions, *SSRN Working paper, id2465840.*

Tversky, A., 1969, Intransitivity of preferences, *Psychological review*, 76, 31.

Tversky, A. And D. Kahneman, 1992, Advances in prospect theory: Cumulative representation of uncertainty, *J. Risk Uncertainty* 5(4) 297-323

Wakker P., 1994, Separating marginal utility and probabilistic risk aversion, *Theory and Decision*, 36(1), 1-44.

Xu Z.Q. and X.Y. Zhou, 2013, Optimal stopping under probability distortion, *Ann. Appl. Prob.*, 23, 1, 251-282.

# Appendix

## A CSC model in a static setting (Cerreia-Vioglio et al (2017))

In this appendix, we present the CSC model in a static setting and give a mixing result of Cerreia-Vioglio et al (2017).

Consider an interval $I \subseteq \mathbb{R}$ of possible monetary prizes. Let $\Delta = \{\nu : \nu \in \mathcal{P}(I)\}$ be the set of lotteries over $I$ and let $Q$ be a subset of $\Delta$. The agent has a set of utility functions $\mathcal{W} \subseteq F^I_\uparrow$. Given a lottery $q \in Q$ and a utility $u \in \mathcal{W}$, we denote by $\mathbb{E}^q(u)$ the expected utility of $u$ with respect to $q$, that is $\mathbb{E}^q(u) = \int_I u(x)q(dx)$. The certainty equivalent of $q$ with respect to $u$ is defined as

$$C^u_q = u^{-1}(\mathbb{E}^q(u)). \tag{A-1}$$

Under the CSC paradigm of Cerreia-Vioglio et al (2015, 2017) (see also earlier work of Maccheroni (2002)) the agent chooses a best lottery from $Q$ by displaying cautious behavior: *the evaluation for any given lottery $q \in Q$ is determined by $V^q = \min_{u\in\mathcal{W}} C^u_q$; the optimal strategy is to choose the lottery $\bar{q} \in Q$ which maximizes $V^q$*. This involves both minimization and maximization steps. Note that typically $I$, $Q$ and $\mathcal{W}$ are taken to be compact so that the optimizers exist.

Now we want to allow the agent to mix over lotteries. Let $co(Q)$ denote the convex hull of Q. Then $\rho = \rho^\lambda \in co(A)$, represents a compound lottery obtained through a randomization $\lambda$ over the lotteries in $Q$. If $\lambda$ is a discrete distribution over $q \in Q$ we have $\rho^\lambda = \sum \lambda_i q_i$; more generally, $\rho^\lambda = \int_Q \lambda(dq)q$ is a measure on $I$ given by $\rho^\lambda(dx) = \int_Q \lambda(dq)q(dx)$. For a lottery $\rho^\lambda \in co(Q)$ we can define the expected utility of $u$ with respect to $\rho^\lambda$ by $\mathbb{E}^{\rho^\lambda}(u) = \int_I u(x)\rho^\lambda(dx) = \int_Q \lambda(dq)\mathbb{E}^q(u)$, and then the certainty equivalent of $\rho^\lambda$ with respect to $u$ is $C^u_{\rho^\lambda} = u^{-1}(\mathbb{E}^{\rho^\lambda}(u))$. Let $V^{\rho^\lambda} = \min_{u\in\mathcal{W}} C^u_{\rho^\lambda}$. Then an optimal randomized lottery is given by

$$\rho^* = \rho^*(Q) \in \underset{\rho^\lambda\in co(A)}{arg\ max}\ V^{\rho^\lambda}.$$

In this static setting, Cerreia-Vioglio et al (2017) show that mixing over two lotteries may improve the worst case certainty equivalent.

Suppose $Q = \{p, q\}$ and $\mathcal{W} = \{u, v\}$. If we have

$$C^u_p > K > C^v_p$$
$$C^u_q < K < C^v_q$$

then, a linear combination of $p$ and $q$ is better than any one of them in terms of smallest certainty equivalent. To see this, note that for $\lambda \in (0, 1)$ and $\rho = \rho^\lambda = \lambda p + (1 - \lambda)q$

$$\mathbb{E}^\rho(u) = \lambda\mathbb{E}^p(u) + (1-\lambda)\mathbb{E}^q(u) = \lambda u(C^u_p) + (1-\lambda)u(C^u_q) > \lambda u(K) + (1-\lambda)u(C^u_q) > u(C^u_q)$$

and it follows that $u^{-1}(\mathbb{E}[u(\lambda p + (1-\lambda)q)]) > u^{-1}(u(C^u_q)) = C^u_q$ since $u$ is strictly increasing. A similar argument gives that $v^{-1}(\mathbb{E}[v(\lambda p + (1-\lambda)q)]) > C^v_p$. Then

$$\min_{w\in\mathcal{W}} w^{-1}(\mathbb{E}^{\rho^\lambda}(w)] > \max\{C^v_p, C^u_q\} = \max_{r\in Q} \min_{w\in\mathcal{W}} C^w_r.$$

It follows that in a static setting it can be optimal to take a mixed strategy.

# B Further results on optimal stopping in the classical case

Under Assumption 1 the process $X$ is a non-negative local martingale, and hence a supermartingale. Further, for any stopping time $\mathbb{E}^x[X_\tau] \leq x$. If $I^X$ is bounded above then $X$ is a martingale and $\mathbb{E}^x[X_\tau] = x$, but for many examples $I^X = (0,\infty)$ or $[0,\infty)$ and then there exist $\tau$ for which $\mathbb{E}^x[X_\tau] < x$. Recall $\mathcal{Q}^X(\mathcal{S})$ is the set of possible laws of the stopped $X$-process, over stopping times in $\mathcal{S}$.

**Lemma 1.** *1. If $I^X$ is bounded then $Q^X(\mathcal{T}) = \{\mathcal{L}(X_\tau) : \tau \in \mathcal{T}\} = \{\nu \in \mathcal{P}(\bar{I}^X) : \int_{\bar{I}^X} z\nu(dz) = x\}$. If $I^X$ is not bounded above then $Q^X(\mathcal{T}) = \{\mathcal{L}(X_\tau) : \tau \in \mathcal{T}\} = \{\nu \in \mathcal{P}(\bar{I}^X) : \int_{\bar{I}^X} z\nu(dz) \leq x\}$.*

*2. If $I^X$ is bounded then $Q^X(\mathcal{T}_T) = \{\mathcal{L}(X_\tau) : \tau \in \mathcal{T}_T\} = \delta_x \cup \left(\{\cup_{\beta<x<\gamma;\ \beta,\gamma\in\bar{I}^X} \chi^x_{a,b}\}\right)$ where $\chi^x_{a,b}$ is the mixture of point masses $\chi^x_{a,b} = \frac{x-a}{b-a}\delta_b + \frac{b-x}{b-a}\delta_a$. If $I^X$ is not bounded above then $Q^X(\mathcal{T}_T) = \{\mathcal{L}(X_\tau) : \tau \in \mathcal{T}_T\} = (\cup_{0\leq\beta\leq x}\delta_\beta) \cup \left(\{\cup_{0\leq\beta<x<\gamma<\infty}\chi^x_{a,b}\}\right)$.*

*3. In both cases $Q^X(\mathcal{T}_R) = Q^X(\mathcal{T})$.*

*Proof.* 1. Suppose $I^X$ is unbounded. The fact that $Q^X(\mathcal{T}) \subseteq \{\nu \in \mathcal{P}(\bar{I}^X) : \int_{\bar{I}^X} z\nu(dz) \leq x\}$ follows from the remarks before the statement of the Lemma. The fact that we have equality follows from the fact that by Skorokhod's Embedding Theorem any $\nu \in \mathcal{P}([0,\infty))$ with $\int z\nu(dz) \leq x$ can be obtained as the law of $X_\tau$ for a stopping time $\tau$. See Pedersen and Peskir (2001) or Cox and Hobson (2004) The case of bounded $X$ has a similar proof.

2. This is immediate from the definition of pure threshold rules.

3. This follows from Lemma 1 of Henderson, Hobson and Zeng (2017).

□

Lemma 1 characterizes the sets $Q^X(\mathcal{S})$ for various sets $\mathcal{S}$. However, the sets $\mathcal{T}$, $\mathcal{T}_T$ and $\mathcal{T}_R$ do not depend on whether we consider stopping times for the process $X$ or $Y$. Hence $\{\nu : \nu \in Q^Y(\mathcal{S})\} = \{\eta_\sharp s : \eta \in Q^X(\mathcal{S})\}$ where, by definition $\eta_\sharp s(A) = \eta(s(A))$.

*Proof of Proposition 1.* This proposition is standard but we provide a short proof which will have parallels to our method in the CSC case. The results will follow if we can show that $\sup_{\tau\in\mathcal{T}_T} \mathbb{E}^x[g(X_\tau)] = \sup_{\tau\in\mathcal{T}} \mathbb{E}^x[g(X_\tau)] = g^{cv}(x)$. Note that $g = U \circ s^{-1}$ is increasing. For $d \geq 0$ suppose $g(z) \leq c + dz$. Then, by the supermartingale property of $X$, $\mathbb{E}^x[g(X_\tau)] \leq c + d\mathbb{E}^x[X_\tau] \leq c + dx$. Taking an infimum over $c, d \geq 0$ for which $c + dz \geq g(z)$ we find $\sup_{\tau\in\mathcal{T}} \mathbb{E}^x[g(X_\tau)] = g^{cv}(x)$. Conversely, either $g(x) = g^{cv}(x)$ and then $\sup_{\tau\in\mathcal{T}_T} \mathbb{E}^x[g(X_\tau)] \geq \mathbb{E}^x[g(X_0)] = g^{cv}(x)$ or there exists a largest interval $I_x$ with endpoints $\{a_x, b_x\}$ such that $x \in I_x$ and $g(z) < g^{cv}(z)$ on $I_x$. If $g(a_x) = g^{cv}(a_x)$ and $g(b_x) = g^{cv}(b_x)$ then $\sup_{\tau\in\mathcal{T}_T} \mathbb{E}^x[g(X_\tau)] \geq \mathbb{E}^x[g(X_{\tau^X_{a,b}})] = g^{cv}(x)$. Otherwise, there exist $a_n \to a_x$ and $b_n \to b_x$ such that $g(a_n) \to g^{cv}(a_x)$ and $g(b_n) \to g^{cv}(b_x)$. Then $\sup_{\tau\in\mathcal{T}_T} \mathbb{E}^x[g(X_\tau)] \geq \limsup \mathbb{E}^x[g(X_{\tau^X_{a_n,b_n}})] = g^{cv}(x)$. □

# C Proofs and auxilliary results for the stylized example

*Proof of Theorem 1.* The result follows immediately from the following lemma. □

**Lemma 2.** *For all* $0 \leq \beta \leq y \leq \gamma$, $\inf_{m \in \mathcal{M}} C^m_{\beta,\gamma} \leq y$.
*Moreover,* $\sup_{\beta,\gamma} \inf_{m \in \mathcal{M}} C^m_{\beta,\gamma} = y$.

*Proof.* If $\gamma \in [y, m_*)$ then $C^m_{\beta,\gamma} = y$ for all $m \in \mathcal{M}$.

If $\gamma \geq m^*$ then using the fact that $C^m_{\beta,\gamma}$ is increasing in $m$ for $m \leq \gamma$ and $\alpha m_* = k m^*$,

$$\inf_{m \in \mathcal{M}} C^m_{\beta,\gamma} = C^{m_*}_{\beta,\gamma} = \frac{1}{k}\left(\alpha m_* \frac{y-\beta}{\gamma-\beta} + k\beta\frac{\gamma-y}{\gamma-\beta}\right) = m^* \frac{y-\beta}{\gamma-\beta} + \beta\frac{\gamma-y}{\gamma-\beta} = y - \frac{(y-\beta)(\gamma-m^*)}{(\gamma-\beta)} \leq y.$$

Finally, if $\gamma \in [m_*, m^*)$ then $\inf_{m \in \mathcal{M}} C^m_{\beta,\gamma} \leq C^{m^*}_{\beta,\gamma} = y$.

The first statement of the lemma follows from consideration of the three possible cases. The second statement follows from the first, given that for all $m$, $C^m_{\beta,y} = y$. □

Next we record some useful properties about $\mathcal{G}^m_{\beta,\gamma}$ $C^m_{\beta,\gamma}$ which follow immediately from the definitions in (9).

**Lemma 3.**
1. *For each* $m \in \mathcal{M}$, $\mathcal{G}^m_{\beta,\gamma}$ *and* $C^m_{\beta,\gamma}$ *are non-increasing in* $\gamma$ *for* $\gamma \geq m^*$. *Hence, for* $\gamma \geq m^*$, $\inf_m C^m_{\beta,\gamma} \leq \inf_m C^m_{\beta,m^*}$.
2. *For each* $m \in \mathcal{M}$, *and* $\gamma \leq m^*$, $\mathcal{G}^m_{\beta,\gamma}$ *and* $C^m_{\beta,m}$ *are non-increasing in* $\beta$ *for* $0 \leq \beta \leq y$. *Hence for* $\gamma \leq m^*$, $\inf_m C^m_{\beta,\gamma} \leq \inf_m C^m_{0,\gamma}$.

Consider the spaces of probability measures $\mathcal{P}(\mathcal{M})$, $\mathcal{P}(\{0\} \times \mathcal{M})$ and $\mathcal{P}([0,y) \times [y,\infty])$. There is an obvious 1-1 correspondence between measures $\tilde{\zeta} \in \mathcal{P}(\mathcal{M})$ and $\hat{\zeta} \in \mathcal{P}(\{0\}\times\mathcal{M})$ given by $\hat{\zeta}(\{0\}\times A) = \tilde{\zeta}(A)$. Write $\tilde{\eta} = p(\hat{\eta})$ for this correspondence. Similarly, we can define a map $P : \mathcal{P}([0,y)\times[y,\infty]) \mapsto \mathcal{P}(\{0\}\times\mathcal{M})$ by $P(\eta) = \hat{\eta}$ where

$$\hat{\eta}(\{0\}, d\gamma) = \begin{cases} \int_\beta \int_{\gamma \leq m_*} \eta(d\beta, d\gamma) & \gamma = m_* \\ \int_\beta \eta(d\beta, d\gamma) & \gamma \in (m_*, m^*) \\ \int_\beta \int_{\gamma \geq m^*} \eta(d\beta, d\gamma) & \gamma = m^* \end{cases}$$

Recall $\mathcal{G}^m_{\beta,\gamma} = \mathbb{E}[u_m(Y_{\tau_{\beta,\gamma}})]$ and $\mathcal{C}^m_{\beta,\gamma} = u_m^{-1}(\mathcal{G}^m_{\beta,\gamma})$. Let $\eta$ be a probability measure on $[0,y)\times[y,\infty]$. We can define a randomized stopping time $\tau = \tau_\eta$ by generating a random variable $\Theta = (\Theta_\beta, \Theta_\gamma)$ with law $\eta$ on $[0,y)\times[y,\infty]$ and setting $\tau = \tau_{\Theta_\beta,\Theta_\gamma}$. Then we define $\mathcal{G}^m_\eta$ to be the expected utility from using the randomized stopping rule $\tau_\eta$:

$$\mathcal{G}^m_\eta = \int_{[0,y)\times[y,\infty]} \eta(d\beta, d\gamma)\mathcal{G}^m_{\beta,\gamma} = \mathbb{E}^\eta[u_m(Y_{\tau_{\beta,\gamma}})]$$

Finally we set $\mathcal{C}^m_\eta$ to be the certainty equivalent: $\mathcal{C}^m_\eta = u_m^{-1}(\mathcal{G}^m_\eta)$.

**Lemma 4.** $\mathcal{G}^m_\eta \leq \mathcal{G}^m_{P(\eta)}$.

*Proof.* From Lemma 3, for $m \in \mathcal{M}$ and $\gamma \geq m^*$, $\mathcal{G}^m_{\beta,\gamma} \leq \mathcal{G}^m_{\beta,m^*} \leq \mathcal{G}^m_{0,m^*}$. Similarly, for $m \in \mathcal{M}$ and $\gamma < m_*$, $\mathcal{G}^m_{\beta,\gamma} = y = \mathcal{G}^m_{\beta,m_*} = \mathcal{G}^m_{0,m_*}$ Then defining $M(\gamma) = (\gamma \vee m_*) \wedge m^*$ we have $\mathcal{G}^m_{\beta,\gamma} \leq \mathcal{G}^m_{0,M(\gamma)}$. It follows that for all $(\beta,\gamma) \in ([0,y) \times [y,\infty])$ we have $\mathcal{G}^m_{\beta,\gamma} \leq \mathcal{G}^m_{0,M(\gamma)}$. Then for any $\eta \in \mathcal{P}([0,y) \times [y,\infty])$, writing $\hat{\eta} = P(\eta)$,

$$\mathcal{G}^m_\eta = \int \eta(d\beta, d\gamma)\mathcal{G}^m_{\beta,\gamma} \leq \int \hat{\eta}(d\beta, d\gamma)\mathcal{G}^m_{\beta,\gamma} = \mathcal{G}^m_{\hat{\eta}}.$$

□

For $\tilde{\zeta} \in \mathcal{P}(\mathcal{M})$ define $\mathcal{C}^m_{\tilde{\zeta}} = \mathcal{C}^m_{\hat{\zeta}}$ where $\hat{\zeta} = p^{-1}(\tilde{\zeta})$.

**Corollary 2.**

$$\sup_{\eta \in \mathcal{P}([0,y)\times[y,\infty])} \inf_{m\in\mathcal{M}} \mathcal{C}^m_\eta = \sup_{\zeta \in \mathcal{P}(\{0\}\times\mathcal{M})} \inf_{m\in\mathcal{M}} \mathcal{C}^m_\zeta = \sup_{\tilde{\zeta} \in \mathcal{P}(\mathcal{M})} \inf_{m\in\mathcal{M}} \mathcal{C}^m_{\tilde{\zeta}}$$

*Proof.* We have

$$\begin{aligned} \sup_{\eta \in \mathcal{P}([0,y)\times[y,\infty])} \inf_{m\in\mathcal{M}} u_m^{-1}\left(\int \eta(d\beta,d\gamma)\mathcal{G}^m_{\beta,\gamma}\right) &\leq \sup_{\hat{\eta} \in \mathcal{P}(\{0\}\times\mathcal{M})} \inf_{m\in\mathcal{M}} u_m^{-1}\left(\int \hat{\eta}(d\beta,d\gamma)\mathcal{G}^m_{\beta,\gamma}\right) \quad \text{(A-2)} \\ &= \sup_{\tilde{\eta} \in \mathcal{P}(\mathcal{M})} \inf_{m\in\mathcal{M}} u_m^{-1}\left(\int \tilde{\eta}(d\gamma)\mathcal{G}^m_{0,\gamma}\right) \end{aligned}$$

Since $\mathcal{P}(\{0\} \times \mathcal{M}) \subseteq \mathcal{P}([0,y) \times [y,\infty])$ the inequality (A-2) is an equality, and the result follows. □

*Proof of Proposition 3.* Let $\zeta$ be a probability measure on $\{0\} \times \mathcal{M} \subseteq [0,y) \times [y,\infty)$. We can identify $\zeta$ with a probability measure $\tilde{\zeta}$ on $\mathcal{M}$ and then $\mathcal{G}^m_{\tilde{\zeta}} = \int_{\mathcal{M}} \tilde{\zeta}(d\gamma)\mathcal{G}^m_{0,\gamma}$ and $\mathcal{C}^m_{\tilde{\zeta}} = u_m^{-1}(\mathcal{G}^m_{\tilde{\zeta}})$.

Corollary 2 shows that $V(\mathcal{T}^0_R) = V(\mathcal{T}_R)$ and it is sufficient to only consider threshold strategies in the mixture with lower bound at 0 and upper bound in $\mathcal{M}$. The fact that $V(\mathcal{T}_R) = V(\mathcal{T})$ follows from Proposition 2 and $Q(\mathcal{T}_R) = Q(\mathcal{T})$. □

Our calculation of the optimal strategy in the CSC setting is based on the following general proposition. Let $\mathcal{Z}$ be a set and let $\mathcal{N}$ be a measurable space. Let $D : \mathcal{Z} \times \mathcal{N} \mapsto \mathbb{R}$ be a map and set $D_*(z) = \inf_{n\in\mathcal{N}} D(z,n)$ and $D^* = D^*(\mathcal{Z}) = \sup_{z\in\mathcal{Z}} D_*(z)$.

**Proposition 4.** *Suppose there exist $\mathcal{Z}_0 \subseteq \mathcal{Z}$, $\mathcal{N}_0 \subseteq \mathcal{N}$, $z^* \in \mathcal{Z}_0$, $\nu \in \mathcal{P}(\mathcal{N}_0)$, a family $(h_n)_{n\in\mathcal{N}_0}$ of strictly increasing functions $h_n : \mathbb{R} \mapsto \mathbb{R}$ and constants $\hat{D}$, $\hat{H}$ such that*

*$D(z^*,n) \geq \hat{D}$ on $\mathcal{N}$ with $D(z^*,n) = \hat{D}$ on $\mathcal{N}_0$*

*$\int_{\mathcal{N}_0} \nu(dn)h_n(D(z,n)) \leq \hat{H}$ on $\mathcal{Z}$ with $\int_{\mathcal{N}_0} \nu(dn)h_n(D(z,n)) = \hat{H}$ on $\mathcal{Z}_0$*

*Then, for any $z \in \mathcal{Z}$*

$$D_*(z) \leq D_*(z^*) = D^*$$

In our interpretation we take $\mathcal{Z}$ to be either the space of stopping rules or the space of attainable laws or the set of randomizations of the levels of lower and upper thresholds. (Since our problem is law invariant, the final result will be equivalent.) $\mathcal{Z}_0$ is a space of relevant stopping rules or attainable laws or randomizations, for example the set of randomized threshold rules for which the upper barrier lies in some interval. $\mathcal{N}$ is a parameterization of the space of utility functions and $\mathcal{N}_0$ is a set of relevant utility functions. We may have $\mathcal{N}_0 \neq \mathcal{N}$ if there are utility functions for which the certainty equivalent is never the lowest over the family of utility functions. See Section 4.4 for an example. Then $D(z,n)$ is the certainty equivalent using utility function $u_n$ and stopping rule $z$; $D_*(z)$ is the CSC value of the stopping rule $z$.

The first idea behind the proof is that we expect the certainty equivalent value of the optimal stopping rule to be constant across the set of (relevant) utility functions. If not, we might expect to be able to improve the certainty equivalent value under the worst utility, at the expense of the certainty equivalent values of those utilities which have a higher certainty equivalent value. This would raise the CSC value. Hence we expect $D(z^*,n)$ is constant on $\mathcal{N}_0$ for the optimal choice $z^*$.

The second idea is that we want there to be only one (randomized threshold) stopping rule for which the certainty equivalent is constant (accross all relevant utilities). This possibility is precluded by a requirement that no stopping rule can achieve a certainty equivalent value which exceeds that of another relevant stopping rule, uniformly across all relevant utilities.

*Proof.* Take $z \in \mathcal{Z}$ and $w$ in $\mathcal{Z}_0$. Suppose $D(z,n) > D(w,n)$ for all $n \in \mathcal{N}_0$. Then $h_n(D(z,n)) > h_n(D(w,n))$ for all $n \in \mathcal{N}_0$ and $\hat{H} \geq \int_{\mathcal{N}_0} \nu(dn) h_n(D(z,n)) > \int_{\mathcal{N}_0} \nu(dn) h_n(D(w,n)) = \hat{H}$ contradicting the hypothesis of the theorem. Hence, for any $z \in \mathcal{Z}, w \in \mathcal{Z}_0$ there exists a non-empty set $\mathcal{N}_{z,w} \subseteq \mathcal{N}_0$ such that $D(z,n) \leq D(w,n)$ on $\mathcal{N}_{z,w}$. Taking $w = z^*$

$$D_*(z) = \inf_{n \in \mathcal{N}} D(z,n) \leq \inf_{n \in \mathcal{N}_{z,z^*}} D(z,n) \leq \inf_{n \in \mathcal{N}_{z,z^*}} D(z^*,n) = \inf_{n \in \mathcal{N}} D(z,n) = D_*(z^*) = \hat{D}.$$

□

*Proof of Theorem 3.* The idea is to apply Proposition 4. To this end take $\mathcal{Z}_0 = \mathcal{Z} = \mathcal{P}(\mathcal{M}), \mathcal{N}_0 = \mathcal{N} = \mathcal{M}$ and set

$$f(\gamma,m) = \mathbb{E}[g_m(Y_{\tau_{0,\gamma}})] = \mathcal{G}^m_{0,\gamma} = \begin{cases} ky & \gamma \in [m_*, m) \\ \frac{\alpha m y}{\gamma} & \gamma \in [m, m^*]. \end{cases}$$

Note that for $\chi \in \mathcal{P}(\mathcal{M})$, $\mathcal{C}^m_\chi = g_m^{-1}\left(\int_{\mathcal{M}} \mathcal{G}^m_{0,\gamma}\chi(d\gamma)\right)$. Take $h_m = g_m$ and $D(\chi,m) = \mathcal{C}^m_\chi$. Then

$$\int \nu(dm) h_m(\mathcal{C}^m_\zeta) = \int \nu(dm)\mathcal{G}^m_\zeta = \int_{\mathcal{M}} \nu(dm) \int_{\mathcal{M}} f(\gamma,m)\zeta(d\gamma) = \int_{\mathcal{M}} \zeta(d\gamma) \int_{\mathcal{M}} f(\gamma,m)\nu(dm).$$

Then by Proposition 4, if we can find $\tilde{\zeta} \in \mathcal{P}(\mathcal{M})$ such that $\mathcal{C}^m_{\tilde{\zeta}}$ does not depend on $m$ and $\nu$ such that $\int_{\mathcal{M}} f(\gamma,m)\nu(dm)$ does not depend on $\gamma$ then $\tilde{\zeta}$ characterizes the optimal mixture of thresholds.

The required conditions follow from the next two lemmas. □

**Lemma 5.** *For $\tilde{\eta}$ as in the statement of Theorem 3, $\mathcal{C}^m_{\tilde{\eta}} = \frac{1}{k}\int_{\mathcal{M}} f(\gamma,m)\tilde{\eta}(d\gamma)$ is independent of $m$.*

*Proof.* It follows from the definition of $C^*$ and $\theta^*$ that

$$1 = \int_{\mathcal{M}} C^* \gamma^{-\frac{\alpha}{\alpha-k}} d\gamma + \theta^*$$

so that $\tilde{\eta}$ is a probability measure on $\mathcal{M}$. Then

$$\begin{aligned}
\frac{1}{ky}\int_{\mathcal{M}} \mathcal{G}^m_\gamma \tilde{\eta}(d\gamma) &= \int_{[m_*,m)} C^*\gamma^{-\frac{\alpha}{\alpha-k}} d\gamma + \int_{[m,m^*)} \frac{\alpha m}{k} C^* \gamma^{-1-\frac{\alpha}{\alpha-k}} d\gamma + \frac{\theta^* \alpha m}{km^*} \\
&= -\frac{\alpha-k}{k} C^* \gamma^{-\frac{k}{\alpha-k}}\Big|_{m_*}^{m} - \frac{\alpha-k}{\alpha}\frac{\alpha m}{k} C^* \gamma^{-\frac{\alpha}{\alpha-k}}\Big|_{m}^{m^*} + \frac{\theta^*\alpha m}{km^*} \\
&= \frac{\alpha-k}{k} C^* m_*^{-\frac{k}{\alpha-k}} - \frac{\alpha-k}{k} C^* m (m^*)^{-\frac{\alpha}{\alpha-k}} + \frac{\theta^*\alpha m}{km^*}
\end{aligned}$$

This does not depend on $m$ since $\theta^* = \frac{\alpha-k}{\alpha} C^* (m^*)^{1-\frac{\alpha}{\alpha-k}}$. □

**Lemma 6.** *Let $\lambda = \frac{\alpha}{k} > 1$. Let $\nu$ be a mixture of an atom of size $\phi = (\lambda^{\frac{\lambda}{\lambda-1}} - \lambda + 1)^{-1}$ at $m_*$ and an absolutely continuous measure $\zeta$ on $\mathcal{M}$ with density $Dm^{\frac{2-\lambda}{\lambda-1}}$ where $D = \frac{\lambda}{\lambda-1}\phi m_*^{-\frac{1}{\lambda-1}}$. Then $\int_{\mathcal{M}} f(\gamma,m)\nu(dm)$ does not depend on $\gamma$.*

*Proof.* Set $\beta = \frac{2-\lambda}{\lambda-1}$. Then $\beta+1 = \frac{1}{\lambda-1}$ and $\beta+2 = \frac{\lambda}{\lambda-1}$ With the absolutely continuous component of $\nu$ having density $Dm^\beta$ we have

$$\begin{aligned}
\frac{1}{y}\int f(\gamma,m)\nu(dm) &= \int_{[m_*,\gamma]} \frac{\alpha m}{k\gamma}\nu(dm) + \int_{(\gamma,m^*]} \nu(dm) \\
&= \frac{\alpha m_*}{k\gamma}\phi + \frac{\alpha}{k\gamma}\int_{(m_*,\gamma]} Dm^{\beta+1} dm + \int_{(\gamma,m^*]} Dm^\beta dm \\
&= \frac{\alpha m_*}{k\gamma}\phi + \frac{\alpha}{k\gamma}\frac{Dm^{\beta+2}}{\beta+2}\Big|_{m_*}^{\gamma} + \frac{Dm^{\beta+1}}{\beta+1}\Big|_{\gamma}^{m^*} \\
&= \frac{m^*}{\gamma}\left[\phi - \frac{Dm_*^{\beta+1}}{(\beta+2)}\right] + \gamma^{\beta+1} D\left[\frac{\alpha}{k(\beta+2)} - \frac{1}{\beta+1}\right] + \frac{D(m^*)^{\beta+1}}{(\beta+1)}
\end{aligned}$$

The two square brackets in this last expression are zero by the choice of $D$ and $\beta$. Hence $\int f(\gamma,m)\nu(dm)$ does not depend on $\gamma$. □

# D Proofs for the generalized stylized example of Section 4.4

*Proof of Theorem 4.* We apply Proposition 4. Let $\mathcal{Z}$ be the set of probability measures on $[0,y)\times[y,\infty]$ and let $\mathcal{Z}_0 \subseteq \mathcal{Z}$ be the set of probability measures with support $\{0\}\times[L,R]$. Let $\mathcal{N} = [L,R]$ and let $\mathcal{N}_0 = \{\alpha : K(\alpha) = J(\alpha)\} \subseteq \mathcal{N}$.

Then $\mathcal{Z}$ is the set of candidate randomizations , and $\mathcal{Z}_0$ is a set of relevant randomizations which are not dominated by some other randomization. $\mathcal{N}$ is a parameterization of the utility functions, and $\mathcal{N}_0$ is a set of utility functions such that no member dominates any other element of $\mathcal{N}$.

Recall the definitions of $\theta$, $\rho$, $\tilde{\eta}$ and $\hat{\eta}$ from the theorem. By the choice of $\theta$, $\tilde{\eta}$ is a probability measure on $[L,R]$. Define $\Delta : [L,R] \mapsto \mathbb{R}$ by

$$\Delta(\alpha) = \frac{\theta}{R} \exp\left( \int_\alpha^R d\beta \frac{K'(\beta)}{K(\beta)-\beta} \right).$$

Then $\Delta$ is differentiable and from the definition of $\rho$ in (13)

$$\Delta'(\alpha) = -\frac{\Delta(\alpha)K'(\alpha)}{(K(\alpha)-\alpha)} = -\frac{\rho(\alpha)}{\alpha}. \tag{A-3}$$

Then, since $\Delta(R) = \frac{\theta}{R}$,

$$\Delta(\alpha) = \frac{\theta}{R} + \int_{(\alpha,R)} \frac{\rho(\beta)}{\beta} d\beta = \int_{(\alpha,R]} \frac{\tilde{\eta}(d\beta)}{\beta}.$$

For $\zeta \in \mathcal{P}([0,y) \times [y,\infty])$ and $\alpha \in \mathcal{N}$ define

$$D(\zeta,\alpha) = \frac{1}{y} u_\alpha^{-1}(\mathbb{E}[u_\alpha(Y_{\tau_\zeta})]) = \frac{1}{y}\mathbb{E}[u_\alpha(Y_{\tau_\zeta})] = \int\int \zeta(d\beta,d\gamma) \frac{u_\alpha(\beta)(\gamma-y)+u_\alpha(\gamma)(y-\beta)}{y(\gamma-\beta)}.$$

For $\hat{\zeta} \in \mathcal{P}(\{0\} \times [L,R])$, this reduces to

$$D(\hat{\zeta},\alpha) = \int \tilde{\zeta}(d\gamma) \frac{u_\alpha(\gamma)}{\gamma}$$

where $\tilde{\zeta} = p(\hat{\zeta})$.

Then, for $\hat{\eta} \in \mathcal{Z}$ as in the statement of the Theorem,

$$\begin{aligned}
\frac{1}{y}\mathbb{E}[u_\alpha(Y_{\tau_{\hat{\eta}}})] &= \tilde{\eta}([L,\alpha)) + J(\alpha) \int_{(\alpha,R]} \frac{\tilde{\eta}(d\beta)}{\beta} \\
&\geq \tilde{\eta}([L,\alpha)) + K(\alpha)\Delta(\alpha) \\
&= \theta + K(L)\Delta(L) + \int_{[L,\alpha)} [\rho(\beta) + \Delta(\beta)K'(\beta) + K(\beta)\Delta'(\beta)]\, d\beta \\
&= \theta + J(L)\Delta(L)
\end{aligned}$$

where we use the first inequality in (A-3) to show that the integrand in the penultimate line is zero.

Then, if $\hat{D} = \theta + J(L)\Delta(L)$ we have for all $\alpha \in \mathcal{N}$,

$$\mathbb{E}[u_\alpha(Y_{\tau_{\hat{\eta}}})] \geq \hat{D}$$

with equality if $\alpha \in \mathcal{N}_0$.

It remains to show that there exists a measure $\nu$ with support in $[L,R]$ and $\hat{H}$ such that $\int_{\mathcal{N}_0} \nu(d\alpha) D(\zeta,\alpha) = \hat{H}$ for $\zeta \in \mathcal{Z}_0$, and $\int_{\mathcal{N}_0} \nu(d\alpha) D(\zeta,\alpha) \leq \hat{H}$ for general $\zeta \in \mathcal{Z}$. (We take $h_\alpha(d) = d$ for all $\alpha \in \mathcal{N}$.)

Recall that $\{z : K(z) = J(z)\} = \cup_{i=1}^N [\ell_i, r_i]$. Let $\nu$ be the measure on $\{z : K(z) = J(z)\}$ such that $\nu$ has atoms of size $\phi_i$ at $\ell_i$ for $i = 1,2,\dots,N$, together with a density $\zeta$ on $\cup_{i=1}^N (\ell_i, r_i)$ given by

$$\zeta(w) = \zeta_i \exp\left( \int_{\ell_i}^w \frac{(2-J'(z))}{(J(z)-z)} dz \right).$$

Here $\phi_1 = 1$ and $\zeta_1 = \frac{\phi_1 J(L)}{L(J(L)-L)}$, and then, proceeding inductively, for $1 \leq i \leq N-1$,

$$\phi_{i+1} = \frac{(\ell_{i+1} - r_i)}{r_i(K(\ell_{i+1}) - \ell_{i+1})} \int_{\alpha \leq r_i} K(\alpha)\nu(d\alpha) \tag{A-4}$$

$$\zeta_{i+1} = \frac{1}{\ell_{i+1}(K(\ell_{i+1}) - \ell_{i+1})} \int_{\alpha \leq \ell_{i+1}} K(\alpha)\nu(d\alpha) \tag{A-5}$$

For any $\tilde{\zeta}$ with support in $[L, R]$ we can define $\hat{\zeta} = p^{-1}(\tilde{\zeta})$. Then

$$\int_{\mathcal{N}_0} \nu(d\alpha) D(\hat{\zeta}, \alpha) = \int_{\mathcal{N}_0} \nu(d\alpha) \int_{[L,R]} \frac{u_\alpha(w)}{w} \tilde{\zeta}(dw) = \int_{[L,R]} \tilde{\zeta}(dw) \left[ \int_{\alpha \leq w} \nu(d\alpha) \frac{K(\alpha)}{w} + \int_{\alpha > w} \nu(d\alpha) \right].$$

First we show that $\Gamma(w) := \int_{\alpha \leq w} \nu(d\alpha) \frac{K(\alpha)}{w} + \int_{\alpha > w} \nu(d\alpha)$ is constant for $w \in \mathcal{N}_0$. For $w \in (\ell_i, r_i)$

$$\Gamma'(w) = \zeta(w) \left[ \frac{K(w)}{w} - 1 \right] - \frac{1}{w^2} \int_{\alpha \leq w} \nu(d\alpha) K(\alpha).$$

We get that $\Gamma$ is constant on $(\ell_i, r_i)$ provided $\Upsilon(w) = 0$ where $\Upsilon(w) = \int_{\alpha \leq w} K(\alpha)\nu(d\alpha) - \zeta(w)w(K(w) - w)$. But $\Upsilon'(w) = \zeta(w)[w(2 - K'(w))] - \zeta'(w)w(K(w) - w) = 0$ by the definition of $\zeta$. Moreover, by the definition of $\zeta_i$ in (A-5) $\Upsilon(\ell_i) = 0$. Hence $\Upsilon \equiv 0$ on $[\ell_i, r_i]$ and $\Gamma(w)$ is constant on this interval.

To prove that $\Gamma$ is constant on $\mathcal{N}_0$ it remains only to show that $\Gamma(r_i) = \Gamma(\ell_{i+1})$. We have

$$\begin{aligned} \Gamma(\ell_{i+1}) - \Gamma(r_i) &= \frac{1}{\ell_{i+1}} \left[ \int_{\alpha \leq r_i} \nu(d\alpha) K(\alpha) + K(\ell_{i+1})\phi_{i+1} \right] - \frac{1}{r_i} \left[ \int_{\alpha \leq r_i} \nu(d\alpha) K(\alpha) \right] - \phi_{i+1} \\ &= \phi_{i+1} \left[ \frac{K(\ell_{i+1})}{\ell_{i+1}} - 1 \right] - \left( \frac{1}{r_i} - \frac{1}{\ell_{i+1}} \right) \int_{\alpha \leq r_i} \nu(d\alpha) K(\alpha) = 0 \end{aligned}$$

by the definition of $\phi_{i+1}$ in (A-4).

Finally, we consider general $\eta \in \mathcal{P}([0, y) \times [y, \infty])$. Recall the definition of $\hat{\zeta} = P(\eta)$. From Lemma 4, $\mathcal{G}^m_\eta \leq \mathcal{G}^m_{P(\eta)}$ for all $m \in [L, M]$. Then $D(\eta, m) \leq D(P(\eta), m)$ for all $x$ and

$$\int_{\mathcal{N}_0} \nu(dm) D(\eta, m) \leq \int_{\mathcal{N}_0} \nu(dm) D(P(\eta), m) = \hat{H}.$$

□