JOURNAL of
MATERIALS and MECHATRONICS:A

e-ISSN 2717-8811
JournalMM, 2023, 4(1), 116-133
https://doi.org/10.55546/jmm.1248851

Araştırma Makalesi / Research Article

Optimization of Generalized Certainty Equivalents on the Finite Horizon

Arnab BASU[1]*

[1] Bilkent University, Department of Industrial Engineering, Anakara, Turkey,
ORCID ID: https://orcid.org/0000-0002-4309-302X, arnab@bilkent.edu.tr

Geliş/ Received: 07.02.2023; **Kabul / Accepted:** 06.03.2023

ABSTRACT: This paper addresses some open issues in optimization of generic certainty equivalents. Such equivalents have been modelled using increasing functionals of the discounted sums of the per-stage unbounded-above cost or reward functions defined on the paths of the underlying controlled Markov chain on general state spaces which models the random dynamics of the system. Examples of such functionals include logarithmic and power utilities as well as the robust Risk-Sensitive preferences among others. The critical results that were obtained were the solutions of this problem for generic unbounded-above per-stage cost minimization and for per stage reward maximization, both satisfying a w-growth (hence unbounded) condition in the nite horizon setup. In the process, we establish certain nontrivial closure properties of the dynamic programming operators. In addition, we provide a real-life example from Portfolio Consumption.

Keywords: Certainty Equivalents, Full-Path Accumulated Setup, Unbounded Rewards and Costs, Discounted Path Sums, Robust or Risk-Sensitive Preferences.

*Sorumlu yazar / Corresponding author: arnab@bilkent.edu.tr

Bu makaleye atıf yapmak için /To cite this article

Basu, A., (2023). Optimization of Generalized Certainty Equivalents on the Finite Horizon. Journal of Materials and Mechatronics: A (JournalMM), 4(1), 116-133.

1. INTRODUCTION

The main motivation of this paper is to address some open issues in optimization of generic (without assuming any specific structure) certainty equivalents which are economic quantities defining the continuation utilities (or costs) in any random dynamic optimization problem from a given time (or stage/period) onwards. Such certainty equivalents have been modelled using increasing functionals of the discounted sums of the generic (only measurability but no continuity assumption) per-stage unbounded above cost or reward functions defined on the paths of the underlying controlled Markov chain on general (Polish) state and action spaces which models the random dynamics of the system. Hence, it includes all usual types of such functions, in particular, the robust risk-sensitive (see, e.g., Hansen and Sargent, 1995; Howard and Matheson, 1972) as well as the usual logarithmic, power etc. functions.

The notion of risk-sensitive preferences in Markov Decision Processes (MDPs) was formally introduced in the seminal paper (Howard and Matheson, 1972) using a stochastic control formulation approach (see also, e.g., (Di Masi and Stettner, 1999), (Jacobson, 1973), (Whittle, 1990) and references therein) though it was first studied by Bellman (Bellman, 1957) (see p. 329 therein). In this criterion, one investigates 'exponential of integral' cost or reward which takes into account the attitude of the controller with respect to risk. In an uncertain situation when the total payoff Z is random, one wishes to select a concave (increasing) utility function $U\ (\cdot)$ for reward maximization (convex for cost minimization), depending on a parameter $\theta > 0$ representing the degree of risk attraction or aversion, which attains the "certainty equivalence", i.e., $EU_\theta\ (Z) = U_\theta\ (EZ)$. This forces $U_\theta\ (z) = \frac{1}{\theta}\ e^{\theta z}$ (see, e.g., Nowak (Nowak, 2005)), wherein the criterion under study becomes of the form $= \frac{1}{\theta}\ln E[e^{\theta Z}]$. This quantity $\theta > 0$ is called the 'absolute risk-aversion parameter'.

In the recent times, several alternatives of risk measurement have come up based on generic certainty equivalents $U^{-1}E[U(Z)]$ given general (increasing) forms of the path functional $U(\cdot)$. We note that the use of certainty equivalents goes back about a hundred years to the 1930s (see, e.g., (Muliere and Parmigiani, 1993)). A study of the generalization of risk-sensitive preferences in MDP setup to generic certainty equivalents in the above sense was recently undertaken in (Bäuerle and Rieder, 2014) by implementing the corresponding utility as a 'full-path accumulated' functional of the underlying random dynamical system, namely, as a certainty equivalent of the time-separable additive von Neumann-Morgenstern discounted utility on the entire (finite or infinite) time horizon. Therein, the authors considered a discrete-time MDP on a Borel state space with bounded one-stage costs over the finite time horizon with general increasing $U(\cdot)$ and on the innite horizon with $U(\cdot)$ increasing and either concave or convex.

On the other hand, the economics literature followed a 'per-stage recursive' robustness approach using the notion of risk-sensitivity as formulated by (Hansen and Sargent, 1995) (see also (Hansen et al., 2006), (Hansen and Sargent, 2007) and (Hansen and Sargent, 2008)), namely, as a sum of the current (per-stage) reward and the one-step discounted certainty equivalent of the future (next-stage) continuation utility wherein this certainty equivalent is computed using a similar concave transformation. The generic form of such a recursive approach was first introduced by (Kreps and Porteus, 1978) and (Kreps and Porteus, 1979) for the infinite horizon and later extended by (Epstein and Zin, 1989) to the infinite horizon. Using the terminology introduced in the recent paper (Basu et al., 2022), we call the robustness framework of (Hansen and Sargent, 1995) as the "per-stage recursive" approach whereas, again as therein, the corresponding original framework of (Howard and Matheson, 1972) is referred to as the "full-path accumulated" approach. These per-stage recursive preferences have been studied extensively in recent years primarily in the context of their connection to robustness analysis for economic model misspecifications (see, e.g., (Anderson et al., 2012; Hansen and Sargent, 2007)), precautionary savings for life-cycle preferences under monotone risk aversion (see, e.g., (Bommier, 2013; Bommier et al., 2017; Bommier and Le Grand, 2019)), climate risk modelling for optimal carbon control (see, e.g., (Bommier et al., 2015; Cai and Lontzek, 2019)) and risk measures in asset pricing/portfolio optimization (see, e.g., (Bäuerle and Jaskiewicz, 2018; Bielecki and Pliska, 2003; Epstein and Zin, 1991; Tallarini, 2000)) among other areas.

Whereas the applications of the per-stage recursive framework have been so diverse and many, the solutions obtained in a vast majority of the above cases (which mainly involved nonlinear certainty equivalents) have been somewhat heuristic and sometimes without formal mathematical justification as to their existence and uniqueness. Whatever formal methods had been deployed were majorly for the linear case which lead to the standard additive von Neumann-Morgenstern preference and the generic nonlinear cases were left open, e.g., preference maximization problems for multiplicative or exponential functionals. Only recently, some formal results in this direction were obtained for a specific problem of stochastic optimal growth (see (Bäuerle and Jaskiewicz, 2018)) under certain growth and stability assumptions on the system. However, to the best of our knowledge, the generic nonlinear case is still largely open.

The only work that endeavoured to address this nonlinear setup albeit in the full-path accumulated setup for cost minimization is the paper by (Bäuerle and Rieder, 2014). This present paper is in some sense the technical successor to (Bäuerle and Rieder, 2014) and aims to address the some of the open issues mentioned therein. Notably, (Bäuerle and Rieder, 2014) made certain strong assumptions on the boundedness of per-stage cost or reward function which usually fail to hold in the real-life scenarios described in the above references. Our paper attempts to address these issues and create an extended optimization theory for certainty equivalents modelled using generic increasing functionals of the discounted path-sums of the unbounded-above per-stage functions defined on general state spaces.

Our main technical contribution is to solve this optimization problem in the finite horizon case for per-stage unbounded costs/rewards satisfying a w-growth condition (see Theorem 3.2 and Corollary 3.3 in Section 3).In the process, a result of much broader (beyond our setup here) potential applicability have been proved, namely, certain important closure properties of the dynamic programming (Bellman) operators, the compactness of optimal action (sub)sets and the corresponding existence of optimizing selectors in Theorem 3.1 in Section 3. We also provide a real-life example from Portfolio Consumption Model to illustrate our ideas (see Proposition 4.3 in Section 4). We choose such a portfolio model as an example because such models are extremely important in business and finance and have regularly been used as canonical examples in certainty equivalent formulations. Hence, we believe that such an example would best serve our purpose of illustration.

We now describe the structure of the paper. Section 2 describes the problem setup as well as clarifies the main technical formulations, notations and assumptions. Section 3 proves a very crucial result, namely Theorem 3.1 involving the closure properties of the underlying dynamic programming operator and then proceeds to the proves the finite horizon results in Theorem 3.2 and Corollary 3.3 for unbounded per-stage functions satisfying a w-growth condition. Section 4 provides a nontrivial application of the obtained results to a real-life Portfolio Consumption problem (see Proposition 4.3). We end with Section 5 which provides some pointers to future directions of work as technical continuation of this work and, in the process, discuss some limitations of this paper.

2. PROBLEM DESCRIPTION

For a metric space S, let $\mathcal{B}$ (S) denote the Borel σ-algebra on S, $C_b(S)$ the set of bounded continuous real functions on S endowed with the uniform/sup norm $\|\cdot\|$, $\mathcal{P}$ (S) denote the set of probability measures on $(S, \mathcal{B}(S))$ endowed with the Prohorov topology (see, e.g., (Borkar, 1995)), i.e., the topology of weak convergence implying, a sequence $\mathcal{P}(S) \supset \{\mu_n\}_{n\geq 1} \overset{w}{\to} \mu \in \mathcal{P}(S)$,

$$\int_S f(x)\mu_n(dx) \overset{n\uparrow\infty}{\longrightarrow} \int_S f(x)\mu(dx),\ \forall f \in \mathcal{C}_b(S). \tag{2.1}$$

Defining $\bar{\mathbb{R}} \overset{def}{=} \mathbb{R} \cup \{\pm\infty\}$, $\mathbb{R}_+ \overset{def}{=} \{x \in \bar{\mathbb{R}} : x \geq 0\}$, $\mathbb{R}_- \overset{def}{=} \{x \in \bar{\mathbb{R}} : x \leq 0\}$, let $f : Y \mapsto \bar{\mathbb{R}}$ for a Borel subset Y of some Polish space. For $\lambda \in \bar{\mathbb{R}}$, consider the level-sets

$$\underline{\mathcal{D}}_f(\lambda; Y) \overset{def}{=} \{y \in Y : f(y) \leq \lambda\}, \qquad \overline{\mathcal{D}}_f(\lambda; Y) \overset{def}{=} \{y \in Y : f(y) \geq \lambda\}. \tag{2.2}$$

We call f lower-semicontinuous (l.s.c) on Y if all such level-sets $\underline{D}$ are closed and inf-compact on Y if they are compact. Similarly, f is upper-semicontinuous (u.s.c) if all such $\overline{D}$ are closed and sup-compact if they are compact. Note that if f is l.s.c (resp. inf-compact) then $-f$ is u.s.c (resp. sup-compact). Consider a discrete-time Markov Decision Process (MDP) with state space $\mathcal{X}$ (Borel subset of a Polish space), action space A (Borel subset of a Polish space) with $\emptyset \neq \mathsf{A}(x) \subseteq \mathcal{A}$ being the Borel set of actions available at $x \in \mathcal{X}$, one-step measurable bounded-below cost function: $(x, a) \in Gr(\mathcal{A}) \overset{def}{=} \{(x, a) : x \in \mathcal{X}, a \in \mathcal{A}(x)\} \mapsto c(x, a) \leq +\infty$, one-step measurable bounded below reward/utility function r: $(x, a) \in Gr(\mathcal{A}) \mapsto r(x, a) \leq +\infty$, and, weakly-continuous transition kernels (laws) $q\ (\cdot\ |x;\ a) \in P(\mathcal{X})$ i.e.

$$\mathcal{X} \times \mathcal{A} \ni (x, a) \mapsto \int_{\mathcal{X}} f(x')q\,(dx'|x, a) \text{ is continuous } \forall f \in \mathcal{C}_b(\mathcal{X}) \tag{2.3}$$

Such that $q(B|\cdot,\cdot)$ is Borel on Gr (A) for all $B \in \mathcal{B}(\mathcal{X})$. We assume, as usual, that $Gr(\mathsf{A})$ is a measurable subset of $\mathcal{X}$ x A i.e. Gr $(\mathsf{A}) \in \mathcal{B}(\mathcal{X}$ x $\mathsf{A})$. We say that a sequence of real-valued functions $\{f_n\}_{n=1,2,\ldots}$ defined on $\mathcal{X}$ is lower semiequicontinuous at a point $x \in \mathcal{X}$ if for each $\epsilon > 0$ there exists a $\delta_\epsilon > 0$ such that for all points $x' \in B_{\delta_\epsilon}(x)$ (δ_ϵ-neighborhood of the point x) we have that $f_n(x') > f_n(x) - \epsilon$ for all $n = 1, 2, \ldots$ (Feinberg et al., 2020, Definition 3.3). This sequence $\{f_n\}_{n=1,2,\ldots}$ is said to be lower semiequicontinuous on $\mathcal{X}$ if it is so at all $x \in \mathcal{X}$. Analogously, a sequence of real-valued functions $\{f_n\}_{n=1,2,\ldots}$defined on $\mathcal{X}$ is called upper semiequicontinuous at a point $x \in \mathcal{X}$ if the sequence $\{-f_n\}_{n=1,2,\ldots}$ is lower semiequicontinuous at that point $x \in X$. Also, we say that a sequence of measurable function $\{f_n: \mathcal{X} \to \overline{\mathbb{R}}\}_{n=1,2,\ldots}$ lower semiconverges (see (Feinberg et al., 2020), Definition 3.2) to a measurable function f_n:$\mathcal{X} \to \overline{\mathbb{R}}$ in some measure μ on $\mathcal{B}(\mathcal{X})$ if, for each $\epsilon > 0$, we have,

$$\lim_{n\to\infty} \mu\left(\{x \in \mathcal{X} : f_n(x) \leq f(x) - \epsilon\}\right) = 0. \tag{2.4}$$

Accordingly, upper semiconvergence is defined as the situation when $\{-f_n\}_{n=1,2,\cdots}$ lower semiconverges to $-f$ or equivalently when,

$$\lim_{n\to\infty} \mu\left(\{x \in \mathcal{X} : f_n(x) \geq f(x) + \epsilon\}\right) = 0. \tag{2.5}$$

The decision system moves as follows:

1. at each time $t \in \mathbb{N}_0 \stackrel{\text{def}}{=} \{0,1,\dots,\}$ the current state $x \in \mathcal{X}$ is observed,
2. the decision maker chooses an action $a \in A(x)$,
3. the cost $c(x,a)$ or reward $r(x,a)$ is incurred, and,
4. the system moves to the next state according to probability $q\ (\cdot\ |x,a)$ at the next time $t\ +1$ and repeats itself till finite or infinite time $0 < T \leq \infty$.

We denote the set of histories up to and including time t as $\mathcal{H}_t \stackrel{\text{def}}{=} (\mathcal{X}+\mathcal{A})^t \times \mathcal{X}$ with $\mathcal{B}(\mathcal{H}_t) = (\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}\ (A)^t \otimes \mathcal{B}\ (\mathcal{X})$. A history-dependent (behavioural) policy $\pi \stackrel{\text{def}}{=} \{\pi_t\}_{t\in\mathbb{N}_0}$ is a sequence of (measurable) maps $\pi_t : \mathcal{H}_t \to A$ such that $\pi_t(h_t) \in A(x_t)$ *where* $\mathcal{H}_t \ni h_t \stackrel{\text{def}}{=} \{x_0, a_0, x_1, a_1, \dots x_{t-1}, a_{t-1}, x_t\}$ is a history up to and including time, t. Let Π denote the set of all such policies. A policy $\phi \stackrel{\text{def}}{=} \{\phi_t\}_{t\in\mathbb{N}_0} \in \prod$ is called Markov if all decisions depend on the current state only i.e. each $\emptyset_t$: $\mathcal{X} \to A$ and is measurable such that $\phi_t(x) \in A(x)$ for all $x \in \mathcal{X}$. Denote by Π^M the set of Markov policies. A Markov policy ϕ is called stationary (Markov) if $\phi_t \equiv \phi$ (repeating notation) for some measurable map $\phi : \mathcal{X} \to \mathcal{A}$ such that $\phi(x) \in A\ (x)$ for all $x \in \mathcal{X}$. Denote by Π^S the set of stationary policies. The Ionescu-Tulcea theorem (see, e.g., (Bertsekas and Shreve, 1996), Proposition 7.28) implies that, for each initial state $x \in \mathcal{X}$ and policy $\pi \in \Pi$, there exists an unique probability measure P_x^π and a stochastic process (X_t, A_t) on the set of all histories (trajectories) $\mathcal{H}_\infty \stackrel{\text{def}}{=} (X \times A)^\infty$ endowed with the product σ-algebra such that $X_t(h_\infty) = x_t$ is the (random) state and $A_t(h_\infty) = a_t$ is the (random) action at time t for $h_\infty \in \mathcal{H}_\infty$. Denote by E_x^π the corresponding expectation.

Let $U: \mathbb{R}_+ \to \overline{\mathbb{R}}$ be continuous, unbounded and strictly increasing function such that U^{-1} exists. This function U shall be used to define overall costs or utilities in our setup as described below. For a finite-horizon $T \in \mathbb{N}_0$ and under a policy $\pi \in \Pi$ define the general certainty equivalent cost/utility criterion as

$$V_{T,\beta}^\pi(x) \stackrel{def}{=} U^{-1}\left(E_x^\pi\left[U\left(F_\beta^T\right)\right]\right),\ x \in \mathcal{X}, \tag{2.6}$$

where $\beta \in (0,1]$ is the discount factor, $F_\beta^t \equiv C_\beta^t \stackrel{\text{def}}{=} \sum_{s=0}^{t-1} \beta^s c(X_s, A_s)$ for cost and $F_\beta^t \equiv R_\beta^t \stackrel{\text{def}}{=} \sum_{s=0}^{t-1} \beta^s r(X_s, A_s)$ for reward for $t \geq 1$. In particular, for our finite horizon case $t = T$. For the undiscounted case, i.e. $\beta = 1$, we write $V_{T,\beta=1}^\pi \equiv V_T^\pi$. Similarly, for the infinite horizon $(T = \infty)$, define the corresponding cost/utility criteria as,

$$V_\beta^\pi(x) \stackrel{def}{=} U^{-1}\left(E_x^\pi\left[U\left(F_\beta^\infty\right)\right]\right),\ x \in \mathcal{X}, \tag{2.7}$$

where $F_\beta^\infty \overset{\text{def}}{=} C_\beta^{T=\infty} = \sum_{t=0}^{\infty} \beta^t c(X_t, A_t)$ for cost, $F_\beta^\infty \overset{\text{def}}{=} R_\beta^{T=\infty} = \sum_{t=0}^{\infty} \beta^t r(X_t, A_t)$ for reward. For any of the above costs $\underline{g}^\pi(x) = V_{T,\beta}^\pi(x)$ (without β for $\beta = 1$) or $V_\beta^\pi(x)$ define the optimal cost as

$$\underline{g}(x) \overset{def}{=} \inf_{\pi \in \Pi} \underline{g}^\pi(x), \; x \in \mathcal{X}. \tag{2.8}$$

Similarly, for any of the above utilities $\overline{g}^\pi(x) = V_{T,\beta}^\pi(x)$(without β for $\beta = 1$) or $V_\beta^\pi(x)$ define the optimal utility as

$$\overline{g}(x) \overset{def}{=} \sup_{\pi \in \Pi} \overline{g}^\pi(x), \; x \in \mathcal{X}. \tag{2.9}$$

A policy $\pi^* \in \Pi$ is said to be optimal for the corresponding cost (resp. utility) criterion if $\underline{g}^{\pi^*}(x) = \underline{g}^\pi(x)$. (resp. $\overline{g}^{\pi^*}(x) = \overline{g}(x)$) for all $x \in \mathcal{X}$. The following Remark 1 is of relevance here.

Remark 1. Given a strategy $\pi \in \Pi$ with an initial state $x \in \mathcal{X}$ and a real-valued random variable Y defined on the probability space $(\mathcal{H}_\infty, B(\mathcal{H}_\infty), P_x^\pi)$) and assuming sufficient regularity for U as given above, we see, by Taylor Expansion, that

$$U^{-1}\left(E_x^\pi\left[U(Y)\right]\right) \approx E_x^\pi[Y] - \frac{1}{2}\mathcal{AP}\left(E_x^\pi[Y]\right) Var[Y]| \tag{2.10}$$

where $\mathcal{AP}(y) \overset{\text{def}}{=} -\frac{U''(y)}{U'(y)}$ is the 'Arrow-Pratt' function of absolute risk-aversion assuming that U is sufficiently regular. Note that $\mathcal{AP}$ term above defines the variability (see, e.g., (Bielecki and Pliska, 2003)). If U is concave then the variance is subtracted and the decision-maker becomes risk-seeking for cost minimization instead becoming risk-averse if U is convex as, in this case, the variance is added. These interpretations are reversed for reward maximization problems.

The point-to-set map $\mathcal{X} \ni x \to \mathcal{A}(x)$ is called upper-semicontinuous in the sense of Kuratowski (u.s.c-K) if $\mathcal{X} \ni x_n \to x \in \mathcal{X}$ and $a_n \in \mathcal{A}(x_n)$ for $n \in \mathbb{N}$ then $\{a_n\}_{n\in\mathbb{N}}$ has a limit point $a \in \mathcal{A}(x)$. Before proceeding further, we make an assumption on the one-step cost / reward function $c(\cdot,\cdot), r(\cdot,\cdot)$ below.

Assumption (A):

1. There exists a measurable function w : $\mathcal{X} \to [\underline{\ell}, \infty)$, a measurable subset $C \subset \mathcal{X}$, constants $0 < \rho < 1, 0 < \underline{\ell} \le 1, c' \ge \underline{\ell}, c'(1-\rho) \ge b > 0$ and $d \ge \frac{c'^2}{\underline{\ell}}$ such that,

$$\int_{\mathcal{X}} w(x')q\left(dx'|x,a\right) \le \rho w(x) + b\mathbf{1}_C(x), \;(x,a) \in Gr(\mathcal{A})$$
$$\text{with} \sup_{x\in C} w(x) = c' < \infty \text{ and } \inf_{x\in C} w(x) = \max\{\frac{c'+d}{(1-\rho)\frac{d}{b}+1}, \underline{\ell}\} \le c'. \tag{2.11}$$

2. For each $x \in \mathcal{X}$ there exists some $a \in \mathcal{A}(x)$ such that $c(x,a) < \infty$ i.e. the situation $c(x,a) = \infty$ for all $(x,a) \in Gr(\mathcal{A})$ is excluded as it is trivial in that all actions are bad.

3. The map $\mathcal{X} \ni x \to \mathcal{A}(x) \subset \mathcal{A}$ is u.s.c and compact-valued.

4. This part has two alternatives:

   (a) For the cost minimization problem (2.8) : c is l.s.c and for all $(x, a) \in \mathrm{Gr}(\mathsf{A})$, $c(x, a) \geq 0$ and $c(x, a) \leq dw(x)$ for d as in Assumption A(1) below.

   (b) For the utility maximization problem (2.9): r is u.s.c and for all $(x, a) \in Gr(\mathsf{A})$, $r(x, a) \geq 0$ and $r(x, a) \leq dw(x)$ where, without loss of generality, d is same as above.

We state the following Remark in regard to the above Assumption (A).

Remark 2. This Assumption (A) is fairly standard in the literature for general state and action spaces (see, e.g, (Bäuerle and Rieder, 2014; (Bäuerle and Jaskiewicz, 2018)) and can usually be shown to be satised in all real-life applications. Assumption A(1) above, referred to as the "drift inequality" (stochastic Lyapunov-type stability condition), and A(4) above, referred to as the "w-growth condition" on the cost or reward function, have been widely used in the MDP and economic literature, e.g., (Altman et al., 1997; Bäuerle and Jasckiewicz, 2018; Meyn and Tweedie, 2009; Nowak and Altman, 2002). Note it is easy to see that, without loss of generality, we can choose the same set of parameters (namely, ρ, b, c', d), the same set C and the same growth function $w(\cdot)$ for both c and r in Assumption (A) above. Finally, it should be noted that the following always holds true for any $d > 0$:

$$\frac{c' + d}{(1-\rho)\frac{d}{b} + 1} \begin{cases} = c' & 0 < b = c'(1-\rho), \\ < c' & 0 < b < c'(1-\rho). \end{cases} \tag{2.12}$$

For a measurable function f: $\mathcal{X} \times \mathsf{A} \to \mathbb{R}_+$ we define the w-norm as follows:

$$||f||_w \stackrel{def}{=} \sup_{x \in \mathcal{X}, a \in \mathcal{A}(x)} \frac{f(x,a)}{w(x)}. \tag{2.13}$$

Note that, in this notation above, by Assumption A(4), we have $\|c\|_w, \|r\|_w \leq d$. As explained in the text following the proof of Lemma 2 in (Feinberg, et al., 2012), it follows that there exists a measurable map $\phi : \mathcal{X} \to A$ such that $\phi(x) \in \mathsf{A}(x)$ for all $x \in \mathcal{X}$ i.e. there exists a (measurable) selector. Since, as explained above, we identify stationary policies with a selector they are the same objects and thus the existence of a selector is a necessary and sufficient condition for the existence of a policy. Again, as explained therein, it is possible that $c(x, a) = \infty$ for some $(x, a) \in Gr(\mathsf{A})$ and then, from a modelling perspective, this state-action pair should be excluded as there is no further reason to optimize once such a pair is reached/selected. So, we can exclude such a situation when c is identically ∞ as we did in Assumption A(2).

We now proceed to the solutions of (2.8) and (2.9) for the various cost or reward functions (as described above) in the next section.

3. FINITE HORIZON PROBLEM

In this section we consider the optimization problems (2.8) and (2.9) for the cost/utility function (2.6) with $\beta \in (0, 1)$. Since U is strictly increasing so is U^{-1} and we can consider the problems (2.8) and (2.9) (resp.) by removing U^{-1} from (2.6) and considering the following equivalent problems (resp.).

$$\underline{v}_{T,\beta}(x) \stackrel{def}{=} \inf_{\pi\in\Pi} E_x^\pi\left[U\left(C_\beta^T\right)\right], \qquad \overline{v}_{T,\beta}(x) \stackrel{def}{=} \sup_{\pi\in\Pi} E_x^\pi\left[U\left(R_\beta^T\right)\right] \tag{3.1}$$

for $x \in \mathcal{X}$. To proceed, we define (as in (Bäuerle and Rieder, 2014)) a MDP on the enhanced state space $\hat{X} \stackrel{\text{def}}{=} \mathcal{X}\ x\ \mathbb{R}_+ \times (0,1]$, corresponding per-stage cost zero and action space A with the selectors now being measurable maps $\phi: \hat{X} \to \mathsf{A}$ such that $\hat{X} \ni (x,y,z) \to \phi(x,y,z) \in \mathsf{A}(x)$. We denote by Φ the set of selectors. Policies are then defined as in the previous section. For $t \in \mathbb{N}$ we define, for $(x,y,z) \in \hat{X}$

$$\underline{v}_{t\pi}(x,y,z) \stackrel{def}{=} E_x^\pi\left[U\left(zC_\beta^t + y\right)\right],\ \pi\in\Pi, \qquad \underline{v}_t(x,y,z) \stackrel{def}{=} \inf_{\pi\in\Pi} \underline{v}_{t\pi}(x,y,z),$$

and

$$\overline{v}_{t\pi}(x,y,z) \stackrel{def}{=} E_x^\pi\left[U\left(zR_\beta^t + y\right)\right],\ \pi\in\Pi, \qquad \overline{v}_t(x,y,z) \stackrel{def}{=} \sup_{\pi\in\Pi} \overline{v}_{t\pi}(x,y,z) \tag{3.2}$$

with terminal cost / utility $\underline{v}_0(x,y,z) = \overline{v}_0(x,y,z) = v_0(x,y,z) \equiv U(y)$ where, as in (Kreps, 1977a) and (Kreps, 1977b), y summarizes the cost or reward that has been accumulated so far and z is a new state variable introduced to keep track of discounting. The corresponding transition kernel $\tilde{q}(d(x',y',z')|x,y,x,a) \equiv q(dx'|x,a) \otimes 1_{(zf(x,a)+y)}(dy') \otimes 1_{z\beta}(dz')$ is defined as:

$$\begin{aligned}
&\int_{\hat{\mathcal{X}}} v(x',y',z')\tilde{q}\left(d(x',y',z')|x,y,z,a\right)\\
&= \int_{\hat{\mathcal{X}}} v(x',y',z')q\left(dx'|x,a\right) \otimes \mathbf{1}_{(zf(x,a)+y)}(dy') \otimes \mathbf{1}_{z\beta}(dz')\\
&= \int_{\mathcal{X}} v(x', zc(x,a)+y, z\beta)q\left(dx'|x,a\right)
\end{aligned} \tag{3.3}$$

where $f = c, r$ as the case may be.

We wish to obtain $v_T(x,0,1) = \underline{v}_{T,\beta}(x), \bar{v}_{T,\beta}(x)$ with the cost/utility interpretation as required. To this end, we define,

$$\mathcal{W} \stackrel{def}{=} \Big\{f : \hat{\mathcal{X}} \mapsto \mathbb{R}_+ :: f \text{ is measurable and } f(x,y,z) \le (d + b\mathbf{1}_{\mathcal{C}}(x))w(x)$$
$$\text{for } (x,y,z) \in \hat{\mathcal{X}}\Big\},$$
$$\mathcal{L}'(\hat{\mathcal{X}}) \stackrel{def}{=} \Big\{f : \hat{\mathcal{X}} \mapsto \mathbb{R}_+ :: f \text{ is l.s.c, } f(x,\cdot,\cdot) \text{ is continuous, non-decreasing for } x \in \mathcal{X}$$
$$\text{and } f(x,y,z) \ge U(y) \text{ for } (x,y,z) \in \hat{\mathcal{X}}\Big\},$$
$$\mathcal{L}(\hat{\mathcal{X}}) \stackrel{def}{=} \mathcal{L}'(\hat{\mathcal{X}}) \bigcap \mathcal{W},$$
$\mathcal{U}'(\hat{\mathcal{X}}) \stackrel{def}{=} \Big\{f : \hat{\mathcal{X}} \mapsto \mathbb{R}_+ :: f \text{ is u.s.c, } f(x,\cdot,\cdot)$ is continuous, non-decreasing for x $\in \mathcal{X}$
$$\text{and } f(x,y,z) \ge U(y) \text{ for } (x,y,z) \in \hat{\mathcal{X}}\Big\},$$

$$\mathcal{U}(\hat{\mathcal{X}}) \overset{def}{=} \mathcal{U}'(\hat{\mathcal{X}}) \bigcap \mathcal{W}. \tag{3.4}$$

Note that, without loss of generality, we can choose the same w, d as in Assumption (A) above. For $v \in \mathcal{L}'(\hat{\mathcal{X}})$ or $U'(\hat{\mathcal{X}})$, $\phi \in \Phi, (x, y, z) \in \hat{\mathcal{X}}$ and $a \in \mathcal{A}$ define

$$\begin{aligned} \eta_f[v](x,y,z,a) &\overset{def}{=} \int_{\mathcal{X}} v(x', zf(x,a)+y, z\beta) q\,(dx'|x,a)\,, \;\; f=c,r, \\ Gr_{\hat{\mathcal{X}}}(\mathcal{A}) &\overset{def}{=} \{(x,y,z,a) : (x,y,z) \in \hat{\mathcal{X}}, a \in \mathcal{A}(x)\} \subseteq \hat{\mathcal{X}} \times \mathcal{A} \end{aligned} \tag{3.5}$$

and

$$\begin{aligned} \mathbf{T}^f_{\phi}[v](x,y,z) &\overset{def}{=} \eta_f[v](x,y,z,\phi(x,y,z)), \;\; f=c,r \\ \underline{\mathbf{T}}[v](x,y,z) &\overset{def}{=} \inf_{a\in\mathcal{A}(x)} \eta_c[v](x,y,z,a), \;\; v \in \mathcal{L}'(\hat{\mathcal{X}}), \\ \overline{\mathbf{T}}[v](x,y,z) &\overset{def}{=} \sup_{a\in\mathcal{A}(x)} \eta_r[v](x,y,z,a), \;\; v \in \mathcal{U}'(\hat{\mathcal{X}}). \end{aligned} \tag{3.6}$$

Remark 3. For $\mathbf{T} = \underline{\mathbf{T}}, \overline{\mathbf{T}}$ or $\mathbf{T} \in \{\mathbf{T}^f{}_{\phi}, \phi \in \Phi\}$are monotone operators in the sense that for $v_1, v_2 \in \mathcal{L}'(\hat{\mathcal{X}})$ or $U'(\hat{\mathcal{X}})$ with $v_1 \le v_2$ it is true that $\mathbf{T}[v_1] \le \mathbf{T}[v_2]$.

We now state and prove a very important result which shall form one of the key foundations of this paper.

Theorem 3.1. Under the Assumption (A) in Section 2 above, we have,

1. Given $v \in U(\hat{\mathcal{X}})$, $\overline{\mathbf{T}}v \in U(\hat{\mathcal{X}})$ the supremum in (3.6) can be replaced by a maximum, i.e.

$$\overline{\mathbf{T}}[v](x,y,z) = \max_{a\in\mathcal{A}(x)} \eta_r[v](x,y,z,a), \tag{3.7}$$

the set of maximizers $\mathcal{A}^*(x, y, z) \overset{def}{=} \{a \in \mathcal{A}(x) : \eta r\,[v](x,y,z,a) = \overline{\mathbf{T}}\,[v](x,y,z)\}$ is nonempty compact and there exists a corresponding maximizing selector i.e. $\phi^* \in \Phi$ such that,

$$\overline{\mathbf{T}}[v](x,y,z) = \mathbf{T}^r_{\phi^*}[v](x,y,z) = \eta_r[v](x,y,z,\phi^*(x,y,z)), \;\; (x,y,z) \in \hat{\mathcal{X}}. \tag{3.8}$$

2. Given $v \in \mathcal{L}(\hat{X}), \underline{\mathbf{T}}[v] \in \mathcal{L}(\hat{X})$, the infimum in (3.6) can be replaced by a minimum i.e.

$$\underline{\mathbf{T}}[v](x,y,z) = \min_{a\in\mathcal{A}(x)} \eta_c[v](x,y,z,a), \tag{3.9}$$

the set of minimizers $\mathcal{A}^*(x, y, z) \overset{def}{=} \{a \in \mathcal{A}(x) : \eta_c\,[v]\,(x,y,z,a) = \underline{\mathbf{T}}[v](x,y,z)$ is nonempty compact and there exists a corresponding minimizing selector i.e. $\phi^* \in \Phi$ such that

$$\underline{\mathbf{T}}[v](x,y,z) = \mathbf{T}^c_{\phi^*}[v](x,y,z) = \eta_c[v](x,y,z,\phi^*(x,y,z)), \;\; (x,y,z) \in \hat{\mathcal{X}}. \tag{3.10}$$

Proof: We prove Part 1 of the above Theorem 3.1. The proof of Part 2 is exactly analogous. Let us prove that the function $\eta_r[v](x,y,z,a)$ is u.s.c. Consider a converging sequence $(x_n, y_n, z_n, a_n) \to (x,y,z,a)$ as $n\to\infty$. Let $v \in U(\hat{\mathcal{X}})$. We define the functions $f(\cdot) \stackrel{\text{def}}{=} v(\cdot, zr(x,a) + y, z\beta)$ and $f_n(\cdot) \stackrel{\text{def}}{=} v(\cdot, z_n r(x_n,a_n) + y_n, z_n\beta)$, where $n = 1, 2, \ldots$. Since the functions v, r are u.s.c and $v(x,\cdot,\cdot)$ is monotonic (nondecreasing actually) the sequence of functions $\{f_n\}_{n=1,2,\ldots}$ is upper semiequicontinuous by (Feinberg et al., 2020, Theorem 3.1(ii)). Again, by upper semicontinuity,

$$f(\cdot) \equiv v(\cdot, zr(x,a) + y, z\beta) \geq \limsup_{n\to\infty} v(\cdot, z_n r(x_n,a_n) + y_n, z_n\beta) = \limsup_{n\to\infty} f_n(\cdot)$$

implying that $\{f_n\}_{n=1,2,\ldots}$ upper semiconverges to f by (Feinberg et al., 2020, Remark 3.4). Also, it is easy to see that, since $(d + b\mathbf{1}_{\mathcal{C}}(\cdot))w(\cdot)$ is the common majorant of $\{f_n\}_{n=1,2,\ldots}$ by Definition (3.4) above, ((Feinberg et al., 2020), Condition (ii) of Theorem 2.2) holds. Hence, by ((Feinberg et al., 2020), Theorem 4.1),

$$\int_{\mathcal{X}} f(x')q(dx'|x,a) \geq \limsup_{n\to\infty} \int_{\mathcal{X}} f_n(x')q(dx'|x_n,a_n)$$

which is equivalent to

$$\eta_r[v](x,y,z,a) \geq \limsup_{n\to\infty} \eta_r[v](x_n,y_n,z_n,a_n).$$

Thus, the function $\eta_r[v]$ is u.s.c and, by ((Aliprantis and Border, 2006), Theorem 2.43), the supremum in (3.6) can be replaced by a maximum and this set of maximizers is nonempty compact since $\mathcal{A}(x)$ is compact given $x \in \mathcal{X}$ by Assumption A(3) above. Thus, by (3.6) and Berge's theorem (Aliprantis and Border, 2006; Lemma 17.30), it follows that $\overline{\boldsymbol{T}}[v]$ is u.s.c. Moreover, by Monotone Convergence Theorem, $(y,z) \to \eta_r[v](x,y,z,a)$ is non-decreasing and continuous (in particular l.s.c). Since the supremum of an arbitrary number of l.s.c functions is also l.s.c, it follows that $(y,z) \to \overline{\mathbf{T}}[v](x,y,z)$ is continuous and non-decreasing. The inequality $\overline{\mathbf{T}}[v](x,y,z) \geq U(y)$ follows directly. By (3.4) since $v(\cdot) \leq (d + b\mathbf{1}_{\mathcal{C}}(\cdot))w(\cdot)$, we have $\overline{\mathbf{T}}[v](x,y,z) \stackrel{by\ (3.7)}{=}$

$$\max_{a\in\mathcal{A}(x)} \int_{\mathcal{X}} v(x', zr(x,a) + y, z\beta)q\,(dx'|x,a)$$

$$\stackrel{a^*\in\mathcal{A}^*(x,y,z)}{=} \int_{\mathcal{X}} v(x', zr(x,a^*) + y, z\beta)q\,(dx'|x,a^*) \leq \int_{\mathcal{X}} (d + b\mathbf{1}_{\mathcal{C}}(x'))w(x')q\,(dx'|x,a^*)$$

$$\stackrel{\text{by Assumption A(1)}}{\leq} d\int_{\mathcal{X}} w(x')q\,(dx'|x,a^*) + b\int_{\mathcal{X}} \mathbf{1}_{\mathcal{C}}(x')w(x')q\,(dx'|x,a^*)$$

$$\stackrel{\text{by Assumption A(1)}}{\leq} d\,(\rho w(x) + b\mathbf{1}_{\mathcal{C}}(x)) + bc'\int_{\mathcal{C}} q\,(dx'|x,a^*) \leq d\,(\rho w(x) + b\mathbf{1}_{\mathcal{C}}(x)) + bc'$$

$$\leq (d + b\mathbf{1}_{\mathcal{C}}(x))\,w(x) + db\mathbf{1}_{\mathcal{C}}(x) + bc' - (1-\rho)dw(x) - b\mathbf{1}_{\mathcal{C}}(x)w(x)$$

$$= (d + b\mathbf{1}_{\mathcal{C}}(x))\,w(x) + \begin{cases} db + bc' - (1-\rho)dw(x) - bw(x) & x \in \mathcal{C}, \\ bc' - (1-\rho)dw(x) & x \notin \mathcal{C}, \end{cases}$$

$$\stackrel{\text{by Assumption A(1)}}{\leq} (d + b\mathbf{1}_{\mathcal{C}}(x))\,w(x) + \begin{cases} b(c'+d) - ((1-\rho)d + b)\left(\frac{c'+d}{(1-\rho)\frac{d}{b}+1}\right) & x \in \mathcal{C}, \\ (1-\rho)(c'^2 - d\underline{\ell}) & x \notin \mathcal{C}, \end{cases}$$

$$\stackrel{\text{by Assumption A(1)}}{\leq} (d + b\mathbf{1}_{\mathcal{C}}(x))\,w(x). \qquad (3.11)$$

Thus $\overline{\mathbf{T}}[v] \in U$ The existence of a maximizing selector follows from the Generalized Selection Theorem of Dubins and Savage (Hinderer, 1970, Theorem 17.9).

We now prove the main Theorem of this section which nontrivially extends Theorem 2 of (Bäuerle and Rieder, 2014) to unbounded per-stage cost or reward satisfying the w-growth condition (see Assumption A) above.

Theorem 3.2. Under Assumption (A), the following results hold:

1. For a policy $\phi \equiv \{\phi_0, \phi_1, ...\} \in \Pi^M$ it holds that $v_{t\pi} = \mathbf{T}^f_{\phi_0} \circ \cdots \circ \mathbf{T}^f_{\phi_{t-1}}[U]$, t = 1, ... T where $v_{t\pi} = \underline{v}_{t\pi}$ when $f = c$ and $v_{t\pi} = \underline{v}_{t\pi}$ when $f = r$.
2. For (x, y, z) $\in \hat{\mathcal{X}}$, $\underline{v}_0(x,y,z) = \overline{v}_0(x,y,z) \equiv v_0(x,y,z) \equiv U(y)$ we have $\underline{v}_t = \underline{\mathbf{T}}[\underline{v}_{t-1}]$ i.e.,

$$\underline{v}_t(x,y,z) = \inf_{a\in\mathcal{A}(x)} \int_{\mathcal{X}} \underline{v}_{t-1}(x', zc(x,a)+y, z\beta) q\,(dx'|x,a)\,, \;\; t = 1,\ldots,T, \quad (3.12)$$

with $\underline{v}_t \in \mathcal{L}(\hat{\mathcal{X}})$ for all t. In particular, $\underline{v}_t = \underline{\mathbf{T}}^t[U]$ for all t. Similarly, we have $\overline{v}_t = \overline{\mathbf{T}}[\overline{v}_{t-1}]$ i.e.,

$$\overline{v}_t(x,y,z) = \sup_{a\in\mathcal{A}(x)} \int_{\mathcal{X}} \overline{v}_{t-1}(x', zr(x,a)+y, z\beta) q\,(dx'|x,a)\,, \;\; t = 1,\ldots,T, \quad (3.13)$$

with $\overline{v}_t \in \mathcal{U}(\hat{\mathcal{X}})$ for all t. In particular, $\overline{v}_t = \overline{\mathbf{T}}^t[U]$ for all t.

3. For each t = 1, 2, ..., T there exists a minimizer $\phi_t^* \in \Phi$ *of* $\underline{v}_{t-1}$ (resp. maximizer $\phi_t^* \in \Phi$ *of* $\overline{v}_{t-1}$) and $\pi^* \equiv (\pi_0^*, \ldots, \pi_{T-1}^*) \in \Pi^M$ with,

$$\pi_0^*(x_0) \equiv \phi_T^*(x_0, 0, 1), \qquad \pi_t^*(h_t) \equiv \phi_{T-t}^*\left(x_t, \sum_{s=0}^{t-1} \beta^s f(x_s, a_s), \beta^t\right) (3.14)$$

is an optimal policy for (3.1) with $f = c$ (resp. $f = r$) in (3.14) above.

4. $\underline{V}_{T,\beta}(x)$(resp. $\overline{V}_{T,\beta}(x)$) is the optimal l.s.c cost (resp. u.s.c utility) for (2.6) under the optimal policy (3.14) above for any $x \in \mathcal{X}$.

Proof: The proofs are provided below.

1. The proof of this part is similar to the proof of Theorem 2(a) of (Bäuerle and Rieder, 2014) and hence is omitted.
2. We prove parts (2) and (3) together. Due to part (1) it follows that for $\pi \in \Pi^M$, the value function in problem (3.1) is the same as the value functions of the original MDP. Moreover, by Theorems 18.1 and 18.4 of (Hinderer, 1970), it suffices to consider Markov policies Π^M i.e. $\underline{v}_t = \inf_{\pi\in\Pi} \underline{v}_{t\pi} = \inf_{\pi\in\Pi^M} \underline{v}_{t\pi}$ or $\overline{v}_t = \sup_{\pi\in\Pi} \overline{v}_{t\pi} = \sup_{\pi\in\Pi^M} \overline{v}_{t\pi}$. Now, since by Theorem 3.1(2) (resp. Theorem 3.1(1)) above $\underline{\mathbf{T}}\,[v] \in \mathcal{L}\,(\hat{\mathcal{X}})$ (resp. $\overline{\boldsymbol{T}}\,[v] \in$ U$(\hat{\mathcal{X}})$ whenever v is and there exists a minimizing (resp. maximizing) selector for v by Theorem 3.1(2) (resp. Theorem 3.1(1)), parts (2) and (3) follow from Theorem 14.4 of (Hinderer, 1970).
3. This part easily follows from the above parts and the fact that $\underline{V}_{T,\beta}(x) = U^{-1} \circ \underline{v}_{T,\beta}(x) = U^{-1} \circ \underline{v}_T(x,0,1)$ (resp. $\overline{V}_{T,\beta}(x) = U^{-1} \circ \overline{v}_{T,\beta}(x) = U^{-1} \circ \overline{v}_T(x,0,1)$) since U^{-1} is continuous and increasing.

Now the following corollary to Theorem 3.2 follows obviously by defining $\hat{\mathcal{X}} \equiv \mathcal{X} \times \mathbb{R}_+$, selectors $\Phi \ni \phi : (x, y) \in \hat{\mathcal{X}} \to \phi(x, y) \in \mathcal{A}(x)$, corresponding Definition (3.2) putting $z = 1$, Definitions (3.4), (3.5) and (3.6) with the corresponding transition kernel $\tilde{q}\,(d(x', y')|x, y, a) \equiv q\,(dx'|x, a) \otimes \mathbf{1}_{(f(x,a)+y)}(dy')$ defined as,

$$\int_{\hat{\mathcal{X}}} v(x', y')\tilde{q}\,(d(x', y')|x, y, a) = \int_{\hat{\mathcal{X}}} v(x', y')q\,(dx'|x, a) \otimes \mathbf{1}_{(f(x,a)+y)}(dy')$$
$$= \int_{\mathcal{X}} v(x', f(x, a) + y)q\,(dx'|x, a)\,, \quad f = c, r \quad (3.15)$$

and the corresponding (simpler) interpretation (putting $z = 1$) of Theorem 3.1. The Corollary 3.3 is a nontrivial extension of Theorem 1 of (Bäuerle and Rieder, 2014).

Corollary 3.3. For the undiscounted case, i.e. $\beta = 1$, exactly same results as in Theorem 3.2 hold for $\underline{V}_T \overset{def}{=} \inf_{\pi\in\Pi} V_T^\pi$ and $\overline{V}_T \overset{def}{=} \sup_{\pi\in\Pi} V_T^\pi$.

We end this section with an important Remark.

Remark 4. It is to be noted here that, under these assumptions of upper semicontinuity of the map $\mathcal{X} \ni x \to \mathcal{A}(x)$in addition to the compactness of $\mathcal{A}(x)$ for each $x \in \mathcal{X}$, (Bäuerle and Rieder, 2014) has proved the analogue of the above Theorem 3.2 (see Theorem 2 and Remark 2 therein) and Corollary 3.3 (see Theorem 1 therein) for bounded costs/rewards. We refer also to our Remark 2 above. Our Theorem 3.2 and Corollary 3.3 above automatically generalizes these corresponding results of (Bäuerle and Rieder, 2014) to any unbounded above cost or rewards satisfying standard *w*-growth conditions of Assumption (A) above.

4. AN APPLICATION

In this section, we provide a real-life example from Portfolio Consumption using the robust risk-sensitive preferences of (Basu et al., 2022) originally from (Howard and Matheson, 1972). For a given fixed but arbitrary risk sensitivity parameter $\theta > 0$, let $U(y) \equiv -\frac{1}{\theta}e^{-\theta y}$ be the concave reward/utility functional. We consider a portfolio of one asset with values $x \in \mathcal{X} \equiv (0, \infty)$. The state space is defined as $\hat{\mathcal{X}} \equiv (0, \theta] \times \mathcal{X}$. The action of a consumer (agent) is to consume an amount $a \in \mathcal{A}(x) \equiv [0, \delta x] \subset \mathcal{A} \equiv \mathcal{X}$ when the value of the asset is x given a fixed but arbitrary $0 \le \delta \le 1$. Thus Assumption A (3) is satisfied. For some given fixed but arbitrary $\gamma > 0$, let $\epsilon \overset{def}{=} \left(1 + \frac{2\sqrt{\pi\gamma}}{2-\sqrt{\pi}}\right)^2$ and $C \overset{def}{=} (0, \theta] \times (0, \epsilon] \subset \hat{\mathcal{X}}$. We choose a fixed but arbitrary $\underline{\ell} \in (0, \frac{1}{2}]$ in Assumption A(1). We define the corresponding $w(\cdot, \cdot)$ as,

$$w(\eta, x) \equiv \sqrt{x} + 1 - \underline{\ell}, \ (\eta, x) \in \hat{\mathcal{X}}. \quad (4.1)$$

We choose, as per Assumption A(4), some u.s.c reward $0 < r\,(\eta, x, a) \equiv r\,(x, a) \le dw\,(\eta, x)$ at state (η, x) for action a where $d \ge \frac{c'^2}{\underline{\ell}}$. Given $(\eta, x, a) \in Gr(\mathcal{A}) \equiv \{(\eta, x, a) : \eta \in (0, \theta], x \in \mathcal{X}; a \in [0, \delta x]\}$ we define the transition kernel $\hat{q}\,(d(\eta', x')|\eta, x, a) \equiv q\,(dx'|x, a) \otimes \mathbf{1}_{\eta\beta}(d\eta')$ as follows,

$$\int_{\hat{\mathcal{X}}} v(\eta', x')\hat{q}\,(d(\eta', x')|\eta, x, a) = \int_{\hat{\mathcal{X}}} v(\eta', x')q\,(dx'|x, a) \otimes \mathbf{1}_{\eta\beta}(d\eta')$$

$$= \int_{\mathcal{X}} v(\eta\beta, x') q\,(dx'|x,a) \tag{4.2}$$

where

$$q\,(B|x,a) \equiv \int_B \left[\frac{2a}{x^2} e^{-\frac{2x'}{x}} + \left(1-\frac{a}{x}\right)\frac{1}{x+\gamma} e^{-\frac{x'}{x+\gamma}}\right] dx',\ \forall B \in \mathcal{B}(\mathcal{X}) \tag{4.3}$$

which models the (controlled) evolution of the asset value under consumptions. Let, $\rho \equiv \frac{\sqrt{\pi}}{4} + \frac{1}{2}$ and $b \equiv \frac{\sqrt{\pi\gamma}}{2} + \left(1 - \frac{\sqrt{\pi}}{2}\right)(1 - \underline{\ell})$.

Now we state and prove the following important Lemma.

Lemma 4.1. The function $w(\cdot,\ \cdot)$ as defined in (4.1) satisfies the conditions of Assumption A (1) above.

Proof: $\omega(\cdot,\cdot) \geq \underline{\ell}$ and $\underline{\ell} < c' = 2\left(1 + \frac{\sqrt{\pi\gamma}}{2-\sqrt{\pi}}\right) - \underline{\ell}$ as required by Assumption A(1). Also, we see that,

$$c'(1-\rho) = \left(1-\frac{\sqrt{\pi}}{2}\right)\left(1+\frac{\sqrt{\pi\gamma}}{2-\sqrt{\pi}}\right) - \underline{\ell}\left(\frac{1}{2}-\frac{\sqrt{\pi}}{4}\right) = \frac{2-\sqrt{\pi}+\sqrt{\pi\gamma}}{2} - \underline{\ell}\left(\frac{1}{2}-\frac{\sqrt{\pi}}{4}\right)$$
$$= 1 - \frac{\sqrt{\pi}}{2} + \frac{\sqrt{\pi\gamma}}{2} - \underline{\ell}\left(\frac{1}{2}-\frac{\sqrt{\pi}}{4}\right) > 1 - \frac{\sqrt{\pi}}{2} + \frac{\sqrt{\pi\gamma}}{2} - \underline{\ell}\left(1-\frac{\sqrt{\pi}}{2}\right) = b \tag{4.4}$$

as required by Assumption A (1). Now, using standard integration results, it can be checked that,

$$\begin{aligned}
\int_{\hat{\mathcal{X}}} w(\eta',x')\hat{q}\,(d(\eta',x')|\eta,x,a) &= \int_{\hat{\mathcal{X}}} w(\eta',x') q\,(dx'|x,a) \otimes \mathbf{1}_{\eta\beta}(d\eta') \\
&= \int_{\mathcal{X}} w(\eta\beta,x') q\,(dx'|x,a) \\
&= \frac{a}{x}\frac{\sqrt{\pi}}{2\sqrt{2}}\sqrt{x} + \left(1-\frac{a}{x}\right)\frac{\sqrt{\pi}}{2}\sqrt{x+\gamma} + 1 - \underline{\ell} \\
&\leq \frac{a}{x}\frac{\sqrt{\pi}}{2}\sqrt{x} + \left(1-\frac{a}{x}\right)\frac{\sqrt{\pi}}{2}(\sqrt{x}+\sqrt{\gamma}) + 1 - \underline{\ell} \\
&\leq \frac{\sqrt{\pi}}{2}(\sqrt{x}+\sqrt{\gamma}) + 1 - \underline{\ell} \\
&= \frac{\sqrt{\pi}}{2} w(x) + \frac{\sqrt{\pi\gamma}}{2} + \left(1-\frac{\sqrt{\pi}}{2}\right)(1-\underline{\ell})
\end{aligned} \tag{4.5}$$

for all $(\eta,\ x,\ a) \in Gr(\mathcal{A})$. It immediately follows from (4.5) above that,

$$\int_{\hat{\mathcal{X}}} w(\eta',x')\hat{q}\,(d(\eta',x')|\eta,x,a) \leq \rho w(\eta,x) + b,\ (\eta,x) \in \mathcal{C},\ a \in \mathcal{A}(x). \tag{4.6}$$

Also, from (4.4) and (4.5) it follows that,

$$\int_{\hat{\mathcal{X}}} w(\eta',x')\hat{q}\,(d(\eta',x')|\eta,x,a) \leq \rho w(\eta,x),\ (\eta,x) \in \hat{\mathcal{X}} \setminus \mathcal{C},\ a \in \mathcal{A}(x). \tag{4.7}$$

Combining the above inequalities (4.6) and (4.7) we get,

$$\int_{\hat{\mathcal{X}}} w(\eta', x')\hat{q}\left(d(\eta', x')|\eta, x, a\right) \leq \rho w(\eta, x) + b\mathbf{1}_{\mathcal{C}}(\eta, x), \ (\eta, x, a) \in Gr(\mathcal{A}) \tag{4.8}$$

as required by Assumption A (1).

We now formulate the risk-sensitive or robust preference problem along the lines of (Basu et al., 2022) and (Howard and Matheson, 1972). We denote by $\mathcal{F}_+$ the set of extended non negative measurable functions f : $\hat{\mathcal{X}} \to \mathbb{R} + \cup \{+\infty\}$ Given $\pi \in \Pi$ the corresponding finite-horizon (i.e. from 0 to T) risk-sensitive discounted utility $V^{\pi}_{T,\beta}(\theta,\cdot)$starting at $(\theta, x_0) \in \hat{\mathcal{X}}$ is given as:

$$V^{\pi}_{T,\beta}(\theta, x_0) \overset{def}{=} -\frac{1}{\theta}\ln E^{\pi}_{x_0}\left[e^{-\theta(\sum_{t=0}^{T-1}\beta^t r(X_t,a_t)+\beta^T g(X_T))}\right], \text{ and } V^{\pi}_{0,\beta}(\cdot, x_T) = g(x_T) \ \ (4.9)$$

under the dynamics generated by π, where $0 < \beta < 1$ is a given discount factor as in Section 2 and $g(\cdot) \in \mathcal{F}_+$ is a given terminal cost function. The corresponding optimization problem is given as:

$$\overline{V}_{T,\beta}(\theta, x_0) \equiv \sup_{\pi\in\Pi} V^{\pi}_{T,\beta}(\theta, x_0). \tag{4.10}$$

As we shall show below in Proposition 4.3, by Theorems 18.1 and 18.4 of (Hinderer, 1970), it suffices to consider Markov policies $\pi \in \Pi^M$ i.e.

$$\overline{V}_{T,\beta}(\theta, x_0) \equiv \sup_{\pi\in\Pi} V^{\pi}_{T,\beta}(\theta, x_0) = \sup_{\pi\in\Pi^M} V^{\pi}_{T,\beta}(\theta, x_0). \tag{4.11}$$

We now provide below a heuristic recursive formulation for the preference function (4.9) which shall allow us to better understand the Dynamic Programming formulation for this maximization problem (4.11) above,

Lemma 4.2. For $1 \leq t \leq T$, $\eta \in (0, \theta]$ and any $\pi \in \Pi$,

$$V^{\pi}_{t,\beta}(\eta, X_{T-t}) = r(X_{T-t}, a_{T-t}) - \frac{1}{\eta}\ln E^{\pi}_{X_{T-t}}\left[e^{-\eta\beta V^{\pi}_{t-1,\beta}(\eta\beta, X_{T-t+1})}\right]$$

with the terminal condition $V^{\pi}_{0,\beta}(\cdot, X_T) = g(X_T)$. (4.12)

Proof: Note that, using (4.9), we can get,

$$\begin{aligned}
V^{\pi}_{t,\beta}(\eta, X_{T-t}) &= -\frac{1}{\eta}\ln E^{\pi}_{X_{T-t}}\left[e^{-\eta(\sum_{s=T-t}^{T-1}\beta^{s-(T-t)}r(X_s,a_s)+\beta^t g(X_T))}\right] \\
&= -\frac{1}{\eta}\ln E^{\pi}_{X_{T-t}}\left[e^{-\eta r(X_{T-t},a_{T-t})-\eta\beta\sum_{s=T-t+1}^{T-1}\beta^{s-(T-t+1)}r(X_s,a_s)-\eta\beta^t g(X_T)}\right] \\
&= -\frac{1}{\eta}\ln E^{\pi}_{X_{T-t}}\left[e^{-\eta r(X_{T-t},a_{T-t})}E^{\pi}_{X_{T-t+1}}\left[e^{-\eta\beta(\sum_{s=T-t+1}^{T-1}\beta^{s-(T-t+1)}r(X_s,a_s)+\beta^{t-1}g(X_T))}\right]\right] \\
&= -\frac{1}{\eta}\ln E^{\pi}_{X_{T-t}}\left[e^{-\eta r(X_{T-t},a_{T-t})}e^{-\eta\beta V^{\pi}_{t-1,\beta}(\eta\beta, X_{T-t+1})}\right] \\
&= r(X_{T-t}, a_{T-t}) - \frac{1}{\eta}\ln E^{\pi}_{X_{T-t}}\left[e^{-\eta\beta V^{\pi}_{t-1,\beta}(\eta\beta, X_{T-t+1})}\right].
\end{aligned} \tag{4.13}$$

The terminal condition is obvious.

$\mathcal{L}_R$As shown by Lemma 4.2 above, there is an explicit time-dependence of the risk aversion parameter through exponentiation of the discounting factor β. Hence we have defined our Markov control process $\{(\eta_t; X_t)\}_{0\leq t\leq T}$ on the enhanced state space $\hat{\mathcal{X}}$, with the selectors now being measurable maps $\emptyset : \hat{\mathcal{X}} \to \mathcal{A}$ such that $\hat{\mathcal{X}} \ni (\eta, x) \to \phi(\eta, x) \in \mathcal{A}(x)$ (see notations above). We again denote by Φ the set of such selectors. The definitions for histories and all types of policies can be defined analogously as above and we keep the notations similar to Section 2 above for ease of understanding. The corresponding transition kernel $\hat{q}(\cdot|\cdot)$ has been defined in (4.2) above. Again, the Ionescu-Tulcea theorem (see, e.g., (Bertsekas and Shreve, 1996, Proposition 7.28) implies that, for each initial state $(\eta, x) \in \hat{\mathcal{X}}$ and policy $\pi \in \Pi$, there exists an unique probability measure $P^{\pi}_{\eta,x}$ and a stochastic process $\{((\eta_t, X_t); A_t)\}_{0\leq t\leq T}$ on the space $(H_T, \mathcal{B}(H_T))$ of all histories (trajectories) such that $(\eta_t, X_t)(h_T) = (\eta_t, x_t)$ is the (random) state and $A_t(h_T) = a_t$ is the (random) action at time t for $h_T \in H_T$. We denote by $E^{\pi}_{\eta,x}$ the corresponding expectation. Now, we can simply consider the preference function (4.9) to be defined on the above state space $\hat{\mathcal{X}}$ with the corresponding optimization problem (4.10) or (4.11) as being defined on the associated state process $\{(\eta_t, X_t)\}_{0\leq t\leq T}$ as explained above. Having understood this, we now proceed to define the corresponding Bellman operators. Motivated by Lemma 4.2, the kernel (4.2) and the discussions around it, we define $\ni (\eta, x, a, f) \to \mathcal{L}_R$: $(\eta, x, a$ $\mathcal{L}_R : \hat{\mathcal{X}} \times \mathcal{A} \times \mathcal{F}_+$ ws:

$$
\begin{aligned}
\mathcal{L}_R(\eta, x, a, f) &\equiv r(x,a) - \frac{1}{\eta} \ln \int_{\hat{\mathcal{X}}} e^{-\eta' f(\eta', x')} \hat{q}\left(d(\eta', x')|\eta, x, a\right) \\
&= r(x,a) - \frac{1}{\eta} \ln \int_{\mathcal{X}} e^{-\eta\beta f(\eta\beta, x')} q\left(dx'|x, a\right). \qquad (4.14)
\end{aligned}
$$

For $\phi \in \Phi$ and $(\eta, x, f) \in \hat{\mathcal{X}} \times \mathcal{F}_+$, let,

$$
\begin{aligned}
\mathbf{T}^{(R)}_{\phi} f(\eta, x) &\overset{def}{=} \mathcal{L}_R(\eta, x, \phi(\eta, x), f), \\
\mathbf{T}^{(R)} f(\eta, x) &\overset{def}{=} \sup_{a\in\mathcal{A}(x)} \mathcal{L}_R(\eta, x, a, f). \qquad (4.15)
\end{aligned}
$$

We now state the following main result (Bellman Dynamic Programming Recursion) of this Section for the cost function (4.9) as follows directly from Theorem 3.2 above. Hence we have proved the following Proposition.

Proposition 4.3. For the Portfolio Consumption Model above, it holds that

1. For any given policy $(\phi_0, \dots \phi_{T-1}) \equiv \pi \in \Pi^M$ with $\phi_t \in \Phi$, $\overline{V}^{\pi}_{T,\beta} = T^{(R)}_{\phi_0} \circ \dots \circ T^{(R)}_{\phi_{T-1}} g.$
2. $\overline{V}_{0,\beta} \equiv g$ *and* $\overline{V}_{t,\beta} = \mathbf{T}^{(R)} \overline{V}_{t-1,\beta}$ i.e. for t = 1, ..., T

$$
\begin{aligned}
\overline{V}_{t,\beta}(\eta, x) &= \sup_{a\in\mathcal{A}(x)} \mathcal{L}_R(\eta, x, a, \overline{V}_{t-1,\beta}) \\
&= \sup_{a\in\mathcal{A}(x)} \left[r(x,a) - \frac{1}{\eta} \ln \int_{\mathcal{X}} e^{-\eta\beta \overline{V}_{t-1,\beta}(\eta\beta, x')} q\left(dx'|x, a\right) \right], \quad (\eta, x) \in \hat{\mathcal{X}}. \quad (4.16)
\end{aligned}
$$

So, $\overline{V}_{t,\beta} = \left(\mathbf{T}^{(R)}\right)^t g$ for all t and in particular $\overline{V}_{T,\beta} = \left(\mathbf{T}^{(R)}\right)^T g.$

3. There exists a maximizing sequence $\{\phi^*_t \in \Phi\}_{0\leq t\leq T-1}$ and $\pi^* \equiv (\pi^*_0, \dots, \pi^*_{T-1}) \in \Pi^M$ such that, for all $t = 0, \dots, T-1$.

$$
\pi^*_t(h_t) \equiv \phi^*_t(\eta_t, x_t) = \arg\max_{a\in\mathcal{A}(x_t)} \mathcal{L}_R(\eta_t, x_t, a, V_{T-1-t,\beta})
$$

$$= \underset{a \in \mathcal{A}(x_t)}{\arg\max} \left[r(x,a) - \frac{1}{\eta_t} \ln \int_{\mathcal{X}} e^{-\eta_t \beta \overline{V}_{T-1-t,\beta}(\eta_t \beta, x')} q\left(dx'|x_t, a\right) \right] \quad (4.17)$$

is an optimal policy for (4.9). In fact $\phi_t^*(\eta_t, x_t) \equiv \phi^*(\eta_0 \beta^t, x_t)$ for some fixed $\phi^* \in \Phi$ where, as in (4.9), $\eta_0 = \theta$ is the given risk-sensitivity parameter.

5. CONCLUSION AND FUTURE DIRECTIONS

Our main technical contribution is to solve a certainty equivalent optimization problem in the finite horizon case for per-stage unbounded costs and rewards satisfying a *w*-growth condition (see Theorem 3.2 and Corollary 3.3 in Section 3). In the process, a result of much broader potential applicability have been proved, namely, certain important closure properties of the dynamic programming (Bellman) operators, the compactness of optimal action (sub)sets and the corresponding existence of optimizing selectors in Theorem 3.1 in Section 3. We also provide a real-life example from Portfolio Consumption Model to illustrate our ideas (see Proposition 4.3 in Section 4) because such models have regularly been used as canonical examples in certainty equivalent formulations. Several more nontrivial applications and examples can be created as separate papers from these theoretical foundations along the lines of the references mentioned in the Introduction. This work can further be extended to noncompact action spaces as well as to generic unbounded-above per-stage costs/rewards without assuming any growth conditions. A further direction of technical interest would be to address these issues for the infinite horizon cost or reward setup as well as the corresponding average cost or reward setups given appropriate geometric ergodicity conditions on the underlying controlled state process. Another direction would be to study the preferences of general Epstein-Zin-Weil structure-types in this framework as a decision problem of significant interest. An interesting research direction would be to study computational algorithms to evaluate such value functions as studied herein for large-scale optimization problems.

6. ACKNOWLEDGMENT

The author most sincerely thanks Prof. Dr. Sukru Talas for his kind support in typesetting and formatting this manuscript as per the requirements of this journal.

7. CONFLICT OF INTEREST

Authors approve that to the best of their knowledge, there is not any conflict of interest or common interest with an institution/organization or a person that may affect the review process of the paper.

8. AUTHOR CONTRIBUTION

Arnab BASU has the full responsibility of the paper about determining the concept of the research, data collection, data analysis and interpretation of the results, preparation of the manuscript and critical analysis of the intellectual content with the final approval.

9. REFERENCES

Aliprantis C. D., Border K. C., Innite Dimensional Analysis: A Hitchhiker's Guide, Third Ed., Springer-Verlag, Berlin, 2006.

Altman E., Hordijk A., Spieksma F, M., Contraction Conditions for Average and α-discount Optimality in Countable State Markov Games with Unbounded Rewards, Math. Oper. Res. 22(3), 588-618, 1997.

Anderson E. W., Hansen L. P., Sargent T. J., Small Noise Methods for Risk Sensitive/Robust Economies, J. Econ. Dyn. Control 36, 468-500, 2012.

Basu A., Lontzek T., Schmedders, K., Zhao Y., The Social Cost of Carbon when we wish for Robustness, Accepted for publication in Management Science, 2022.

Bäuerle N., Rieder U., More Risk-Sensitive Markov Decision Processes, Math. Oper. Res. 39(1), 105-120, 2014.

Bäuerle N., Jasckiewicz A., Stochastic Optimal Growth Model with Risk-Sensitive Preferences. J. Eco, Theory 173, 181-200, 2018.

Bellman R., Dynamic Programming, Princeton University Press, Princeton, New Jersey, U.S.A, 1957.

Bertsekas D. P., Shreve, S. E., Stochastic Optimal Control: The Discrete-Time Case, Athena Scientic, Belmont, Massachusetts, 1996.

Bielecki T., Pliska S. R., Economic Properties of the Risk-Sensitive Criterion for Portfolio Management, Rev. Account. Fin. 2(2), 3-17, 2003.

Bommier A., Life-Cycle Preferences Revisited, J. Eur. Econ. Assoc. 11(6), 1290-1319, 2013.

Bommier, Lanz A. B., Zuber S., Models-as-usual for Unusual Risks? On the Value of Catastrophic Climate Change, J. Environ. Econ. Manag. 74, 1-22, 2015.

Bommier A., Kochov A., Le Grand F., On Monotone Recursive Preferences, Econometrica 85(5), 1433-1466, 2017.

Bommier, A. and F. Le Grand., Risk Aversion and Precautionary Savings in Dynamic Settings, Management Science 65(3), 1386-1397, 2019.

Borkar V. S., Probability Theory: An Advance Course, Springer-Verlag, New York,1995.

Cai Y., Lontzek T. S., The Social Cost of Carbon with Economic and Climate Risks, J. Pol. Econ. 127(6): 2684-2734, 2019.

Di Masi G. V., Stettner L., Risk-sensitive Control of Discrete-time Markov Processes with Infinite Horizon, SIAM J. Control Optim., 38(1): 61-78, 1999.

Epstein L. G., Zin S. E., Substitution, Risk Aversion, and the Temporal Behaviour of Consumption and Asset Returns: A Theoretical Framework, Econometrica 57(4): 937-969, 1989.

Epstein L. G., Zin S. E., Substitution, Risk Aversion, and the Temporal Behaviour of Consumption and Asset Returns: An Empirical Analysis, J. Pol. Econ. 99(2):263-286, 1991

Feinberg E. A., Kasyanov P. O., Zadoianchuk N. V., Average Cost Markov Decision Processes with Weakly Continuous Transition Probabilities, Math. Oper. Res. 37(4), 591-607. 2012

Feinberg E. A., Kasyanov P. O., Liang Y., Fatou's Lemma in Its Classical Form and Lebesque's Convergence Theorems for Varying Measures with Applications to Markov Decision Processes, Theory Probab. Appl. 65(2), 270-291, 2020.

Hansen L. P., Sargent T. J., Discounted Linear Exponential Quadratic Gaussian Control, IEEE Trans. on Auto. Control 40(5): 968-971, 1995.

Hansen L. P., Sargent T. J., Turmuhambetova G., Williams N., Robust Controland Model Misspecication, J. Econ. Theory 128, 45-90, 2006.

Hansen L. P., Sargent T. J., Recursive Robust Estimation and Control without Commitment, J. Econ. Theory 136(1): 1-27, 2007.

Hansen L. P., Sargent T. J., Robustness. Princeton Univ. Press, 2008.

Hinderer K., Foundations of Nonstationary Dynamic Programming with Discrete Time Parameter, Springer-Verlag, Berlin, 1970.

Howard R. A., Matheson J. E., Risk-sensitive Markov Decision Processes, Management Sc. 18(7), 356-369, 1972.

Jacobson D. H., Optimal Stochastic Linear Systems with Exponential Performance Criteria and Their Relation to Deterministic Dierential Games, IEEE Trans. Aut. Control, AC-18(2): 124-131, 1973.

Kreps D. M., Decision Problems with Expected Utility Criteria, I: Upper and Lower Convergent Utility, Math. Oper. Res. 2(1), 45-53, 1977.

Kreps D. M., Decision Problems with Expected Utility Criteria, II: Stationarity Math. Oper. Res. 2(3), 266-274, 1977.

Kreps D. M., Porteus E. L., Temporal Resolution of Uncertainty and Dynamic Choice Theory, Econometrica, 46(1): 185-200, 1978.

Kreps D. M., Porteus E. L., Dynamic Choice Theory and Dynamic Programming, Econometrica, 47(1): 91-100, 1979.

Meyn S. P., Tweedie R. L., Markov Chains and Stochastic Stability, Second Ed., Cambridge University Press, 2009.

Muliere P., Parmigiani G., Utility and Means in the 1930s, Statistic. Sc. 8(4), 421-432, 1993.

Nowak A. S., Notes on risk-sensitive Nash equilibria, in Advances in Dynamics, 95-109, Ann. Internatl. Soc. Dyn. Games, 7, Birkhauser, Boston, 2005.

Nowak A. S., Altman E., ε-Equilibria for Stochastic Games with Uncountable State Space and Unbounded Costs, SIAM J. Control Optim. 40(6), 1821-1839, 2002.

Tallarini Jr. T. D., Risk-Sensitive Business Cycles, J. Mon. Econ. 45, 507-532, 2000.

Whittle P., Risk-Sensitive Optimal Control, Wiley, Chichester, England, and New York, 1990.