Persistently optimal policies in stochastic dynamic programming with generalized discounting

Jaśkiewicz, Anna; Matkowski, Janusz and Nowak, Andrzej S.
University of Zielona Góra

21. June 2011

Persistently optimal policies in stochastic dynamic programming with generalized discounting

Anna Jaśkiewicz,[1] Janusz Matkowski,[2] Andrzej S. Nowak[2]

Abstract In this paper we study a Markov decision process with a non-linear discount function. Our approach is in spirit of the von Neumann-Morgenstern concept and is based on the notion of expectation. First, we define a utility on the space of trajectories of the process in the finite and infinite time horizon and then take their expected values. It turns out that the associated optimization problem leads to a non-stationary dynamic programming and an infinite system of Bellman equations, which result in obtaining persistently optimal policies. Our theory is enriched by examples.

Key words Stochastic dynamic programming, Persistently optimal policies, Variable discounting, Bellman equation, Resource extraction, Growth theory

JEL Classification Numbers C61 and D90

1. Introduction

This paper is devoted to recursive utilities with a new measure of impatience given exogenously by a real-valued function. Our theory is developed for a vast class of decision processes under uncertainty and includes various stochastic growth models as specific examples. To explain the basic idea behind, we first refer to a simple case. Let us consider a standard deterministic growth model with some fixed production technology and a bounded instantaneous utility (in other words, subutility) function u. If $c := \{c_t\}$ is a feasible consumption sequence, then the standard *time additive and separable utility* is given by

$$U(c) := \sum_{t=1}^{\infty} \beta^{t-1} u(c_t) \tag{1}$$

where $\beta \in (0, 1)$ is a fixed discount factor, see Samuelson (1937). Koopmans (1960) proposed a more general approach to construction of recursive utilities via the so-called *aggregator*. Such an aggregator, roughly speaking, is a function $G(a, r)$ of two real variables. Then, a recursive utility U^* is a unique solution to the following equation

$$U^*(c_t, c_{t+1}, ...) = G(u(c_t), U^*(c_{t+1}, c_{t+2}, ...)) \tag{2}$$

for any t and $c = \{c_\tau\}$. This equation indicates that utility enjoyed from period t on depends on current consumption c_t and the aggregate utility $U^*(c_{t+1}, c_{t+2}, ...)$ from period $t+1$ on. Clearly, U in (1) can be obtained by applying the aggregator $G(x, y) = a+\beta r$. In the literature, equation

[1] *Institute of Mathematics and Computer Science, Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland, e-mail: ajaskiew@im.pwr.wroc.pl*

[2] *Faculty of Mathematics, Computer Science and Econometrics, University of Zielona Góra, Podgórna 50, 65-246 Zielona Góra, Poland, e-mails: j.matkowski@wmie.uz.zgora.pl, a.nowak@wmie.uz.zgora.pl (corresponding author)*

(2) is referred to as Koopmans' equation. One of the commonly used conditions imposed on G is the following

$$|G(a, r_1) - G(a, r_1)| \leq \beta |r_1 - r_2| \tag{3}$$

for each a, r_1, r_2 and some $\beta \in (0, 1)$ (see Denardo (1967); Lucas and Stokey (1984); Stokey et al. (1989).[3]) It ensures (by the Banach contraction principle) the existence of a unique bounded solution to equation (2). Recursive utilities derived in this way need not possess time additivity and separability properties, see Becker and Boyd III (1997); Boyd III (1990)

The key idea of this paper is to introduce a *variable discount factor* into a stochastic framework. Namely, we propose to replace the constant β in (3) by an increasing real-valued function δ satisfying some natural conditions such as $\delta(0) = 0$, $\delta(r) < r$ for all $r > 0$. In the deterministic case such models were studied in Jaśkiewicz et al. (2011). An extension of the Banach contraction principle given by Matkowski (1975) enables us to construct a much larger class of recursive utilities. In the simple case with aggregator $G(a, r) = a + \delta(r)$, we obtain a new utility

$$U_\delta^*(c) = u(c_1) + \delta(u(c_2) + \delta(u(c_3) + \cdots)), \quad c = \{c_t\}. \tag{4}$$

Clearly, (4) is not separable. It reduces to (1), when $\delta(r) = \beta r$ for all r. But, if δ is merely piecewise linear, that is, for some constants β_1, $\beta_2 \in (0, 1)$, $\delta(r) = \beta_1 r$ for $r \geq 0$, and $\delta(r) = \beta_2 r$ for $r < 0$, then we deal with a non-separable case. A detailed discussion of different discount functions δ and their applications in the deterministic case is included in Jaśkiewicz et al. (2011). In this paper, we consider the recursive utilities on the space of sample paths of the stochastic decision process and then use their expected values (the von Neumann-Morgenstern approach). We would like to emphasize that our results can be applied to a number of stochastic growth models, and therefore embraces also those studied by Brock and Mirman (1972), Stokey et al. (1989). Unfortunately, within such a framework we cannot expect to obtain a stationary or Markov optimal policy. However, we prove the existence of *persistently optimal policies* and give their characterization by a system of Bellamn's optimality equations. Related results were obtained by Dubins and Savage (1976) (in gambling theory), Kertz and Nachman (1979); Schäl (1975, 1981) (in stochastic dynamic programming), and by Nowak (1986) (in dynamic games). For linear functions δ, our results reduce to those of Blackwell (1965), Bertsekas (1977); Bertsekas and Shreve (1978) (in discounted dynamic programming) and Brock and Mirman (1972) (in the theory of optimal economic growth). An excellent survey of different criteria in dynamic programming can be found in Feinberg (2002) and Feinberg and Shwartz (2002). In the first step, we study models with a bounded subutility function u. Using a natural truncation technique, we also obtain some optimality results for u bounded from above, generalizing the papers by Strauch (1966) and Schäl (1975). Basic convergence results for our derivation of the Bellman equations are given in Section 6.

2. Preliminaries

We start with some preliminaries. By R we denote the set of all real numbers and $\underline{R} = R \cup \{-\infty\}$. Let S, A be non-empty Borel (subsets of complete separable metric) spaces. Assume

[3] A similar contraction assumption for stochastic dynamic programming was used by Bertsekas (1977) and Porteus (1982).

that Δ is a Borel subset of $S \times A$ such that

$$A(s) := \{c \in A : \quad (s,c) \in \Delta\}$$

is non-empty and compact. Then, it is well-known that there exists a Borel mapping $\varphi : S \mapsto A$ such that $\varphi(s) \in A(s)$ for each $s \in S$, see Corollary 1 in Brown and Purves (1973). A set-valued mapping $s \to A(s)$ (induced by the set Δ) is called *upper semicontinuous* if $\{s \in S : A(s) \cap K \neq \emptyset\}$ is closed for each closed set $K \subset A$.

Lemma 1 (a) *Let $g : \Delta \mapsto R$ be a Borel measurable function such that $c \mapsto g(s,c)$ is upper semicontinuous on $A(s)$ for each $s \in S$. Then,*

$$g^*(s) := \max_{c \in A(s)} g(s,c)$$

is measurable and there exists a measurable mapping $f^* : S \mapsto A$ such that

$$f^*(s) \in \arg \max_{c \in A(s)} g(s,c)$$

for all $s \in S$. (b) If, in addition, we assume that $s \mapsto A(s)$ is upper semicontinuous and g is upper semicontinuous on Δ, then g^* is also upper semicontinuous.

Part (a) follows from Corollary 1 in Brown and Purves (1973) and (b) is a corollary to Berge's maximum theorem, see pages 115-116 in Berge (1963).

3. The dynamic programming model

In this section, we examine a dynamic programming model with non-separable utility. Our approach is inspired by the works on non-stationary dynamic programming (Hinderer (1970); Kertz and Nachman (1979); Schäl (1975, 1981)), and therefore includes also models studied by Blackwell (1965) and Brock and Mirman (1972).

Let Y be a metric space, and let $B(Y)$ stand for the the space of all bounded from above real-valued Borel measurable functions on Y. Further, let $\underline{R} := R \cup \{-\infty\}$

A *discrete-time decision process* is specified by the following objects:

(i) X is the *state space* and is assumed to be a Borel space.

(ii) A is the *action space* and is assumed to be a Borel space.

(iii) D is non-empty Borel subset of $X \times A$. We assume that for each $x \in X$, the non-empty x-section

$$A(x) := \{c \in A : \quad (x,c) \in D\}$$

of D represents the set of *actions available in state* x. (In the context of growth theory, $c \in A(x)$ will be often referred to as a feasible consumption when the stock is in state $x \in X$.) We presume that $A(x)$ is compact for each $x \in X$.

(iv) q is a *transition probability* from D to X.

Let $\hat{H}_0 = X$ and

$$\hat{H}_n = \underbrace{(X \times A) \times \ldots \times (X \times A)}_{n} \times X, \quad \hat{H} = (X \times A) \times (X \times A) \times \ldots$$

for $n \geq 0$. Then, $\hat{H}_n$ is the set of histories up to the nth period, and $\hat{H}$ is the set of infinite histories. We assume that $\hat{H}_n$ and $\hat{H}$ are equipped with product Borel σ-algebras. Let x_k and c_k describe the state and action at period k. By $h_n = (x_0, c_0, \ldots, x_n)$ and $h = (x_0, c_0, \ldots)$ we denote the elements of $\hat{H}_n$ and $\hat{H}$, respectively. Let H_n and H be the sets of feasible histories h_n and h, respectively, where each $c_k \in A(x_k)$. Clearly, H_n (H) is a Borel subset of $\hat{H}_n$ $(\hat{H})$.

(v) $u \in B(D)$ is a *utility function.*

(vi) $\delta : \underline{R} \mapsto \underline{R}$ is a *discount function.*

A *policy*[4] $\pi = \{p_n\}$ is defined as a sequence of Borel measurable mappings $p_n : H_n \mapsto A$ such that $p_n(h_n) \in A(x_n)$, $n \geq 0$, $h_n \in H_n$. We write Π to denote the set of all policies. Since $A(x)$ is compact, we note that Π is non-empty (see Corollary 1 in Brown and Purves (1973)).

We now make our basic assumption on the discount function δ.

(A1) There exists a continuous increasing function $\gamma : [0, \infty) \to [0, \infty)$ such that $\gamma(z) < z$ for each $z > 0$ and

$$|\delta(z_1) - \delta(z_2)| \leq \gamma(|z_1 - z_2|) \tag{5}$$

for all $z_1, z_2 \in D$.

(A2) δ is continuous and nondecreasing.

Assumption (A1) implies that $\gamma(0) = 0$.

For $n \geq 0$ we define H^n as the space of future feasible histories of the process from period n onwards. An element of H^n is denoted by $h^n = (c_n, x_{n+1}, c_{n+1}, x_{n+2}, \ldots)$. According to the Ionescu-Tulcea theorem (Proposition V.1.1 in Neveu (1965)), for any policy $\pi \in \Pi$, there exists a unique conditional probability measure $P^\pi(\cdot|x)$ on H^0 given an initial state $x_0 = x$. The expectation operator corresponding to this measure is E_x^π.

Let $H'_m = \underbrace{D \times \cdots \times D}_{(m+1)}$ for $m \geq 0$. An element of H'_m is denoted by $h'_m = (x_0, c_0, \cdots, x_m, c_m)$. Clearly, every function $w \in B(H'_m)$ can be regarded as a function from the space $B(H)$.

For a history $h'_n \in H'_n$ we define

$$U_n(h'_n) = u(x_0, c_0) + \delta(u(x_1, c_1) + \delta(u(x_2, c_2) + \ldots + \delta(u(x_n, c_n)) \ldots)). \tag{6}$$

[4] For convenience, we restrict our attention to non-randomized policies. We wish to emphasize that no improvements of the results can be obtained by allowing for randomized ones.

From Proposition 2 we infer that

$$U(x,h^0) := \lim_{n\to\infty} U_n(x,c_0,\ldots,x_0,c_0) = u(x,c_0) + \delta(u(x_1,c_1) + \delta(u(x_2,c_2) + \ldots))$$

exists in the set $\underline{R}$. Therefore, for an initial state $x_0 = x$ and $\pi \in \Pi$, we may define the *expected utility* as follows

$$V(x,\pi) := E_x^\pi U = \int_{H^0} U(x,h^0) P^\pi(dh^0|x). \tag{7}$$

In the classical set-up with constant discount factor $\beta \in (0,1)$, the discount function $\delta(r) = \beta r$ and the expected utility is of the form

$$V(x,\pi) = E_x^\pi \left(\sum_{t=0}^{n} \beta^t u(x_t,c_t) \right).$$

Let $\pi = \{p_n\} \in \Pi$ be any policy. If w is a function of $c \in A$, then $p_k w(h_k) := w(p_k(h_k))$, $h_k \in H_k$, for any $k \geq 0$. If $w \in B(H_{k+1})$, then we define

$$qw(h'_k) := \int_X w(h'_k, x_{k+1}) q(dx_{k+1}|x_k,c_k)$$

for $k \geq 0$. Note that if $w' \in B(H'_{k+1})$, then

$$qp_{k+1}w'(h'_k) := \int_X w(h'_k, x_{k+1}, p_{k+1}(h_{k+1})) q(dx_{k+1}|x_k,c_k).$$

According to the Ionescu-Tulcea theorem (Proposition V.1.1 in Neveu (1965)), for any policy $\pi \in \Pi$ and $m \geq 0$, there exists a unique conditional probability measure $P^\pi(\cdot|h_m)$ on H^m given $h_m \in H_m$. By $E_{h_m}^\pi$ we denote the expectation operator corresponding to the measure $P^\pi(\cdot|h_m)$. If $n > m$ and $w \in B(H'_n)$, then

$$E_{h_m}^\pi w := \int_{H^m} w(h_m,h^m) P^\pi(dh^m|h_m) = p_m q p_{m+1} \cdots q p_n w(h_m).$$

For a given history $h_m \in H_m$ and $\pi \in \Pi$, we define the *expected utility from period m onwards* as follows

$$V_m(h_m,\pi) := E_{h_m}^\pi U.$$

We note that $V_0 = V$ in (7).

For each $m \geq 0$ we define

$$P(h_m) := \{\nu : \nu = P^\pi(\cdot|h_m) \quad \text{for some} \quad \pi \in \Pi\}$$

as the set of attainable probability measures on the future given a history $h_m \in H_m$.

Definition 1 A policy $\hat{\pi} \in \Pi$ is called *persistently optimal* if

$$E_{h_m}^{\hat{\pi}} U = V_m(h_m) := \sup_{\nu \in P(h_m)} \int_{H^m} U(h_m,h^m) \nu(dh^m)$$

for $h_m \in H_m$ and each $m \geq 0$.

We shall need the following two sets of assumptions, which will be used alternatively.

Conditions (S):
(S1) For each $x \in X$ and every Borel set $\widetilde{X} \subset X$, the function $q(\widetilde{X}|x,\cdot)$ is continuous on $A(x)$.
(S2) The function $u(x,\cdot)$ is upper semicontinuous on $A(x)$ for every $x \in X$.

Conditions (W):
(W0) The set-valued mapping $x \mapsto A(x)$ is upper semicontinuous.
(W1) The transition law q is *weakly continuous*, that is,

$$\int_X \phi(y)q(dy|x,c)$$

is a continuous function of $(x,c) \in D$ for each bounded continuous function ϕ.
(W2) The function u is upper semicontinuous on D.

Note that Conditions (W) also allow us to consider *deterministic optimal growth models* with a continuous production function as in Stokey et al. (1989).

The dynamic programming models considered in this section were also studied by Hinderer (1970); Kertz and Nachman (1979); Schäl (1975, 1981) and extended to stochastic games by Nowak (1986).

4 Main results

Theorem 1 *Assume (A1) and (A2). If additionally either the set of Conditions (S) or (W) is satisfied, then there exists a persistently optimal policy.*

Proof Let us first assume (S).

PART I In this part we consider the bounded utility function:

$$u^l(x,c) = \max\{u(x,c), -l\} \quad \text{for } (x,c) \in D \text{ and } l \geq 1.$$

All functions related to u^l and defined in Section 3 will be denoted with the superscript l. Let $h_m \in H_m$. For any $n \geq m$, consider a finite horizon decision problem from period m till period n with the utility function U_n^l introduced in (6) with u replaced by u^l. A policy in this model is a sequence $\{p_m, \ldots, p_n\}$, with p_k defined in Section 3. By the backward induction, making use of Lemma 1(a) and Conditions (S), we conclude that there exists an optimal policy $\{p_m^o, \ldots, p_n^o\}$, i.e.,

$$v_{m,n}^l(h_m) := \sup_{\nu \in P(h_m)} \int_{H^m} U_n^l(h_m, h^m)\nu(dh^m) = p_m^o q p_{m+1}^o \ldots q p_n^o U_n^l(h_m).$$

This implies that $v_{m,n}^l$ is a Borel measurable function of h_m. We claim that $\{v_{m,n}^l(h_m)\}$ converges (as $n \to \infty$) to

$$V_m^l(h_m) = \sup_{\nu \in P(h_m)} \int_{H^m} U^l(h_m, h^m)\nu(dh^m) \tag{8}$$

uniformly in $h_m \in H_m$. Indeed, note that

$$
\begin{aligned}
&\sup_{h_m\in H_m} |v^l_{m,n}(h_m) - V^l_m(h_m)| \\
&= \sup_{h_m\in H_m} \left| \sup_{\nu\in P(h_m)} \int_{H^m} U^l_n(h_m,h^m)\nu(dh^m) - \sup_{\nu\in P(h_m)} \int_{H^m} U^l(h_m,h^m)\nu(dh^m) \right| \\
&\le \sup_{h_m\in H_m} \sup_{\nu\in P(h_m)} \int_{H^m} \left| U^l_n(h_m,h^m) - U^l(h_m,h^m) \right| \nu(dh^m) \\
&\le \sup_{h_m\in H_m} \sup_{\nu\in P(h_m)} \int_{H^m} \sup_{h\in H} \left| U^l_n(h_m,h^m) - U^l(h_m,h^m) \right| \nu(dh^m) \\
&= \sup_{h\in H} \left| U^l_n(h) - U^l(h) \right|.
\end{aligned}
$$

Now making use of Theorem 1 in Jaśkiewicz et al. (2011) , we conclude that $\sup_{h\in H} \left| U^l_n(h) - U^l(h) \right| \to 0$ as $n\to\infty$. Hence, V^l_m is Borel measurable. Using the uniform convergence showed above one can easily see that

$$
V^l_m(h_m) = \max_{c\in A(x_m)} \int_X V^l_{m+1}(h_{m+1}) q(dx_{m+1}|x_m,c) \tag{9}
$$

for any m. (Recall that x_k is the last component of h_k, $k=m,m+1$.) By Lemma 1(a), for any m there exists a Borel measurable function π^*_m of $h_m\in H_m$ such that the maximum is attained in (9) at the point $\pi^*_m(h_m)$. Thus, we have

$$
V^l_m(h_m) = \pi^*_m q V^l_{m+1}(h_m) = \int_X V^l_{m+1}(h_{m+1}) q(dx_{m+1}|x_m,\pi^*_m(h_m)), \quad h_m\in H_m. \tag{10}
$$

Iterating (10), we obtain for any m and $k>m$ that

$$
V^l_m(h_m) = \pi^*_m q \pi^*_{m+1} q \cdots \pi^*_k q V^l_{k+1}(h_m). \tag{11}
$$

Let $\pi^* := \{\pi^*_m\}$ with π^*_m defined by (10) for each m. We show that

$$
\lim_{k\to\infty} \pi^*_m q \pi^*_{m+1} q \cdots \pi^*_k q V^l_{k+1}(h_m) = E^{\pi^*}_{h_m} U^l. \tag{12}
$$

By the triangle inequality, we have that

$$
\begin{aligned}
&|\pi^*_m q \pi^*_{m+1} q \cdots \pi^*_k q V^l_{k+1}(h_m) - E^{\pi^*}_{h_m} U^l| \\
&\le \left| \pi^*_m q \pi^*_{m+1} q \cdots \pi^*_k q V^l_{k+1}(h_m) - \pi^*_m q \pi^*_{m+1} q \cdots \pi^*_k U^l_k(h_m) \right| \\
&\quad + \left| \pi^*_m q \pi^*_{m+1} q \cdots \pi^*_k U^l_k(h_m) - E^{\pi^*}_{h_m} U^l \right|.
\end{aligned} \tag{13}
$$

Making use again of Theorem 1 in Jaśkiewicz et al. (2011), the second term in (13) tends to 0 uniformly on H_m. Now let us consider the the first term in (13) and note that

$$
\begin{aligned}
&\left| \pi^*_m q \pi^*_{m+1} q \cdots \pi^*_k q V^l_{k+1}(h_m) - \pi^*_m q \pi^*_{m+1} q \cdots \pi^*_k U^l_k(h_m) \right| \\
&\le \pi^*_m q \pi^*_{m+1} q \cdots \pi^*_{k-1} q \left| \pi^*_k q V^l_{k+1} - \pi^*_k U^l_k \right| (h_m).
\end{aligned} \tag{14}
$$

Using (8) and (14), we get

$$
\begin{aligned}
&\left|\pi_k^* q V_{k+1}^l(h_k) - \pi_k^* U_k^l(h_k)\right| \\
&= \left|\int_X \sup_{\nu\in P(h_{k+1})} \int_{H^{k+1}} U^l(h_{k+1},h^{k+1})\nu(dh^{k+1}) q(dx_{k+1}|x_k,\pi_k^*(h_k)) - U_k^l(h_k,\pi^*(h_k))\right| \\
&\le \int_X \sup_{\nu\in P(h_{k+1})} \int_{H^{k+1}} \left|U^l(h_{k+1},h^{k+1}) - U_k^l(h_k,\pi^*(h_k))\right| \nu(dh^{k+1}) q(dx_{k+1}|x_k,\pi_k^*(h_k)) \\
&\le \sup_{h\in H} |U^l(h) - U_k^l(h)| \to 0.
\end{aligned}
$$

The former inequality in the above expression is due to the fact that U_k^l depends neither on x_{k+1} nor h^{k+1}. From (8), (11) and (12), it follows that π^* is persistently optimal.

Part II Note that $u^l \searrow u$. Since, δ is nondecreasing by (A2), it follows that $\{V_m^l\}$ is non-increasing for every $m \ge 1$. Therefore,

$$V_m(h_m) := \lim_{l\to\infty} V_m^l(h_m)$$

exists in $\underline{R}$ for $h_m \in H_m$. Moreover, letting $l \to \infty$ in (9) and making use of Proposition 10.1 in Schäl (1975) and the dominated convergence theorem we conclude that

$$
\begin{aligned}
V_m(h_m) &= \lim_{l\to\infty} \max_{c\in A(x_m)} \int_X V_{m+1}^l(h_{m+1}) q(dx_{m+1}|x_m,c) \\
&= \max_{c\in A(x_m)} \lim_{l\to\infty} \int_X V_{m+1}^l(h_{m+1}) q(dx_{m+1}|x_m,c) = \max_{c\in A(x_m)} \int_X V_{m+1}(h_{m+1}) q(dx_{m+1}|x_m,c).
\end{aligned}
$$

By Lemma 1(a), for any m there exists a Borel measurable function $\hat{\pi}_m$ of $h_m \in H_m$ such that the maximum is attained at the point $\hat{\pi}_m(h_m)$ in the above display Thus, we have

$$V_m(h_m) = \hat{\pi}_m q V_{m+1}(h_m) = \int_X V_{m+1}(h_{m+1}) q(dx_{m+1}|x_m,\hat{\pi}_m(h_m)), \quad h_m \in H_m. \tag{15}$$

Iterating this equality, we obtain for any m and $k > m$ that

$$V_m(h_m) = \hat{\pi}_m q \hat{\pi}_{m+1} q \cdots \hat{\pi}_k q V_{k+1}(h_m).$$

Let $\hat{\pi} := \{\hat{\pi}_m\}$ with $\hat{\pi}_m$ defined by (15) for each m. Since δ is nondecreasing, it follows that

$$\hat{\pi}_m q \hat{\pi}_{m+1} q \cdots \hat{\pi}_k q V_{k+1}(h_m) \le \hat{\pi}_m q \hat{\pi}_{m+1} q \cdots \hat{\pi}_k q V_{k+1}^l(h_m) \quad \text{for } l \ge 1. \tag{16}$$

Now letting $k \to \infty$ in (16) and using part I (see with (12)), we deduce that

$$\lim_{k\to\infty} \hat{\pi}_m q \hat{\pi}_{m+1} q \cdots \hat{\pi}_k q V_{k+1}(h_m) \le E_{h_m}^{\hat{\pi}} U^l \quad \text{for } l \ge 1.$$

From Proposition 3, it follows that $U_l \searrow U$, and therefore, the dominated convergence theorem yields

$$\lim_{k\to\infty} \hat{\pi}_m q \hat{\pi}_{m+1} q \cdots \hat{\pi}_k q V_{k+1}(h_m) \le E_{h_m}^{\hat{\pi}} U \quad \text{for } l \ge 1. \tag{17}$$

On the other hand, from (8), we obtain

$$V_m^l(h_m) \ge \int_{H^m} U^l(h_m,h^m)\nu(dh^m) \quad \text{for each } \nu \in P(h_m).$$

Letting $l \to \infty$ and making use of Proposition 3 and the dominated convergence theorem, we conclude that

$$V_m(h_m) \geq \int_{H^m} U(h_m, h^m)\nu(dh^m) \quad \text{for each } \nu \in P(h_m). \tag{18}$$

From (17) and (18) it follows that $\hat{\pi}$ is persistently optimal policy.

The proof under Conditions (W) makes use of Lemma 1(b) and proceeds analogously as in the case of Conditions (S). In fact, the existence of a persistently optimal policy under (W) for a bounded utility function was established by Kertz and Nachman (1979) (see Theorem 3.4). However, their proof is more complicated than ours. □

Remark 1 It is worth emphasizing that the existence of persistently optimal policies under (W) was already shown by Kertz and Nachman (1979), see Theorem 5.2. However, their proof proceeds along different lines than ours, since they do not assume (C). In consequence, their route to existence is more involved. On the other hand, the result under Conditions (C) and (S) has not been clearly presented so far, and only some remarks were given by Schäl (1981).

Remark 2 Persistently optimal policies for non-stationary decision processes are counterparts of stationary optimal ones in stationary dynamic programming. This concept was extensively used in gambling theory, see Dubins and Savage (1976).

5. Examples

Example 1 Theorem 1 can be applied to the theory of stochastic optimal growth. We have in mind classical models studied Stokey et al. (1989), but with generalized discounting. Let $X = [0, s]$ be the set of all *capital stocks* where $s > 1$. If x_t is a capital stock at the beginning of period t, then consumption c_t in this period belongs to $A(x_t) := [0, x_t]$. The utility of consumption c_t is $u(c_t)$ where $u : X \mapsto \underline{R}$ is a fixed function. The evolution of the state process is described by some function f of the investment for the next period $y_t := x_t - c_t$ and some random variable ξ_t. In the literature, f is called *production technology*, see Stokey et al. (1989). We shall view this model as a decision process with $X = [0, s]$, $A(x) = [0, x]$, and $u(x, c) = u(c)$, $x \in X$, $c \in A(x)$. Suppose that $\{\xi_t\}$ are independent and have a common probability distribution μ with support included in $[0, z]$ for some $z > 1$. Assume that

$$x_{t+1} = f(x_t - c_t)\xi_t, \quad \text{for } t = 0, 1, \ldots,$$

where $f : X \mapsto R$ is a continuous and increasing function, $f(0) = 0$,

$$(0, \infty) \ni y \to \frac{f(y)}{y} \quad \text{is strictly decreasing;} \tag{19}$$

$$\lim_{y \to 0+} \frac{f(y)}{y} > 1 \quad \text{and} \quad \frac{f(s)}{s} < 1. \tag{20}$$

Conditions (19)-(20) imply that there exists $y_0 > 0$ such that

$$f(y) > y \quad \text{for all} \quad y \in (0, y_0) \quad \text{and} \quad f(y) < y \quad \text{for all} \quad y > y_0. \tag{21}$$

We assume that $f(s)z \leq s$. Then for any $x_t \in X$, $c_t \in A(x_t)$ and $\xi_t \in [0, z]$, $x_{t+1} = f(x_t - c_t)\xi_t \in X$. Observe that the transition probability q is of the form,

$$q(B|x, c) = \int_0^s 1_B(f(x - c)\xi)\mu(d\xi),$$

where $B \subset X$ is any Borel set, $x \in X$, $c \in A(x)$. Here, 1_B is the indicator function of the set B. If v is a continuous function on X, then then the integral

$$\int_X v(y)q(dy|x, c) = \int_0^s v(f(x - c)\xi)\mu(d\xi)$$

depends continuously on (x, c). This example allows for $u(c) = \log c$ as a utility function where $\log 0 = -\infty$. For any non-linear discount function δ satisfying (A1) and (A2), there exists a persistently optimal policy.

Example 2 The inventory model. A manager sells a certain amount of goods each period $t = 0, 1, \ldots$ at price p. If he has $x_t \geq 0$ units in stock, he can sell $\min\{x_t, D_t\}$, where $D_t \geq 0$ is a *continuous* random variable representing an unknown demand. He can also order any amount c_t of new goods to be delivered at the beginning of next period at a cost $l(c_t)$ paid now. It is assumed that l is continuous, increasing and $l(0) = 0$. The system equation is of the form:

$$x_{t+1} = x_t - \min\{x_t, D_t\} + c_t, \quad \text{for } t = 0, 1, \ldots,$$

where $\{D_t\}$ is a sequence of i.i.d. random variables such that each D_t follows a distribution F and $ED_t < \infty$. The manager discounts his revenues according to a function δ satisfying (A1)-(A2).

This model can be viewed as a dynamical system, in which $X := [0, \infty)$ is the state space (i.e., the set of possible levels of stock), $A = A(x) := [0, K]$ is the action space, where $K > 0$ and $u(x, c) := Ep \min\{x, D\} - l(c)$ is the immediate return function, where $D \sim F$.

Clearly, $u(x, c) \leq pED$. Next note that he transition probability q is of the form,

$$q(B|x, c) = \int_0^\infty 1_B(x - \min\{x, y\} + c)dF(y),$$

where $B \subset X$ is any Borel set, $x \in X$, $c \in A$. If v is a continuous function on X, then the integral

$$\begin{aligned}\int_X v(y)q(dy|x, c) &= \int_0^\infty v(x - \min\{x, y\} + c)dF(y) = \int_0^x v(x - y + c)dF(y) + \int_x^\infty v(c)dF(y) \\ &= \int_0^x v(x - y + c)dF(dy) + v(c)(1 - F(x))\end{aligned}$$

depends continuously on (x, c). Hence, the model satisfies Conditions (W). Therefore, for any non-linear discount function δ satisfying (A1) and (A2), there exists a persistently optimal policy.

6 Basic convergence results

Let assumptions (A1) and (A2) be satisfied. By $\{r_n\}_0^\infty$ we denote a sequence such that $r_n \in \underline{R}$ and $r_n \le l$ for each $n \ge 0$ and some $l > 0$. Define the following functions

$$w_0(r_0) = r_0, \quad w_n(r_0, r_1, \ldots, r_n) = r_0 + \delta(w_{n-1}(r_1, \ldots, r_n)) \quad \text{for } n \ge 1.$$

Note that w_n defined above is a function of $n+1$ variables. We first use the functions w_n for the sequence $\{r_n\}_0^\infty$ with $r_n = l$ for all $n \ge 0$. Put $l_{n+1} := (r_0, ..., r_n)$ if $r_t = l$ for $t = 0, ..., n$, $n \ge 0$.

In the proof of our first result we use a simple argument from Cho and O'Regan (2008).

Proposition 1 *There exists* $L := \lim_{m\to\infty} w_m(l_{m+1}) = \sup_{m\ge 1} w_m(l_{m+1}) < \infty$.

Proof Note that (since the function δ is non-decreasing), for each $m \ge 1$,

$$w_m(l_{m+1}) \ge w_{m-1}(l_m).$$

Hence, the sequence $\{w_m(l_{m+1})\}$ is non-decreasing. We show that its limit is finite. Indeed, observe that by (A1)

$$w_1(l,l) - w_0(l) = l + \delta(l) - l \le \gamma(l),$$
$$w_2(l,l,l) - w_1(l,l) = l + \delta(w_1(l,l)) - l - \delta(w_0(l)) \le \gamma(w_1(l,l) - w_0(l)) \le \gamma^{(2)}(l).$$

Continuing this procedure one can see that

$$w_m(l_{m+1}) - w_{m-1}(l_m) \le \gamma^{(m)}(l).$$

Let $\epsilon > 0$ be fixed. Since $\gamma^{(m)}(l) \to 0$ as $m \to \infty$, there exists $m \ge 1$ such that

$$w_m(l_{m+1}) - w_{m-1}(l_m) \le \epsilon - \gamma(\epsilon).$$

Note now that

$$\begin{aligned}
w_{m+1}(l_{m+2}) - w_{m-1}(l_m) &= w_{m+1}(l_{m+2}) - w_m(l_{m+1}) + w_m(l_{m+1}) - w_{m-1}(l_m)\\
&\le l + \delta(w_m(l_{m+1})) - l - \delta(w_{m-1}(l_m)) + \epsilon - \gamma(\epsilon)\\
&\le \gamma(w_m(l_{m+1}) - w_{m-1}(l_m)) + \epsilon - \gamma(\epsilon)\\
&\le \gamma(\epsilon - \gamma(\epsilon)) + \epsilon - \gamma(\epsilon) \le \gamma(\epsilon) + \epsilon - \gamma(\epsilon) = \epsilon.
\end{aligned}$$

Similarly,

$$\begin{aligned}
w_{m+2}(l_{m+3}) - w_{m-1}(l_m) &= w_{m+2}(l_{m+3}) - w_m(l_{m+1}) + w_m(l_{m+1}) - w_{m-1}(l_m)\\
&\le l + \delta(w_{m+1}(l_{m+2})) - l - \delta(w_{m-1}(l_m)) + \epsilon - \gamma(\epsilon)\\
&\le \gamma(w_{m+1}(l_{m+2}) - w_{m-1}(l_m)) + \epsilon - \gamma(\epsilon)\\
&\le \gamma(\epsilon) + \epsilon - \gamma(\epsilon) = \epsilon.
\end{aligned}$$

Thus, by induction we obtain that

$$w_{m+k}(l_{m+k+1}) - w_{m-1}(l_m) \le \epsilon$$

for all $k \ge 0$. Hence, $w_{m+k}(l_{m+k+1})) \le w_{m-1}(l_m) + \epsilon$. Since $w_{m-1}(l_m)$ is finite, it follows that L is finite. □

Consider a sequence $\{u_n\}_0^\infty$ of utilities $u_n \in \underline{R}$ such that $u_n \le l$ for each $n \ge 0$.

For $m \ge 1$ and $n \ge 0$ let us introduce the following notation

$$W_{n,m}(u_0, \dots, u_n, l_m) := w_{n+m}(u_0, \dots, u_n, l_m).$$

Note that $W_{n,m}$ is a function of a $(n + m + 1)$-dimensional vector. For example,

$$W_{2,3}(u_0, u_1, u_2, l_3) = W_{2,3}(u_0, u_1, u_2, l, l, l) = u_0 + \delta(u_1 + \delta(u_2 + \delta(l + \delta(l + \delta(l))))),$$

$$W_{3,2}(u_0, u_1, u_2, u_3, l_2) = W_{4,2}(u_0, u_1, u_2, u_3, l, l) = u_0 + \delta(u_1 + \delta(u_2 + \delta(u_3 + \delta(l + \delta(l))))).$$

Proposition 2 *The limit of the sequence $\{w_n(u_0, \dots, u_n)\}$ exists in $\underline{R}$.*

Proof We first study the case where $u_n > -\infty$ for all $n \ge 0$. Note that for each $n \ge 0$ and $m \ge 1$

$$w_n(u_0, \dots, u_n) \le W_{n,m}(u_0, \dots, u_n, l_m). \tag{22}$$

Moreover,

$$\begin{aligned}&W_{n,m}(u_0, \dots, u_n, l_m) - w_n(u_0, \dots, u_n)\\ &= u_0 + \delta(W_{n-1,m}(u_1, \dots, u_n, l_m)) - u_0 - \delta(w_{n-1}(u_1, \dots, u_n))\\ &\le \gamma(W_{n-1,m}(u_1, \dots, u_n, l_m) - w_{n-1}(u_1, \dots, u_n)).\end{aligned}$$

Similarly,

$$\begin{aligned}&W_{n-1,m}(u_1, \dots, u_n, l_m) - w_{n-1}(u_1, \dots, u_n)\\ &= u_0 + \delta(W_{n-2,m}(u_2, \dots, u_n, l_m)) - u_0 - \delta(w_{n-2}(u_2, \dots, u_n))\\ &\le \gamma(W_{n-2,m}(u_2, \dots, u_n, l_m) - w_{n-2}(u_2, \dots, u_n)).\end{aligned}$$

Thus

$$W_{n,m}(u_0, \dots, u_n, l_m) - w_n(u_0, \dots, u_n) \le \gamma^{(2)}(W_{n-2,m}(u_2, \dots, u_n, l_m) - w_{n-2}(u_2, \dots, u_n)).$$

Continuing in this way and using Proposition 1, we obtain

$$\begin{aligned}W_{n,m}(u_0, \dots, u_n, l_m) - w_n(u_0, \dots, u_n) &\le \gamma^{(n)}(W_{0,m}(u_n, l_m) - w_0(u_n))\\ &= \gamma^{(n)}(\delta(w_{m-1}(l_m)) \le \gamma^{(n+1)}(L).\end{aligned}$$

for all $m \geq 1$. Let $\epsilon > 0$ be fixed. Then, for sufficiently large n, say $n > N_1$,

$$W_{n,m}(u_0, \ldots, u_n, l_m) \leq w_n(u_0, \ldots, u_n) + \epsilon \tag{23}$$

for all $m \geq 1$. Clearly, for any $m \geq 1$, we have

$$\begin{aligned} &W_{n,m}(u_0, \ldots, u_n, l_m) \leq W_{n,m+1}(u_0, \ldots, u_n, l_{m+1}) \\ &\quad \leq w_{n+m+1}(l_{n+m+2}) \leq \sup_{t \geq 1} w_t(l_{t+1}) = L < \infty. \end{aligned}$$

Therefore, $\lim_{m\to\infty} W_{n,m}(u_0, \ldots, u_n, l_m)$ exists and is bounded from above by L. Let us denote this limit by G_n. From (22) and (23) we conclude that

$$G_n - \epsilon \leq w_n(u_0, \ldots, u_n) \leq G_n$$

for $n > N_1$. Observe that $\{G_n\}$ is decreasing and $G_* := \lim_{n\to\infty} G_n$ exists in $\underline{R}$. Hence, the limit

$$\lim_{n\to\infty} w_n(u_0, \ldots, u_n)$$

also exists and equals G_*. Assume now that $u_n = -\infty$ for some $n \geq 0$. Then

$$w_n(u_0, \ldots, u_n) = -\infty$$

and

$$W_{n,m}(u_0, \ldots, u_n, l_m) = -\infty$$

for all $m \geq 1$. Therefore,

$$G_n = -\infty \quad \text{and} \quad \lim_{n\to\infty} w_n(u_0, \ldots, u_n) = G_* = \lim_{n\to\infty} G_n = -\infty.$$

□

For $k \geq 1$ let us define $u_n^k := \max\{u_n, -k\}$. Then, we have arrived at our final result in this section.

Proposition 3 $\lim_{n\to\infty} w_n(u_0, \ldots, u_n) = \inf_{k\geq 1} \lim_{n\to\infty} w_n(u_0^k, \ldots, u_n^k)$.

Proof Assume first that $\lim_{n\to\infty} w_n(u_0, \ldots, u_n) > -\infty$. Let $\epsilon > 0$ be fixed. Then, by (23), we have

$$w_n(u_0, \ldots, u_n) \geq W_{n,m}(u_0, \ldots, u_n, l_m) - \epsilon$$

for all $m \geq 1$ and $n > N_1$. Moreover, there exists N_2 such that for all $n > N_2$

$$\lim_{t\to\infty} w_t(u_0, \ldots, u_t) \geq w_n(u_0, \ldots, u_n) - \epsilon. \tag{24}$$

Let us now fix $n > \max\{N_1, N_2\}$. Since $u_i^k \to u_i$ for $i \in \{0, \ldots, n\}$ as $k \to \infty$, then there exists $K_1 > 0$ such that for $k > K_1$

$$u_i^k - u_i \leq \frac{\epsilon}{n+1} \quad \text{for each } i \in \{0, \ldots, n\}.$$

Fix any $k > K_1$. By assumption (A1) we obtain that

$$\begin{aligned} w_n(u_0^k, \dots, u_n^k) - w_n(u_0, \dots, u_n) &\le u_0^k - u_0 + \gamma(w_{n-1}(u_1^k, \dots, u_n^k) - w_{n-1}(u_1, \dots, u_n)) \\ &\le \frac{\epsilon}{n+1} + w_{n-1}(u_1^k, \dots, u_n^k) - w_{n-1}(u_1, \dots, u_n). \end{aligned}$$

Similarly,

$$\begin{aligned} w_{n-1}(u_1^k, \dots, u_n^k) - w_{n-1}(u_1, \dots, u_n) &\le u_1^k - u_1 + \gamma(w_{n-2}(u_2^k, \dots, u_n^k) - w_{n-2}(u_2, \dots, u_n)) \\ &\le \frac{\epsilon}{n+1} + w_{n-2}(u_2^k, \dots, u_n^k) - w_{n-2}(u_2, \dots, u_n). \end{aligned}$$

Hence

$$w_n(u_0^k, \dots, u_n^k) - w_n(u_0, \dots, u_n) \le \frac{2\epsilon}{n+1} + w_{n-2}(u_2^k, \dots, u_n^k) - w_{n-2}(u_2, \dots, u_n).$$

Proceeding along this line, we finally obtain

$$w_n(u_0^k, \dots, u_n^k) - w_n(u_0, \dots, u_n) \le \frac{n\epsilon}{n+1} + w_0(u_n^k) - w_0(u_n) = \frac{n\epsilon}{n+1} + u_n^k - u_n \le \epsilon.$$

Similarly, for any $m \ge 1$, we have that

$$\begin{aligned} &W_{n,m}(u_0^k, \dots, u_n^k, l_m) - W_{n,m}(u_0, \dots, u_n, l_m) \\ &= w_{n+m}(u_0^k, \dots, u_n^k, l_m) - w_{n+m}(u_0, \dots, u_n, l_m) \le \epsilon. \end{aligned}$$

Hence, we infer that

$$\begin{aligned} w_n(u_0, \dots, u_n) &\ge W_{n,m}(u_0, \dots, u_n, l_m) - \epsilon \\ &\ge W_{n,m}(u_0^k, \dots, u_n^k, l_m) - 2\epsilon \\ &= w_{n+m}(u_0^k, \dots, u_n^k, l_m) - 2\epsilon \\ &\ge w_{n+m}(u_0^k, \dots, u_n^k, u_{n+1}^k, \dots, u_{n+m}^k) - 2\epsilon. \end{aligned} \tag{25}$$

By Proposition 2, we deduce that

$$\lim_{m\to\infty} w_{n+m}(u_0^{k'}, \dots, u_n^{k'}, u_{n+1}^{k'}, \dots, u_{n+m}^{k'}) =: G_*^{k'}$$

exists for any positive integer k'. Therefore, by (25), we have

$$w_n(u_0, \dots, u_n) \ge G_*^k - 2\epsilon, \quad k > K_1.$$

Now applying (24) we obtain that

$$\lim_{t\to\infty} w_t(u_0, \dots, u_t) \ge G_*^k - 3\epsilon \ge \inf_{k'\ge 1} G_*^{k'} - 3\epsilon.$$

Since $\epsilon > 0$ was arbitrary, we get that

$$\lim_{t\to\infty} w_t(u_0, \dots, u_t) \ge \inf_{k'\ge 1} G_*^{k'}.$$

On the other hand, it is obvious that

$$\lim_{t\to\infty} w_t(u_0,\dots,u_t) \le \inf_{k\ge 1}\lim_{t\to\infty} w_t(u_0^k,\dots,u_t^k) = \inf_{k\ge 1} G_*^k.$$

Combining the last two inequalities we get the conclusion.

If $\lim_{n\to\infty} w_n(u_0,\dots,u_n) = -\infty$, then for any $M < 0$ there exists N_3 such that for all $n > N_3$

$$M > w_n(u_0,\dots,u_n).$$

Proceeding as above we obtain that $\inf_{k\ge 1} G_*^k = -\infty$. If $u_n = -\infty$ for some $n \ge 0$, then the proof that

$$\inf_{k\ge 1}\lim_{n\to\infty} w_n(u_0^k,\dots,u_n^k) = \lim_{n\to\infty} w_n(u_0,\dots,u_n) = -\infty,$$

is very simple. □

References

Becker, R.A., Boyd III, J.H., 1997. Capital Theory, Equilibrium Analysis and Recursive Utility. Blackwell Publishers, New York.

Berge, C., 1963. Topological Spaces. MacMillan, New York.

Bertsekas, D.P., 1977. Monotone mappings with application in dynamic programming, SIAM Journal on Control and Optimization 15, 438-464.

Bertsekas, D.P., Shreve, S.E., 1978. Stochastic Optimal Control: the Discrete Time Case. Academic Press, New York.

Blackwell, D., 1965. Discounted dynamic programming. Annals of Mathematical Statistics 36, 226-235.

Boyd III, J.H., 1990. Recursive utility and the Ramsey problem. Journal of Economic Theory 50, 326-345.

Brock, W.A., Mirman, L.J., 1972. Optimal economic growth and uncertainty: the discounted case. Journal of Economic Theory 4, 479-513.

Brown, L.D., Purves, R., 1973. Measurable selections of extrema. Annals of Statistics 1, 902-912.

Cho, Y., O'Regan, D., 2008. Fixed point theory for Volterra contractive operators of Matkowski type in Frechet spaces. Dynamics of Continuous, Discrete and Impulsive Systems, Series A: Mathematical Analysis 15, 871-884.

Denardo, E.V., 1967. Contraction mappings in the theory underlying dynamic programming. SIAM Review 9, 165-177.

Dubins, L.E., Savage, L.J., 1976. Inequalities for Stochastic Processes (How to Gamble if You Must). Dover. New York.

Feinberg, E.A., 2002. Total reward criteria. In: Feinberg, E.J., Shwartz, A. (Eds) Handbook of Markov decision processes: theory and methods, Kluwer Academic Publishers, Dordrecht, The Netherlands, 173-208.

Feinberg, E.A., Shwartz, A., (Eds) 2002. Handbook of Markov decision processes: theory and methods. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Grandmont, J.M., 1977. Temporary general equilibrium theory. Econometrica 45, 535-572.

Hicks, J.R., 1965. Capital and Growth. Oxford University Press, Oxford.

Hinderer, K., 1970. Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter. Lecture Notes in Operations Research 33, Springer-Verlag, NY.

Jaśkiewicz, A., Matkowski, J., Nowak, A., 2011. On variable discounting in dynamic programming: applications to resource extraction and other economic models.

Kertz, R.P., Nachman, D.C., 1979. Persistently optimal plans for nonstationary dynamic programming: the topology of weak convergence case. Annals of Probability 7, 811-826.

Koopmans, T.C., 1960. Stationary ordinal utility and impatience. Econometrica 28, 287-309.

Kreps, D.M., Porteus, E.L., 1978. Temporal resolution of uncertainty and dynamic choice theory. Econometrica 46, 185-200.

Lucas Jr., R.E., Stokey, N., 1984. Optimal growth with many consumers. Journal of Economic Theory 32, 139-171.

Matkowski, J., 1975. Integral solutions of functional equations. Dissertationes Mathematicae 127, 1-68.

Neveu, J., 1965. Mathematical Foundations of the Calculus of Probability. Holden-Day, San Francisco.

Nowak, A.S., 1986. Semicontinuous nonstationary stochastic games. Journal of Mathematical Analysis and Applications 117, 84-99.

Porteus, E., 1982. Conditions for characterizing the structure of optimal strategies in infinite-horizon dynamic programs. Journal of Optimization Theory and Applications 36, 419-431.

Samuelson, P., 1937. A note on measurement of utility. Review of Economic Studies 4, 155-161.

Schäl M., 1975. Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal. Z Wahrsch verwandte Gebiete 32, 179-196.

Schäl, M., 1975. On dynamic programming: compactness of the space of policies. Stochastic Processes and their Applications 3, 345-364.

Schäl, M., 1981. Utility functions and optimal policies in sequential decision problems. In: Game Theory and Mathematical Economics (Moeschlin, O., Pallaschke, D., eds.), North-Holland, Amsterdam, 357-365.

Stokey, N.L., Lucas, Jr., R.E., Prescott, E., 1989. Recursive Methods in Economic Dynamics. Harvard University Press, Cambridge, MA.

Strauch, R., 1966. Negative dynamic programming. Annals of Mathematical Statistics 37, 871–890.