KYBERNETIKA — VOLUME *52* (2016), NUMBER 3, PAGES 403–426

DISCRETE-TIME MARKOV CONTROL PROCESSES WITH RECURSIVE DISCOUNT RATES

Yofre H. García and Juan González-Hernández

This work analyzes a discrete-time Markov Control Model (MCM) on Borel spaces when the performance index is the expected total discounted cost. This criterion admits unbounded costs. It is assumed that the discount rate in any period is obtained by using recursive functions and a known initial discount rate. The classic dynamic programming method for finite-horizon case is verified. Under slight conditions, the existence of deterministic non-stationary optimal policies for infinite-horizon case is proven. Also, to find deterministic non-stationary $\epsilon-$optimal policies, the value-iteration method is used. To illustrate an example of recursive functions that generate discount rates, we consider the expected values of stochastic processes, which are solutions of certain class of Stochastic Differential Equations (SDE) between consecutive periods, when the initial condition is the previous discount rate. Finally, the consumption-investment problem and the discount linear-quadratic problem are presented as examples; in both cases, the discount rates are obtained using a SDE, similar to the Vasicek short-rate model.

Keywords: dynamic programming method, optimal stochastic control

Classification: 49L20, 93E20

1. INTRODUCTION

This work deals with a discrete-time MCM with non-stationary discount rate and possibly unbounded cost on Borel spaces. The performance index considered is

$$J = E\Big[\sum_{n<T} e^{-S_n} c(x_n, a_n)\Big], \tag{1}$$

where $S_n = r_0 + \cdots + r_{n-1}$ is the sum of the discount rates applied in previous periods, and the control a_n depends on the state x_n and the discount rate r_n. The discount rates satisfy the recursive relations

$$r_n := R_n(r_{n-1}),$$

where, R_n is a measurable function, and r_0 is the initial discount rate, $n = 1, 2, \ldots$

This class of MCMs can be used to build models for small investors, businessmen or entrepreneurs, where it is assumed that the discount rates are exogenous variables,

DOI: 10.14736/kyb-2016-3-0403

changing in each period depending on financial market conditions and the previous discount rate. Hence, non-stationary discount rates turn out to be a better modelling option in these cases, and deterministic non-stationary policies are the solution for this type of Markov control problem. To the best of our knowledge, this is the first work on discounted MCMs, where the discount rates satisfies a type of recursive relation. An example of these recursive relations for the discount rates, can be obtained with a certain type of interest rate models, best known in financial literature as Short-Rate Models, where the interest rate evolution follows a Stochastic Differential Equation (SDE) (see [20] and Remark 3.9, part 3), and the recursive functions are obtained as the expected value of the solution of the SDEs defined between consecutive periods.

Among the discounted MCMs in discrete-time with non-constant rates are the following:

Feinberg and Shwartz [11] consider an MCM with a finite number of discount rates. The performance index criterion is defined by

$$J(\pi, x) = \sum_{i=1}^{k} E_x^{\pi}\left[\sum_{t=0}^{\infty} (\beta_i)^t c_i(x_t, a_t)\right], \quad x \in X, \quad \pi \in \Pi.$$

They establish conditions for the existence of $\epsilon-$optimal strategies. Subsequently, Carmon and Shwartz [7] proposed a discount function $h(t)$

$$J(\pi, x) = E_x^{\pi}\left[\sum_{t=0}^{\infty} h(t) c(x_t, a_t)\right], \quad x \in X, \quad \pi \in \Pi,$$

where h satisfies the condition $|h(t)| \leq k\beta^t$, for some $0 < \beta < 1$. They prove the existence of $\epsilon-$optimal strategies under stationary optimal control tools.

Della Vecchia et al. [8] proposed a similar non-stationary MCM with non-constant deterministic bounded discount factor $\lambda_t \leq \rho^{t+1},\ 0 < \rho < 1$, with the performance index criterion

$$J(\pi, x) = E_x^{\pi}\left[\sum_{t=0}^{\infty} \lambda_t c(x_t, a_t)\right], \quad x \in X, \quad \pi \in \Pi. \tag{2}$$

As this MCM is time-dependent and under the hypothesis that the cost are uniformly bounded, they transform this model to a stationary MCM. Also, they define an appropriate dynamic operator on this new model, and prove, by means of fixed-point of particular contraction operator, the existence of deterministic stationary optimal policies. These stationary optimal policies, obtained for the stationary MCM, generate Markov optimal policies in the original one.

Hinderer [18] and Schäl [25] consider the case when the cost function is bounded and the discount factors depend on states and controls:

$$J(\pi, x) = E_x^{\pi}\Big[c(x_0, a_0) + \sum_{t=0}^{\infty} \prod_{i=0}^{t} \beta(x_i, a_i, x_{i+1}) c(x_t, a_t, x_{t+1})\Big], \quad x \in X, \ \pi \in \Pi.$$

They give conditions to guarantee the existence of an optimal policy. In this sense, Wei and Guo [27] propose a similar case when the cost can be unbounded and the discount rates depend only on the current state

$$J(\pi, x) = E_x^\pi \left[c(x_0, a_0) + \sum_{t=1}^{\infty} \prod_{i=0}^{t-1} \beta(x_i) c(x_t, a_t) \right], \quad x \in X, \quad \pi \in \Pi,$$

and they give conditions for guaranteeing the existence of an optimal solution. Guo et al. [15] consider a non-stationary dynamic where the performance index is a first passage type. Ye and Guo [28] consider the continuous-time case. A similar model using convex programming is proposed by Zhang [29]. Minjarez-Sosa [17] works on discrete-time Markov control models with non-constant discount factors of the form

$$\Gamma_n = \prod_{k=0}^{n-1} \alpha(x_k, a_k, \xi_{k+1}), \quad n \in \mathbb{N}, \text{ and } \Gamma_0 = 1,$$

where x_k, a_k, and ξ_{k+1} are the state, the action, and a random disturbance at time n,

$$V(\pi, x) = E_x^\pi \left[\sum_{n=0}^{\infty} \Gamma_n c(x_n, a_n) \right], \quad x \in X, \quad \pi \in \Pi.$$

It is assumed that the random disturbance process $\{\xi_k\}$ is formed by observable independent and identically distributed random variables and the distributions are unknown. Minjarez-Sosa introduces an estimation and control procedure to find asymptotically optimal policies. Also he studies the minimax control problems when the random disturbance process is non-observable.

The assumptions in the last four discounted MCMs can be used to optimize systems where the controls and previous states affect the current discount rates. For example, can be used to model the actions of very big economic agents such as central banks or monopolies, but are not appropriate to model the actions of small economic agents.

González-Hernández et al. [12] assume MCMs where the discount factor has an exponential form and the discount rate is modelled as a non-negative Markov chain $\{r_n : n \in \mathbb{N}\}$ over $(0, \infty)$:

$$J(\pi, x, r) = E_{(x,r)}^\pi \left[\sum_{t=0}^{\infty} e^{-S_t} c(x_t, a_t) \right], \quad x \in X, \quad \pi \in \Pi, \tag{3}$$

$S_t = \sum_{i=0}^{t-1} r_t$ and $S_0 = 0$. They give conditions for the existence of optimal policies and establish the dynamic programming algorithm. In González-Hernández et al. [13] they use the empirical distributions to prove the existence of asymptotically optimal polices and in González-Hernández et al. [14] they introduce three approximation algorithms. These MCMs correspond to stationary case.

The rest of the paper is organized as follows: In Section 2, the MCM is presented and the dynamical programming method is verified, hence, the existence of non-stationary optimal policies for finite-horizon case is proved; moreover, the measurable selector condition is verified as well. In Section 3, slight conditions for MCMs in order to guarantee the existence of non-stationary optimal polices for the infinite-horizon case are provided. Finally, in Section 4, the consumption-investment problem and the linear-quadratic example are presented, where recursive discount rates are obtained by means of expected value of solution of SDEs associated to scalar Short-Rate Models, which are defined between the periods of the MCM.

2. MARKOV CONTROL MODEL (MCM)

Let us consider the discrete-time Markov Control Model

$$\Big(X', A, \{A(x,r) : (x,r) \in X'\}, r_0, \{\, R_n\}_{n\in\mathbb{N}}, Q, c\Big), \tag{4}$$

where,

1. $X' = X \times (0,\infty)$ is the state space, where X is a Borel space and $(0,\infty)$ is the discount rate space.
2. A is the action space and is a Borel space.
3. $\{A(x,r) : (x,r) \in X'\}$ is a family of non-empty measurable subsets of A. Each $A(x,r)$ is the set of all admissible controls at state $(x,r) \in X'$. The set of all admissible state-action pairs

$$\mathbb{K} := \{(x,r,a) : a \in A(x,r), (x,r) \in X'\} \tag{5}$$

   is supposed to be a measurable set.
4. $r_0 > 0$ is the initial discount rate.
5. $\{R_n\}_{n\in\mathbb{N}}$ is a sequence of measurable functions $R_n : (0,\infty) \to (0,\infty)$, where $r_{n+1} := R_n(r_n)$, $n = 1, 2, \ldots$
6. Q is a stochastic kernel on X given $\mathbb{K}$ and let us define

$$\overline{Q}_n(E \times F \mid x_n, a_n, r_n) := Q(E \mid x_n, a_n, r_n) I_F(R_n(r_n)),$$

   $E \in \mathfrak{B}(X),\ F \in \mathfrak{B}\big((0,\infty)\big)$, $n = 0, 1, \ldots$ where $I_F(\cdot)$ is the indicator function of the set F. The stochastic kernel $\overline{Q}_n$ represents the transition law.
7. $c : X \times A \longrightarrow \mathbb{R}$ is a non-negative measurable function.

Note that the discount rates are given by construction.

More details on MCMs, stochastic kernels and measurable selectors can be found in [4] and [16].

Assumption 1. The set $\mathbb{K}$ contains the graph of a measurable function from X' to A.

Histories and policies

The space of admissible histories up to time n is given by

$$H_n := \mathbb{K}^n \times X', \text{ for } n = 1, 2, \ldots \tag{6}$$

and $H_0 := X'$. A generic element in H_n is of the form

$$h_n := (x_0, r_0, a_0, \ldots, x_{n-1}, r_{n-1}, a_{n-1}, x_n, r_n), \tag{7}$$

where $(x_i, r_i, a_i) \in \mathbb{K}$ for $i = 0, \ldots, n-1$, $(x_n, r_n) \in X'$.

Definition 2.1.

1. A policy $\pi := \{\pi_n\}_{n\in\mathbb{N}}$ is a sequence of stochastic kernels on A given H_n such that for all $h_n \in H_n$,
$$\pi_n(A(x_n, r_n) \mid h_n) = 1,$$
where $h_n = (x_0, r_0, a_0, \ldots, x_{n-1}, r_{n-1}, a_{n-1}, x_n, r_n)$. The set all policies is denoted by Π.

2. A policy π is a **Markov** policy if there exists a sequence of stochastic kernels $\{\phi_n\}_{n\in\mathbb{N}}$ on A given X' such that
$$\pi_n(D \mid h_n) = \phi_n(D \mid x_n, r_n),$$
for all $h_n \in H_n$, $D \in \mathfrak{B}(A)$ and $n \in \mathbb{N}$. The set of all Markov policies is denoted by $\mathbb{M}$.

3. A Markov policy π is a **deterministic** non-stationary policy if there exists a sequence $\{g_n\}_{n\in\mathbb{N}}$ of measurable functions (or **selectors**) $g_n : X' \to A$ such that
$$\phi_n(D \mid x_n, r_n) = I_D[g_n(x_n, r_n)], \quad \forall (x_n, r_n) \in X',\ D \in \mathfrak{B}(A) \text{ and } n \in \mathbb{N}.$$
The set of all deterministic policies is denoted by $\mathbb{D}$.

The relation among these sets is $\mathbb{D} \subset \mathbb{M} \subset \Pi$, and, by Assumption 1, they are non-empty sets.

The canonical construction of the process

Let us consider $(\Omega, \mathfrak{F})$ as the product space where $\Omega := (X' \times A)^\infty$ and $\mathfrak{F}$ is the corresponding product σ-algebra on Ω. The subset $H_\infty := \mathbb{K}^\infty$ is the set of all admissible trajectories.

For a given policy $\pi = \{\pi_n\}$ and $(x_0, r_0) \in X'$ the Ionescu Tulcea Theorem [4, Prop. 7.28] guarantees the existence of a probability measure $P^{\pi}_{(x_0,r_0)}$ on $(\Omega, \mathfrak{F})$ such that

$$P^{\pi}_{(x_0,r_0)}(dx_0 dr_0 da_0 dx_1 dr_1 da_1 \ldots) := \pi_0(da_0 \mid x_0, r_0)\overline{Q}_0(d(x_1, r_1) \mid x_0, r_0, a_0) \cdots$$

This probability measure satisfies for all $B \in \mathfrak{B}(X')$, $C \in \mathfrak{B}(A)$, $h_n \in H_n$, and for $n = 0, 1, \ldots$

1. $P^{\pi}_{(x_0,r_0)}(\mathbb{H}_\infty) = 1$;
2. $P^{\pi}_{(x_0,r_0)}\big((x_0, r_0) \in B\big) = I_B(x_0, r_0)$;
3. $P^{\pi}_{(x_0,r_0)}\big(a_n \in C \mid h_n\big) = \pi_n(C \mid h_n)$;
4. $P^{\pi}_{(x_0,r_0)}\big((x_{n+1}, r_{n+1}) \in B \mid h_n, a_n\big) = \overline{Q}_n(B \mid x_n, r_n, a_n)$.

The stochastic process

$$(\Omega, \mathfrak{F}, P^{\pi}_{(x_0,r_0)}, \{x_n, r_n\}_{n\in\mathbb{N}})$$

is called the discrete-time Markov control process. The expectation operator associated with $P^{\pi}_{(x_0,r_0)}$ is denoted by $E^{\pi}_{(x_0,r_0)}$.

Interpretation. Let $r_0 \in (0, \infty)$ the initial discount rate and x_0 the initial state. Next, an action a_0 with distribution $\pi_0(\cdot \mid x_0, r_0)$ is applied. The process moves to (x_1, r_1), where $r_1 = R_0(r_0)$ and x_1 has the distribution $Q(\cdot \mid x_0, r_0, a_0)$. The process continues in this way.

2.1. Finite-horizon problem

Consider now a MCM as given in (4) operating in N periods and $c_N : X \to \mathbb{R}$ which represents the non-negative terminal cost in the period N.

Definition 2.2. For any $\pi \in \Pi$ and any $(x, r) \in X'$, the measurable function $J : \Pi \times X' \longrightarrow \mathbb{R}$ given by

$$J(\pi, x, r) := E^{\pi}_{(x,r)}\left[\sum_{n=0}^{N-1} e^{-S_n} c(x_n, a_n) + e^{-S_N} c_N(x_N)\right], \tag{8}$$

where $S_0 := 0$ and S_n is defined as

$$S_n := \sum_{i=0}^{n-1} r_i, \quad n = 1, 2, \ldots \tag{9}$$

J is called the expected total discounted cost when the horizon is finite. The expected value is conditioned with respect to $(x, r) = (x_0, r_0)$ and under the policy $\pi \in \Pi$.

The expression

$$J^*(x,r) := \inf_{\Pi} J(\pi, x, r), \quad x \in X, \tag{10}$$

is the value function.

The control problem is to find a policy $\pi^* \in \Pi$ such that

$$J(\pi^*, x, r) = J^*(x,r) \text{ for all } x \in X, \text{ and } r = r_0.$$

The next theorem is known as the Dynamic Programming Theorem. The backward induction method is used in the proof.

Theorem 2.3. Let us define

$$J_N(x, r_N) := c_N(x), \quad x \in X \tag{11}$$

and for $n = 0, 1, \ldots, N-1$,

$$J_n(x, r_n) := \min_{A(x,r_n)} \Big[c(x,a) + e^{-r_n} \int_X J_{n+1}(y, r_{n+1}) Q(\mathrm{d}y \mid x, r_n, a) \Big], \; x \in X. \tag{12}$$

Let us suppose that the functions J_n are measurable for any $n = 0, \ldots, N$, and there exist measurable selectors $f_n \in \mathbb{F}$ such that

$$J_n(x, r_n) = c(x, f_n) + e^{-r_n} \int_X J_{n+1}(y, r_{n+1}) Q(\mathrm{d}y \mid x, r_n, f_n),$$

for $n = 0, 1, \ldots, N-1$.

Then, the policy $\pi^* = \{f_0, f_1 \ldots, f_{N-1}\}$ is optimal and the optimal value J^* coincides with J_0, that is, if $r = r_0$,

$$J^*(x,r) = J_0(x,r) = J(\pi^*, x, r), \quad \forall x \in X. \tag{13}$$

P r o o f. It is similar to the proof of Theorem 3.2.1 of [16] with obvious changes. □

The measurable selector condition

The existence of a sequence of measurable selectors in the previous Theorem is supposed. Now we give conditions on MCM (4) in order to guarantee the existence of such selectors.

Definition 2.4. A function $\nu : \mathbb{K} \to \mathbb{R}$ is called *inf-compact* on $\mathbb{K}$, if for each $(x,r) \in X'$ and $z \in \mathbb{R}$, the set

$$\{a \in A(x,r) : \nu(x,r,a) \leq z\}$$

is compact.

Condition 2.5. *Measurable Selector Condition.*

For each measurable function $u : X' \to \mathbb{R}$, the function u^* of X' to $\mathbb{R}$

$$u^*(x, r_n) = \inf_{a \in A(x,r_n)} \Big[c(x,a) + e^{-r_n} \int_X u(y, r_{n+1}) Q(\mathrm{d}y \mid x, r_n, a) \Big], \tag{14}$$

is measurable for each n, and there exists a measurable selector $f_n : X' \to A$ such that the right side in Equation (14) attains the minimum at $f_n(x, r_n) \in A(x, r_n)$ for all $x \in X,\ r_n \in (0, \infty)$ and $n = 1, 2, \ldots$, that is,

$$u^*(x, r_n) := c(x, f_n(x, r_n)) + e^{-r_n} \int_X u(y, r_{n+1}) Q(\mathrm{d}y \mid x, r_n, f_n(x, r_n)), \quad n = 1, 2, \ldots \tag{15}$$

Condition 2.6.

1. The cost functions c and c_N are l.s.c., and inf-compact on $\mathbb{K}$ for all $(x, r) \in X'$.
2. The transition kernel Q is strongly continuous, that is, the function

$$\bar{u}(x, r_n, a) := \int_X u(y, r_{n+1}) Q(\mathrm{d}y \mid x, r_n, a) \tag{16}$$

   is continuous and bounded on $\mathbb{K}$, for all bounded measurable function u on X', $n = 0, 1, 2, \ldots$

Theorem 2.7. Let us suppose that an MCM (4) satisfies the Conditions 2.6 for any measurable function u on X'. Then, the Condition 2.5 is valid.

Proof. It is similar to Theorem 3.3.5 of [16]. □

3. THE INFINITE-HORIZON PROBLEM

In this case, the MCM of (4), has the performance index as

$$V(\pi, x, r) := E^{\pi}_{(x,r)} \left[\sum_{n=0}^{\infty} e^{-S_n} c(x_n, a_n) \right], \tag{17}$$

where $\pi \in \Pi,\ x \in X,\ r = r_0 \in (0, \infty)$ and S_n are defined in Equation (9).

For $r = r_0$, the value function for the infinite-horizon is defined by

$$V^*(x, r) = \inf_{\Pi} V(\pi, x, r) \quad \forall x \in X. \tag{18}$$

The optimal control problem with infinite-horizon consists in finding $\pi^* \in \Pi$ such that

$$V(\pi^*, x, r) = V^*(x, r) \quad \forall x \in X, \text{ and } r = r_0.$$

The existence and measurability of value function V^* is neither a consequence of fixed-point theorems nor an interchange of limits and minima in a particular discounted cost optimality equation, as occurs for stationary discount case. The existence and measurability of V^* is guaranteed by approximating with a sequence of finite horizon control problems. Then, we shall suppose that all conditions for the finite-horizon case hold and we need to introduce some additional notation.

Definition 3.1. Let m and n be non-negative numbers, such that $m \leq n$. The expected total cost from time m up to time n conditioned by initial pair (x_0, r_0) is defined by

$$V_m^n(\pi, x, r) = E_{(x_0,r_0)}^{\pi} \left[\sum_{t=m}^{n} e^{-(S_t - S_m)} c(x_t, a_t) \mid x_m = x, r_m \right], \tag{19}$$

for any $\pi \in \Pi$ and $x \in X$. The value function from time m up to time n is

$$V_m^{n*}(x, r) := \inf_{\Pi} \; V_m^n(\pi, x, r), \quad \forall x \in X, \; r = r_m. \tag{20}$$

The control problem from time m up to time n is to find a policy $\pi^{(m,n)*} \in \Pi$ such that

$$V_m^{n*}(\pi^{(m,n)*}, x, r) = V_m^{n*}(x, r), \quad \forall x \in X, \; r = r_m. \tag{21}$$

If $m = 0$, $V_0^n(\pi, x, r) := V^n(\pi, x, r)$. If $n \to \infty$, $V_m^n(\pi, x, r) := V_m(\pi, x, r)$ for m fixed and $\pi \in \Pi$.

Condition 3.2.

1. The cost function c is l.s.c. and inf-compact on $\mathbb{K}$ for all $(x, r) \in X'$.
2. The transition kernel Q is strongly continuous.
3. There exists a policy π in Π such that $\lim_{k\to\infty} V_k(\pi, x, r_k) = 0$ for any $x \in X$. The set of policies that satisfy this condition is denoted by Π^1.

Definition 3.3. Given $\epsilon > 0$. A policy $\pi \in \Pi^1$ is $\epsilon-$optimal if

$$V(\pi, x, r) - V^*(x, r) < \epsilon, \; \forall x \in X, \; r = r_0.$$

Lemma 3.4. For any $x \in X$, $r = r_k$ and $k = 0, 1, 2, \ldots,$

$$V_k^{n*}(x, r_k) \to V_k^*(x, r_k) \tag{22}$$

as $n \to \infty$.

P r o o f. Let us define for each $k = 0, 1, 2, \ldots$ the measurable functions on X'

$$u^0(x, r_k) := \inf_{a \in A(x, r_k)} \Big[c(x, a) \Big], \tag{23}$$

and

$$u^n(x,r_k) := \inf_{a\in A(x,r_k)} \Big[c(x,a) + e^{-r_k} \int_X u^{(n-1)}(y,r_{k+1})Q(\mathrm{d}y \mid x,r_k,a)\Big],\ n = 1,2,\ldots \tag{24}$$

We shall show that for all n, $u^n(x,r_k)$ generates the optimal value of the next n steps when the initial step is k, i. e., $u^n(x,r_k) = V_k^{(k+n)*}(x,r_k)$ for all $k = 0,1,2,\ldots,\ x \in X$. It is proved by induction over n.
For $n = 0$, note that

$$u^0(x,r_k) \le V_k^k(\pi,x,r_k), \quad \forall\pi \in \Pi,\ x \in X,\ k = 0,1,2,\ldots$$

Taking the infimum over Π,

$$u^0(x,r_k) \le V_k^{k*}(x,r_k),\ k = 0,1,2,\ldots \tag{25}$$

On the other hand, by Measurable Selector Condition, there exists for each $k = 0,1,2,\ldots,$ the selector $f_k^{(0)}$ such that

$$\begin{aligned} u^0(x,r_k) &= c(x,f_k^{(0)}(x,r_k)) \\ &= V_k^k(\pi^{(k,k)},x,r_k) \\ &\ge V_k^{k*}(x,r_k), \end{aligned} \tag{26}$$

where $\pi^{(k,k)}$ represents the policy $\pi^{(k,k)} = \{\pi_0,\ldots,\pi_{k-1},f_k^{(0)},\pi_{k+1},\ldots\}$, and $\pi = \{\pi_0,\ldots,\pi_k,\ldots\} \in \Pi^1$. Then, by (25) and (26), $V_k^{k*} = u^0$ for all $x \in X$ and any $k = 0,1,2,\ldots,$ is concluded.

Now, let us assume that for each $k = 0,1,2,\ldots,$ the induction hypothesis holds for n, i. e.,

$$u^n(x,r_k) = V_k^{(k+n)*}(x,r_k),$$

which implies that there exists a measurable selector $f_k^{(n)}$ such that

$$u^n(x,r_k) = c(x,f_k^{(n)}) + e^{-r_k} \int_X u^{(n-1)}(y,r_{k+1})Q(\mathrm{d}y \mid x,r_k,f_k^{(n)}), \tag{27}$$

and there exists a policy $\pi^{(k,k+n)} = \{\pi_0,\ldots,\pi_{k-1},f_k^{(n)},f_{k+1}^{(n-1)},\ldots,f_{k+n}^{(0)},\pi_{k+n+1},\ldots\}$, that satisfies

$$V_k^n(\pi^{(k,k+n)},x,r_k) = V_k^{n*}(x,r_k).$$

For $n+1$, by induction hyphotesis, note that for any $k = 0,1,2,\ldots$

$$u^{(n+1)}(x,r_k) \le V_k^{(n+1)}(\pi,x,r_k), \quad \forall\pi \in \Pi,\ x \in X,$$

and taking the infimum over Π, is obtained that

$$u^{n+1}(x,r_k) \le V_k^{(n+1)*}(x,r_k). \tag{28}$$

As $u^n(x, r_k)$ is the optimal value in n steps for any initial step k, in particular is valid for the initial step $k+1$, that is, $u^n(x, r_{k+1}) = V_{k+1}^{(k+1+n)*}(x, r_{k+1})$. Again, by Measurable Selector Condition there exists for each $k = 0, 1, 2, \dots$, a measurable selector $f_k^{(n+1)}$ such that

$$\begin{aligned}
u^{n+1}(x, r_k) &= c(x, f_k^{(n+1)}) + e^{-r_k} \int\limits_X u^n(y, r_{k+1}) Q(\mathrm{d}y \mid x, r_k, f_k^{(n+1)}), \\
&= c(x, f_k^{(n+1)}) + e^{-r_k} \int\limits_X V_{k+1}^{(k+1+n)*}(y, r_{k+1}) Q(\mathrm{d}y \mid x, r_k, f_k^{(n+1)}), \\
&= V_k^{(k+n+1)}(\pi^{(k,n+1)}, x, r_k), \\
&\geq V_k^{(n+1)*}(x, r_k),
\end{aligned}$$

where $\pi^{(k,n+1)}$ is the policy

$$\pi^{(k,k+n+1)} = \{\pi_0, \dots, \pi_{k-1}, f_k^{(n+1)}, f_{k+1}^{(n)}, \dots, f_{k+n}^{(1)}, f_{k+n+1}^{(0)}, \pi_{k+n+2}, \dots\}.$$

Hence, it is concluded that $V_k^{(k+n+1)*}(x, r_k) = u^{(n+1)}(x, r_k)$ for any $x \in X$, $k = 0, 1, 2, \dots$ and $n = 0, 1, 2, \dots$

Moreover, by construction of u^n

$$0 \leq u^n(x, r_k) = V_k^{n*}(x, r_k) \leq u^{(n+1)}(x, r_k) \leq V_k^*(x, r_k), \quad \text{for } k, n = 0, 1, 2, \dots$$

then, there exists the measurable function $\overline{U}_k$ over X' such that $\overline{U}_k(x, r_k) \leq V_k^*(x, r_k)$, $k = 0, 1, 2, \dots$ and $u^n \uparrow \overline{U}_k$. Hence, as the functions V_m^* tends to 0, if $m \to \infty$, therefore, by the part (3) in the Condition 3.2, there exist measurable selectors $f_k, f_{k+1}, \dots$ such that, the policy $\pi^{(k,\infty)} = \{\pi_0, \dots, \pi_{k-1}, f_k, f_{k+1}, \dots\} \in \Pi^1$ satisfies the inequality

$$\overline{U}_k(x, r_k) \geq V_k(\pi_k^{(k,\infty)}, x, r_k)$$

for all $x \in X$, $k = 0, 1, 2, \dots$

Moreover by definition of V_k^* is obtained that

$$V_k^*(x, r_k) \leq V_k(\pi_k^{(k,\infty)}, x, r_k) \leq \overline{U}_k(x, r_k),$$

for all $k = 0, 1, 2, \dots$ and all $x \in X$. Hence, $\overline{U}_k(x, r_k) = V_k^*(x, r_k)$, for all $k = 0, 1, 2, \dots$ and $x \in X$. □

Note that, by former lemma, the solution of the infinite-horizon optimal Markov control problem is obtained if $k = 0$.

Theorem 3.5. Suppose valid the conditions 3.2, then, there exists a deterministic non-stationary policy $\pi \in \Pi^1$, such that,

$$V(\pi, x, r) = V^*(x, r), \quad \text{for all } (x, r) \in X.$$

P r o o f. For any natural n, and $(x,r) \in X'$, by lemma 3.4 the functions V_n^* are measurable.

We shall prove that

$$V_n^*(x,r_n) = \inf_{a\in A(x,r_n)} \Big[c(x,a) + e^{-r_n} \int\limits_X V_{n+1}^*(y,r_{n+1}) Q(\mathrm{d}y \mid x,r_n,a)\Big], \tag{29}$$

for all $n = 0,1,2,\dots$ Let us define

$$\nu(x,r_n) := \inf_{a\in A(x,r_n)} \Big[c(x,a) + e^{-r_n} \int\limits_X V_{n+1}^*(y,r_{n+1}) Q(\mathrm{d}y \mid x,r_n,a)\Big],\ n = 0,1,2,\dots$$

For any policy $\pi \in \Pi$ is valid that

$$V_n(\pi,x,r_n) \geq V_n^*(x,r_n), \quad V_n(\pi,x,r_n) \geq \nu(x,r_n),\ \forall n,$$

therefore

$$\begin{aligned} V_n(\pi,x,r_n) &= E^\pi_{(x_0,r_0)} \left[c(x,a) + e^{-r_n} \int\limits_X V_{n+1}(y,r_{n+1}) Q(\mathrm{d}y \mid x,r_n,a)\right] \\ &\geq E^\pi_{(x,r_n)} \left[c(x,a) + e^{-r_n} \int\limits_X V_{n+1}^*(y,r_{n+1}) Q(\mathrm{d}y \mid x,r_n,a)\right] \\ &\geq \inf_{a\in A(x,r_n)} \Big[c(x,a) + e^{-r_n} \int\limits_X V_{n+1}^*(y,r_{n+1}) Q(\mathrm{d}y \mid x,r_n,a)\Big], \end{aligned}$$

and taking the infimum over Π,

$$V_n^*(x,r_n) \geq \nu(x,r_n). \tag{30}$$

Let $\bar\pi \in \Pi^1$. Note that

$$V_n^*(x,r_n) \leq V_n(\bar\pi,x,r_n) \downarrow 0, \quad \text{when } n \to \infty.$$

On the other hand, by Measurable Selector Condition, there exists a measurable selector f_k such that

$$\nu(x,r_k) = c(x,f_k) + e^{-r_k} \int\limits_X V_{k+1}^*(y,r_{k+1}) Q(\mathrm{d}y \mid x,r_k,f_k),\ \forall k \geq n.$$

Consider now the policy $\bar\pi^{(n,\infty)} = \{\bar\pi_0,\dots,\bar\pi_{n-1},f_n,f_{n+1},\dots\}$. By last lemma, for any $\epsilon > 0$, there exists m_0 such that, if $m > m_0$

$$\begin{aligned} V_n(\bar\pi^{(n,\infty)},x,r_n) &< V_n^m(\bar\pi^{(n,\infty)},x,r_n) + \epsilon \\ &= V_n^{m*}(x,r_n) + \epsilon. \end{aligned}$$

for all n. Then, if $m \to \infty$, $V_n(\overline{\pi}^{(n,\infty)}, x, r_n) \leq V_n^*(x, r_n)$ for all n. By definition of V_n^*, $V_n(\overline{\pi}^{(n,\infty)}, x, r_n) \geq V_n^*(x, r_n)$, therefore

$$V_n(\overline{\pi}^{(n,\infty)}, x, r_n) = V_n^*(x, r_n), \text{ for all } n.$$

Moreover

$$\begin{aligned}\nu(x, r_n) &= c(x, f_n) + e^{-r_n} \int_X V_{n+1}^*(x_{n+1}, r_{n+1}) Q(\mathrm{d}x_{n+1} \mid x, r_n, f_n) \\ &= c(x, f_n) + e^{-r_n} \int_X V_{n+1}(\overline{\pi}^{(n,\infty)}, x_{n+1}, r_{n+1}) Q(\mathrm{d}x_{n+1} \mid x, r_n, f_n) \\ &\geq V_n^*(x, r_n),\end{aligned}$$

and by inequality (30), the equality (29) holds. Hence, the deterministic non-stationary policy

$$\overline{\pi}^{(0,\infty)} := \{f_0, f_1, f_2, \dots\}$$

satisfies for $n = 0$ in (29) that

$$V^*(x, r) = V_0^*(x, r_0) = c(x, f_0) + e^{-r_0} \int V_1^*(x_1, r_1) Q(\mathrm{d}x_1 \mid x, r_0, f_0) = V(\overline{\pi}^{(0,\infty)}, x, r_0).$$

□

Theorem 3.6. Suppose valid the conditions 3.2. If there exists a policy $\overline{\pi}$ such that

$$V_k(\overline{\pi}, x, r_k) = \inf_{a \in A(x, r_k)} \Big[c(x, a) + e^{-r_k} \int_X V_{k+1}(\overline{\pi}, y, r_{k+1}) Q(\mathrm{d}y \mid x, r_k, a) \Big], \tag{31}$$

for all $x \in X$, $k = 0, 1, 2, \dots$, and satisfies

$$\lim_{k \to \infty} e^{-S_k} E_{(x,r)}^{\pi} V_k(\overline{\pi}, x_k, r_k) = 0, \quad \forall \pi \in \Pi^1 \text{ and } (x, r) \in X', \tag{32}$$

then, $V(\overline{\pi}, x, r) = V^*(x, r)$ for all $(x, r) \in X'$.

P r o o f. If Equation (31) holds, by definition of V^*, $V(\overline{\pi}, x, r) \geq V^*(x, r)$ for all $(x, r) \in X'$. For reverse inequality, from any $\pi \in \Pi$, $(x, r) \in X'$ and by Markov Property,

$$\begin{aligned}&E_{(x,r)}^{\pi}[e^{-S_{k+1}} V_{k+1}(\overline{\pi}, x_{k+1}, r_{k+1}) \mid h_k, a_k] \\ &= e^{-S_{k+1}} \int_X V_{k+1}(\overline{\pi}, y, r_{k+1}) Q(\mathrm{d}y \mid x_k, r_k, a_k) \\ &= e^{-S_k} \left[c(x_k, a_k) + e^{-r_k} \int_X V_{k+1}(\overline{\pi}, y, r_{k+1}) Q(\mathrm{d}y \mid x_k, r_k, a_k) - c(x_k, a_k) \right] \\ &\geq e^{-S_k} \left[V_k(\overline{\pi}, x_k, r_k) - c(x_k, a_k) \right],\end{aligned}$$

therefore,

$$e^{-S_k}c(x_k,a_k) \geq -E^{\pi}_{(x,r)}\left[e^{-S_{k+1}}V_{k+1}(\overline{\pi},x_{k+1},r_{k+1}) - e^{-S_k}V_k(\overline{\pi},x_k,r_k) \mid h_k,a_k)\right].$$

Thus, applying expectations $E^{\pi}_{(x,r)}$ and adding over $i=0,1,\ldots,n-1$,

$$E^{\pi}_{(x,r)}\sum_{i=0}^{n-1}e^{-S_i}c(x_i,a_i) = V_0(\overline{\pi},x,r) - e^{-S_k}E^{\pi}_{(x,r)}V_n(\overline{\pi},x_n,r_n),$$

for all n. Taking $n\to\infty$ and using the Equation (32), it follows that $V(\pi,x,r)\geq V(\overline{\pi},x,r)$ for all $(x,r)\in X'$, and $V^*(x,r)\geq V(\overline{\pi},x,r)$ holds $(x,r)\in X'$. In consequence, $V^*(x,r)=V(\overline{\pi},x,r)$ holds for $(x,r)\in X'$. □

Proposition 3.7. For the affirmations:

(i) The cost function c is bounded, i. e., there exists constant m, such that $0\leq c(x,a)\leq m$ for all $(x,r,a)\in\mathbb{K}$.

(ii) Suppose that the discount rate space is reduced to $[d_1,z)$ for some z positive number, $1\leq z\leq e^r$, there exist $m>0$ and a non-negative measurable function w on X' such that

$$c(x,a)\leq mw(x,r_k), \quad \text{and} \quad \int w(y,r_{k+1})Q(\mathrm{d}y\mid x,r_k,a)\leq zw(x,r_k),$$

for $(x,r_k,a)\in\mathbb{K},\ k=0,1,2,\ldots$

(iii) $C(x,r_j):=\sum_{k=0}^{\infty}e^{-(S_{j+k}-S_j)}c_k(x,r_{k+j})<\infty$, for all $(x,r_j)\in X'$, where $c_0(x,r_k):=\sup_{A(x,r_k)}c(x,a)$, $k=0,1,2,\ldots$ and

$$c_k(x,r_j)=\sup_{A(x,r_j)}\int_X c_{k-1}(y,r_{j+1})Q(\mathrm{d}y\mid x,r_j,a), \quad k>1,\ j=0,1,2,\ldots$$

(iv) $\lim_{n\to\infty}e^{-S_n}E^{\pi}_{(x,r)}V_n(\overline{\pi},x_n,r_n)=0$, for all $\pi,\overline{\pi}\in\Pi^1$ and $(x,r)\in X'$.

Are valid the relations:

$$(i)\Longrightarrow(ii)\Longrightarrow(iii)\Longrightarrow(iv)\Longrightarrow(32), \tag{33}$$

and

$$(i)\Longrightarrow(ii)\Longrightarrow(iii)\Longrightarrow\Pi^1=\Pi. \tag{34}$$

Moreover, if any of the conditions (i) to (iv) holds, then a policy π^* is optimal if and only if $V(\pi^*,x,r)$ satisfies the Equation (31).

P r o o f. To prove $(i) \Longrightarrow (ii)$, use $w \equiv 1$ for example.
(ii) implies (iii). By induction on k $c_k(x, r_k) \leq mz^k w(x, r_k)$ is obtained for all $(x, r_k) \in X'$ and $k = 0, 1, 2, \ldots$ Therefore,

$$C(x, r_j) \leq mw(x, r_j)/(1 - ze^{-d_1}) < \infty,$$

for each $(x, r_j) \in X'$.

(iii) implies (iv). Suppose that (iii) holds. First, we shall prove the following inequalities

$$V_n(\pi, x, r_n) \leq C(x, r_n), \ \forall (x, r_n) \in X', \pi \in \Pi, \tag{35}$$

$n = 0, 1, 2, \ldots$ and

$$E^{\pi}_{(x,r)} C(x_n, r_n) \leq \sum_{k=n}^{\infty} e^{-(S_k - S_n)} c_k(x, r) \ \forall (x, r) \in X', \pi \in \Pi, \ n = 0, 1, \ldots \tag{36}$$

To prove (35), by the Markov property

$$E^{\pi}_{(x,r)}[c_0(x_{k+1}, r_{k+1}) \mid h_k, a_k] = \int c_0(y, r_{k+1}) Q(\mathrm{d}y \mid x_k, r_k, a_k) \leq c_1(x_k, r_k),$$

for all $k \geq 0$, then $E^{\pi}_{(x,r)} c_0(x_{k+1}, r_{k+1}) \leq E^{\pi}_{(x,r)} c_1(x_k, r_k)$ holds. Iterating the last argument, the inequalities

$$E^{\pi}_{(x,r)} c_0(x_k, r_k) \leq E^{\pi}_{(x,r)} c_1(x_{k-1}, r_{k-1}) \leq \cdots \leq E^{\pi}_{(x,r)} c_k(x_0, r_0) = c_k(x, r), \tag{37}$$

holds. Since $c(x_k, a_k) \leq c_0(x_k, r_k)$, then

$$E^{\pi}_{(x,r)} c(x_k, a_k) \leq E^{\pi}_{(x,r)} c_0(x_k, r_k) \leq c_k(x, r) \text{ for all } k,$$

and if for each $k = 0, 1, 2, \ldots$, the inequality $E^{\pi}_{(x,r)} c(x_k, a_k) \leq c_k(x, r)$ is multiplied by e^{-S_k} and these inequalities are summed beginning from $k = n$, $V_n(\pi, x, r_n)$ and $C(x, r_n)$ are obtained, implying the validity of (35).

To prove (36), if $n = 0$, this equation follows from the definition of C. Now for $n \geq 1$, using the Markov property

$$\begin{aligned} E^{\pi}_{(x,r)}[C(x_n, r_n) \mid h_{n-1}, a_{n-1}] &= \int C(y, r_n) Q(\mathrm{d}y \mid x_{n-1}, r_{n-1}, a_{n-1}) \\ &= \sum_{k=0}^{\infty} e^{-S_k} \int c_k(y, r_n) Q(\mathrm{d}y \mid x_{n-1}, r_{n-1}, a_{n-1}) \\ &\leq \sum_{k=0}^{\infty} e^{-S_k} c_k(x_{n-1}, r_{n-1}), \end{aligned}$$

and applying $E^{\pi}_{(x,r)}$, the inequality $E^{\pi}_{(x,r)} C(x_n, r_n) \leq \sum\limits_{k=0}^{\infty} e^{-S_k} c_{k+1}(x_{n-1}, r_{n-1})$ holds.

Now, similar to (37), we can see

$$E^{\pi}_{(x,r)}c_{k+1}(x_{n-1},r_{n-1}) \leq E^{\pi}_{(x,r)}c_{k+2}(x_{n-2},r_{n-2}) \leq \cdots \leq E^{\pi}_{(x,r)}c_{k+n}(x_0,r_0) = c_{k+n}(x,r),$$

therefore $E^{\pi}_{(x,r)}C(x_n,r_n) \leq \sum_{k=0}^{\infty} e^{-S_k}c_{k+n}(x,r)$, and (36) is concluded.

To prove (iv), consider π and $\overline{\pi}$ arbitrary policies. By (35) $V(\overline{\pi},x_n,r_n) \leq C(x_n,r_n)$ for all n, and using (36)

$$E^{\pi}_{(x,r)}V(\overline{\pi},x_n,r_n) \leq E^{\pi}_{(x,r)}C(x_n,r_n) \leq \sum_{k=n}^{\infty} e^{-(S_k-S_n)}c_k(x,r).$$

Finally, as $C(x,r) < \infty$ and if $n \to \infty$,

$$e^{-S_n}E^{\pi}_{(x,r)}V(\overline{\pi},x_n,r_n) \leq \sum_{k=n}^{\infty} e^{-S_k}c_k(x,r) \to 0,$$

then, it follows (iv).

Former paragraph shows that (iv) implies (32). The affirmation (34) follows of (33) and the inequality $V \leq C$. Finally, $\Pi^1 = \Pi$ follows of (33) and Theorem 3.6. □

Proposition 3.8. There can be found deterministic non-stationary $\epsilon-$optimal policies (Definition 3.3).

Proof. Let $\pi \in \Pi^1$ be an initial policy, $\epsilon > 0$ and $x \in X$. For any $(x,r) \in X'$, consider the measurable functions

$$u^0(x,r_0) = \inf_{a\in A(x,r_0)} c(x,a) \tag{38}$$

and

$$u^n(x,r) = \inf_{a\in A(x,r)}\left[c(x,a) + e^{-r}\int_X u^{n-1}(y,r_1)Q(\mathrm{d}y \mid x,r,a)\right], \tag{39}$$

for $n = 1,2,\ldots$ Using these functions, by Condition (3.2) and Lemma 3.4, the sequence of measurable functions $\overline{a}_0,\overline{a}_1,\ldots,\overline{a}_n,\ldots$, is obtained. Additionally, if we suppose that

$$u^n(x,r) - u^{n-1}(x,r_1) < \frac{\epsilon}{2^n},\ (x,r) \in X',$$

then

$$u^n(x,r) - V^*(x,r_{n-1}) < \epsilon\Big[\frac{1-(1/2)^{n+1}}{1-1/2}\Big],$$

and if n is great enough, the policy

$$\overline{\pi} = \{\overline{a}_0,\overline{a}_1,\ldots,\overline{a}_n,\pi_{n+1},\pi_{n+2},\ldots\}$$

is a deterministic non-stationary $\epsilon-$optimal. □

Remark 3.9.

1. By former lemma, as $V^{n*}(x,r)$ is measurable for each n, $(x,r) \in X'$, and the limit $V^*(x,r)$ is measurable for $(x,r) \in X'$. It is also measurable for all $m \in \mathbb{N}$ and $x \in X$, $V^n_m(x,r_m)$, where r_m is obtained using the recursive functions $R_m, R_{m-1}, \ldots, R_0$ from any initial discount rate $r \in (0,\infty)$.

2. Observe that for $\pi \in \Pi^1$ and any natural number n

$$\lim_{k\to\infty}\left[\exp\left\{-\sum_{i=n}^{k} r_i\right\}V^{n+k}(\pi, x_{n+k}, r_{n+k})\right] = 0. \tag{40}$$

3. The most common models for discount rate used in finance, economics and administration are generally considered as stochastic processes whose dynamics follow a stochastic differential equation, as the Vasicek short-rate Model [26], for example. A case of recursive discount rate relation $\{R_n\}_{n\in\mathbb{N}}$ can be obtained when in each period the applied discount rate is the expected value r_n of a solution of

$$\mathrm{d}z_n = \kappa_n[\theta_n - z_n]\,\mathrm{d}t + \sigma_n\,\mathrm{d}W_n, \quad t \in [n-1, n] \tag{41}$$

under the initial condition r_{n-1}, where for all n there exist $\epsilon > 0$ and $\gamma > 0$ such that $\theta_n \in [\theta_{\min}, \theta_{\max}]$, $\kappa_n \in [\kappa_{\min}, \kappa_{\max}]$, and

$$\epsilon < \theta_{\min}, \quad \theta_{\max} < 1-\epsilon, \quad \gamma < \kappa_{\min} \quad \kappa_{\max} < 1-\gamma. \tag{42}$$

By classical Itô conditions (see [1]), the solution for each SDE, $n = 1, 2\ldots,$ is

$$z_t = r_{n-1}e^{-(t-(n-1))\kappa_n} + \theta_n[1 - e^{-(t-(n-1))\kappa_n}] + \sigma_n e^{-t\kappa_n}\int_{n-1}^{t} e^{u\kappa_n}\mathrm{d}W_u. \tag{43}$$

Its expected value is

$$r_t := r_{n-1}e^{-(t-(n-1))\kappa_n} + \theta_n[1 - e^{-(t-(n-1))\kappa_n}] \tag{44}$$

and evaluated in $t = n$, it generates the recursive discount rate relation

$$R_n(r_{n-1}, \theta_n, \kappa_n) = Er_n = r_{n-1}e^{-\kappa_n} + \theta_n[1 - e^{-\kappa_n}]. \tag{45}$$

Using the restrictions for κ_n and θ_n given by the equations (42) it is easy to see that the sums S_n in (9) diverge. Moreover, when $n \to \infty$

$$\lim_{n\to\infty} e^{-S_n} = 0. \tag{46}$$

This limit guarantees that for $\pi \in \Pi^1$,

$$\lim_{n\to\infty} e^{-S_n} E^{\pi}_{(x,r_0)}[V^{(n+1)*}(x_{n+1}, r_{n+1})] = 0.$$

Note that the drift in the SDE defined in the equation (41), satisfies the *reversion property* to θ_n, with $t \in [n-1, n]$, $n = 1, 2, \ldots$ This property is present in many short-rate models as [6, 9, 21, 26], for example. In a short-rate model, an SDE has a mean reverting property if the process tends to drift towards its long-term mean. This means that if any time $t \in [n-1, n]$, $n = 1, 2, \ldots$, the rate satisfies $r_n > \theta_n$, then the drift of the SDE is negative and causes that the rate decrease. Similarly, if the rate $r_n < \theta_n$, the drift causes the rate to increase. The expression θ_n is the mean value or tendency of the rate, when $t \in [n-1, n]$, $n = 1, 2, \ldots$ This tendency of mean reversion in the SDE is real because of regulation policies in financial markets related with credits and investments, where the interest rate revolves around the ideal value $\theta_n > 0$.

Other short rate models used in finance are

$$\mathrm{d}z_t = \mu z_t \,\mathrm{d}t + \sigma z_t \,\mathrm{d}W_t, \quad t \in [n-1, n], \tag{47}$$

with μ real constant, σ non-negative constant and initial condition r_{n-1}. It corresponds to Dothan and Rendleman Bartter models (see Dothan [10], Brigo and Mercurio [5] and Rendleman and Bartter [23]). The expected value of the solution evaluated in $t = n$ defines the recursive relation

$$R_n(r_{n-1}) = r_n = r_{n-1} e^{\mu} = r_0 e^{n\mu}, \ n = 1, 2, \ldots, \tag{48}$$

and it is clear that for $\mu > 0$,

$$\lim_{n\to\infty} r_0 e^{n\mu} = \infty.$$

See Equation (46). Hence, we can use these discount rates to find $\epsilon-$optimal non-stationary policies.

Similar results can be obtained with Ho-Lee and Hull-White models (see [19] and [21]). They consider the SDE

$$\mathrm{d}z_t = \theta(t)\,\mathrm{d}t + \sigma \mathrm{d}W_t, \tag{49}$$

where θ is a function of t, σ is constant and $t \in [n-1, n]$, $n = 1, 2, \ldots$ The expected value of the solution evaluated at $t = n$ generates the recursive relation

$$R(r_{n-1}) = r_n = r_{n-1} + \int_{n-1}^{n} \theta(u)\,\mathrm{d}u = r_0 + \int_0^n \theta(u)\,\mathrm{d}u, \ n = 1, 2, \ldots \tag{50}$$

and, if we suppose that $\theta(\cdot)$ is such that

$$\lim_{n\to\infty} r_0 + \int_0^n \theta(u)\,\mathrm{d}u = r_0 + \int_0^\infty \theta(u)\,\mathrm{d}u = \infty,$$

then $\lim\limits_{n\to\infty} S_n = \infty$.

4. EXAMPLES

Consumption-investment model

Let us suppose the consumption-investment problem (see [16], section 3.6). The variable x_n represents the capital accumulated for investment at time n, a_n represents the quantity of capital invested at time n, and $x_n - a_n$ the capital used for consumption at time n, with $n = 0, \ldots, N$.

Then, $X = [0, \infty)$, and $A(x, r) = [0, x]$. Additionally, let us suppose for any $n > 0$, the capital grows according to

$$x_{n+1} = a_n + \xi_n, \quad n = 0, 1, \ldots,$$

where the random variables ξ_n are i.i.d. and independent from x_0. Let $\overline{m} := E\xi_0 < \infty$. The problem is to find the optimal quantity that can be invested for maximizing the utility of consumption u, defined by

$$u(x - a) = b(x - a),$$

where b is a non-negative constant. The terminal consumption is the quantity $d > 0$ and is considered fixed. The dynamic of the discount rate follows the SDE in Equation (45) and satisfies the restrictions given in the Equation (42), $n = 1, 2, \ldots$ Therefore for the initial discount rate $r := r_0 \in [d_1, d_2]$, $r_i = R_{i-1}(r_{i-1}, \kappa_i, \theta_i)$, $i = 1, \ldots, N-1$

$$J(\pi, x, r) = E^{\pi}_{(x,r)} \left[\sum_{t=0}^{N-1} e^{-S_n} u(x_n - a_n) \right], \ r = r_0.$$

where S_n is defined in Equation (9).

The optimality Equation for this problem is

$$J_N(x, r) = u(x_N) = e^{-r_N} d, \quad \forall (x, r) \in X \times (0, \infty).$$

$$J_n(x, r) = \max_{A(x,r)} \left[b(x - a) + e^{-r_n} E J_{n+1}(a + \xi_n, r) \right],$$

$n = N-1, \ldots, 0$, and $r_n = R_n(r_{n-1}, \kappa_n, \theta_n)$ can be represented in terms of r_0 as in the Equation (45). For $t = N-1$, we have

$$\begin{aligned} J_{N-1}(x, r) &= \max_{A(x, r_{N-1})} \left[b(x - a) + e^{-r_{N-1}} E J_N(a + \xi_{N-1}, r) \right] \\ &= bx + de^{-(r_{N-1} + r_N)}, \end{aligned}$$

with $a = 0 = f_{N-1}(x, r)$. If $t = N-2$,

$$\begin{aligned} J_{N-2}(x, r) &= \max_{A(x, r_{N-2})} \left[b(x - a) + e^{-r_{N-2}} E J_{N-1}(a + \xi_0, r) \right] \\ &= \max_{A(x, r_{N-2})} \left[b(x - a) + e^{-r_{N-2}} [b(a + \overline{m}) + de^{-(r_{N-1} + r_N)}] \right] \\ &= \max_{A(x, r_{N-2})} \left[ab(e^{-r_{N-2}} - 1) + b(x + \overline{m} e^{-r_{N-2}}) + de^{-r_{N-1}} \right] \\ &= b(x + \overline{m} e^{-r_{N-2}}) + de^{-r_{N-1}} \end{aligned}$$

for $a = 0 = f_{N-2}(x,r)$. In a similar way, the policy is $f_0 = f_1 = \cdots = f_{N-1} = 0$, for the next periods and we have

$$J_0(x,r) = b\left(x + \overline{m}\sum_{i=1}^{N-1} e^{-S_i}\right) + de^{-S_N}.$$

The linear-quadratic model

(see [16], section 3.7). Let us suppose an MCM with $X = A = \mathbb{R}$, where the evolution of states is given by the linear equation

$$x_{n+1} = bx_n + ca_n + \xi_n, \quad n = 0, 1, 2, \dots,$$

b and c are real constants. The random variables ξ_n are supposed i.i.d. and independent from x_0, with

$$E\xi_0 = 0, \quad 0 < E\xi_0^2 = \sigma^2 < \infty.$$

The cost function is given by

$$c(x,a) = dx^2 + ga^2, \quad d \geq 0, \quad g > 0.$$

The finite-horizon problem in this case consists of finding the discounted optimal policy until period time N when the dynamic of the discount rate r_n is given by Equation (41), the expected solution is given by Equation (44), κ_n and θ_n are positive constants such that the restrictions in Equation (42) $n = 1, 2, \dots,$ are satisfied. For any initial discount rate $r = r_0 \in [d_1, d_2]$, the objective is to maximize

$$\bar{J}(\pi, x, r) = E^{\pi}_{(x,r)}\left[\sum_{n=0}^{N-1} e^{-S_n}(dx^2 + ga^2) + d_N x_N^2\right]$$

where the constant $d_N > 0$, $d_N x_N^2$ is the terminal cost, and S_n is presented in Equation (9).

The dynamic programming equations are

$$\bar{J}_N(x,r) = d_N e^{-r_N} x_N^2,$$
$$\bar{J}_n(x,r) = \min_{A(x,r)}\left[(dx^2 + ga^2) + e^{-r_n} E\bar{J}_{n+1}(bx_n + ca_n + \xi_n, r)\right]$$

$x \in X, \quad n = N-1, \dots, 0$. For $n = N-1$, we have that

$$\begin{aligned}\bar{J}_{N-1}(x,r) &= \min_{A(x,r)}\left[(dx^2 + ga^2) + e^{-(r_{N-1}+r_N)} d_N E[bx + ca + \xi_{N-1}]^2\right] \\ &= \min_{A(x,r)}\Big[a^2(c^2 d_N e^{-(r_{N-1}+r_N)} + g) + a(2bcd_N x e^{-(r_{N-1}+r_N)}) \\ &\qquad + x^2(d + b^2 d_N e^{-(r_{N-1}+r_N)}) + \sigma^2 d_N e^{-(r_{N-1}+r_N)}\Big].\end{aligned}$$

The optimal policy $f_{N-1}(x,r)$ is obtained using the first and second derivative criterion for $\bar{J}_{N-1}$,

$$a_{N-1} = -\frac{bcd_N e^{-(r_{N-1}+r_N)}x}{c^2 d_N e^{-(r_{N-1}+r_N)} + g}.$$

To replace the last value in $\bar{J}_{N-1}(x,r)$, we have that

$$\bar{J}_{N-1}(x,r) = \left[d + b^2 e^{-r_{N-1}} H_N - \frac{b^2 c^2 e^{-2r_{N-1}} H_N^2}{c^2 e^{-r_{N-1}} H_N + g} \right] x^2 + \sigma^2 e^{-r_{N-1}} H_N$$

$x \in X$, $r = r_0$ where $H_n = d_N e^{-r_N}$. For $n = N-1, \ldots, 0$, define

$$H_n := d + b^2 e^{-r_n} H_{n+1} - \frac{b^2 c^2 e^{-2r_n} H_{n+1}^2}{c^2 e^{-r_n} H_{n+1} + g},$$

and

$$P_n := \frac{bcH_{n+1}e^{-r_n}}{c^2 H_{n+1} e^{-r_n} + g}.$$

These terms allow to rewrite the optimal value in each period as

$$\bar{J}_n(x,r) = H_n x^2 + \sigma^2 e^{-r_n} H_{n+1} \quad n = 0, \ldots, N-1$$

and the optimal policy as

$$f_n(x,r) = -P_n x.$$

The optimal value is $\bar{J}_0(x,r)$.

The discounted cost criterion for the *infinite-horizon* case is

$$\bar{V}(\pi, x, r) = E^{\pi}_{(x,r)} \left[\sum_{n=0}^{\infty} e^{-S_n} (dx^2 + ga^2) \right], \ r \in [d_1, d_2].$$

To obtain a non-stationary $\epsilon-$optimal policy we consider the non-stationary policy $\pi' = \{f'_0, f'_1, f'_2, \ldots\}$ defined by the measurable selectors

$$f'_i(x,r) = \sqrt{\frac{d}{g}}x, \ i = 0, 1, 2, \ldots$$

which satisfy

$$\bar{V}(\pi', x, r) = E^{\pi'}_{(x,r)} \left[\sum_{n=0}^{\infty} e^{-S_n} (dx^2 + ga^2) \right] = 2dx^2 E^{\pi'}_{(x,r)} \left[\sum_{n=0}^{\infty} e^{-S_n} \right] < \infty.$$

By finite-horizon case, the finite non-stationary policy $\pi = \{f_0, f_1, \ldots, f_n\}$, where

$$f_i(x,r) = -P_n x, \ i = 1, 2 \ldots, n,$$

is optimal in the first n periods. Now, we consider the policy

$$\overline{\pi} = \{f_0, f_1, \ldots, f_n, f'_{n+1}, f'_{n+2}, f'_{n+3}, \ldots\}.$$

Given $\epsilon > 0$, it is possible to find a natural number n_0 such that if $n > n_0$

$$\bar{V}^{(n)}(\overline{\pi}, x, r_n) - \bar{V}^{(n+1)}(\overline{\pi}, x, r_n e^{-\kappa_n} + \theta_n[1 - e^{-\kappa_n}]) < \epsilon/2^n,$$

i. e., $V^{(n)}(\overline{\pi}, x, r_n)$ approximate to $V^*(x,r)$ and $\overline{\pi}$ is a non-stationary $\epsilon-$ optimal.

CONCLUSION

For discounted MCMs, with discount rates independent of controls and generated by a different recursive function in each period, the dynamic programming algorithm is valid when the horizon is finite. In the infinite-horizon case, using the Measurable Selector Condition, we prove the existence of deterministic non-stationary optimal policies by finite approximation to the value function. Hence, deterministic $\epsilon-$optimal policies can be constructed. Additionally, we give sufficient conditions (Proposition 3.7) to guarantee the existence of deterministic non-stationary optimal policies. A particular case of the discount rate recursive functions is obtained from the expected value of solution of scalar SDEs, defined between the periods of MCM.

An open problem is extend the discounted MCM with recursive discount rates, to semi-Markov control case. This work is in process.

(Received May 12, 2015)

REFERENCES

[1] L. Arnold: Stochastic Differential Equations: Theory and Applications. John Wiley and Sons, New York 1973. DOI:10.1002/zamm.19770570413

[2] R. Ash and C. Doléans-Dade: Probability and Measure Theory. Academic Press, San Diego 2000.

[3] R. Bellman: Dynamic Programming. Princeton Univ. Press, New Jersey 1957.

[4] D. Bertsekas and S. Shreve: Stochastic Optimal Control: The Discrete Time Case. Athena Scientific, Massachusetts 1996.

[5] D. Brigo and F. Mercurio: Interest Rate Models Theory and Practice. Springer-Verlag, New York 2001. DOI:10.1007/978-3-662-04553-4

[6] F. Black and P. Karasinski: Bond and option pricing when short rates are lognormal. Financ. Anal. J. *47* (1991), 4, 52–59. DOI:10.2469/faj.v47.n4.52

[7] Y. Carmon and A. Shwartz: Markov decision processes with exponentially representable discounting. Oper. Res. Lett. *37* (2009), 51–55. DOI:10.1016/j.orl.2008.10.005

[8] E. Della Vecchia, S. Di Marco, and F. Vidal: Dynamic programming for variable discounted Markov decision problems. In: Jornadas Argentinas de Informática e Investigación Operativa (43JAIIO) – XII Simposio Argentino de Investigación Operativa (SIO), Buenos Aires, 2014, pp. 50–62.

[9] J. Cox, J. Ingersoll, and S. Ross: A theory of the term structure of interest rates. Econometrica *53* (1985), 385–407. DOI:10.2307/1911242

[10] U. Dothan: On the term structure interest rates. J. Financ. Econ. *6* (1978), 59–69. DOI:10.1016/0304-405x(78)90020-x

[11] E. Feinberg and A. Shwartz: Markov decision models with weighted discounted criteria. J. Finan. Econ. *19* (1994), 152–168. DOI:10.1287/moor.19.1.152

[12] J. González-Hernández, R. López-Martínez, and J. Pérez-Hernández: Markov control processes with randomized discounted cost. Math. Method Oper. Res. *65* (2006), 1, 27–44. DOI:10.1007/s00186-006-0092-2

[13] J. González-Hernández, R. López-Martínez, and A. Minjarez-Sosa: Adaptive policies for stochastic systems under a randomized discounted cost criterion. Bol. Soc. Mat. Mex. *14* (2008), 3, 149–163.

[14] J. González-Hernández, R. López-Martínez, and A. Minjarez-Sosa: Approximation, estimation and control of stochastic systems under a randomized discounted cost criterion. Kybernetika *45* (2008), 5, 737–754.

[15] X. Guo, A. Hernández-Del-Valle, and O. Hernández-Lerma: First passage problems for a non-stationary discrette-time stochastic control systems. Eur. J. Control *15* (2012), 7, 528–538. DOI:10.3166/ejc.18.528-538

[16] O. Hernández-Lerma and J. B. Lasserre: Discrete-Time Markov Control Processes. Basic Optimality Criteria. Springer–Verlag, New York 1996. DOI:10.1007/978-1-4612-0729-0

[17] J. Minjarez-Sosa: Markov control models with unknown random sate-action-dependent discount factors. TOP *23* (2015), 3, 743–772. DOI:10.1007/s11750-015-0360-5

[18] K. Hinderer: Foundations of non-stationary dynamical programming with discrete time parameter. In: Lecture Notes Operations Research (M. Bechmann and H. Künzi, eds.), Springer-Verlag *33*, Zürich 1970. DOI:10.1007/978-3-642-46229-0

[19] T. Ho and S. Lee: Term structure movements and pricing interest rate claims. J. Financ. *41* (1986), 1011–1029. DOI:10.1111/j.1540-6261.1986.tb02528.x

[20] J. Hull: Options, Futures and other Derivatives. Sixth edition. Prentice Hall, New Jersey 2006.

[21] J. Hull and A. White: Pricing interest rate derivative securities. Rev. Financ. Stud. *3* (1990), 573–592. DOI:10.1093/rfs/3.4.573

[22] F. Mercurio and J. Moraleda: A family of humped volatility models. Eur. J. Finance *7* (2001), 93–116. DOI:10.1080/13518470151141440

[23] R. Rendleman and B. Bartter: The pricing of options on debt securities. J. Financ. Quant. Anal. *15* (1980), 11–24. DOI:10.2307/2979016

[24] U. Rieder: Measurable selection theorems for optimization problems. Manuscripta Math. *24* (1978), 115–131. DOI:10.1007/bf01168566

[25] M. Schäl: Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal. Z. Wahrscheinlichkeit *32* (1975), 179–196. DOI:10.1007/bf00532612

[26] O. Vasicek: An equilibrium characterization of the term structure. J. Financ. Econ. *5* (1977), 177–188. DOI:10.1016/0304-405x(77)90016-2

[27] Q. Wei and X. Guo X: Markov decision processes with state-dependent discount factors and unbounded rewards costs. Oper. Res. Lett. *39* (2011), 369–374. DOI:10.1016/j.orl.2011.06.014

[28] L. Ye and X. Guo: Continuous-time Markov decision processes with state-dependent discount factors. Acta Appl. Math. *121* (2012), 1, 5–27. DOI:10.1007/s10440-012-9669-3

[29] Y. Zhang: Convex analytic approach to constrained discounted Markov decision processes with non-constant discount factors. TOP *21* (2013), 2, 378–408. DOI:10.1007/s11750-011-0186-8

Yofre H. García, Av. Instituto Politécnico Nacional S/N, Edificio 9, Unidad Profesional Adolfo López Mateos, Zacatenco, Distrito Federal. México.

e-mail: ygarciag1104@alumno.ipn.mx

Juan González-Hernández, Avenida Universidad 3000, Ciudad Universitaria, Circuito Escolar S/N, IIMAS-UNAM, Departamento de Probabilidad y Estadística, Coyoacán, C.P. 04510, Distrito Federal. México.
e-mail: juan@sigma.iimas.unam.mx