

Asymptotic Evaluations of the Stability Index for a Markov Control Process with the Expected Total Discounted Reward Criterion

Jaime Eduardo Martínez-Sánchez

Departamento de Ciencias del Tecnológico de Monterrey, Campus Monterrey (ITESM-CM), Ciudad de Monterrey, N.L., México
Email: meduardo@tec.mx

How to cite this paper: Martínez-Sánchez, J.E. (2021) Asymptotic Evaluations of the Stability Index for a Markov Control Process with the Expected Total Discounted Reward Criterion. *American Journal of Operations Research*, 11, 62-85.
<https://doi.org/10.4236/ajor.2021.111004>

Received: December 21, 2020

Accepted: January 25, 2021

Published: January 28, 2021

Copyright © 2021 by author(s) and

Scientific Research Publishing Inc.

This work is licensed under the Creative

Commons Attribution International

License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this work, for a control consumption-investment process with the discounted reward optimization criteria, a numerical estimate of the stability index is made. Using explicit formulas for the optimal stationary policies and for the value functions, the stability index is explicitly calculated and through statistical techniques its asymptotic behavior is investigated (using numerical experiments) when the discount coefficient approaches 1. The results obtained define the conditions under which an approximate optimal stationary policy can be used to control the original process.

Keywords

Control Consumption-Investment Process, Discrete-Time Markov Control Process, Expected Total Discounted Reward, Probabilistic Metrics, Stability Index Estimation

1. Introduction

In a standard way (see [1] [2] for definitions), let \mathcal{M} be a Markov control process at discrete time with infinite horizon (also called Markov decision processes) and let $\tilde{\mathcal{M}}$ be its approximation. We will use the performance criterion (objective function) called expected total discounted reward. Suppose that the optimal control problem for $\tilde{\mathcal{M}}$ has a solution, that is, we can find an optimal solution (\tilde{f}_*) for the approximate process $\tilde{\mathcal{M}}$. Now, if for some reasons (some of these causes are discussed later), it is not possible to find an optimal policy for the original process \mathcal{M} , we could use the policy (\tilde{f}_*) to control the original process \mathcal{M} . The use of such approximation will cause a reduction in the total discounted reward, this reduction is measured by the stability index (Δ) , see [3] [4] [5] for

definition. The importance of this stability index is that it allows us to calibrate the use of (\tilde{f}_*) to control the original process \mathcal{M} .

Clearly, if this stability index is very high ($\Delta \rightarrow \infty$), it would imply that it is not optimal to use the optimal policy (\tilde{f}_*) to control the process \mathcal{M} , on the other hand if this stability index is low ($\Delta \rightarrow 0$), then the use of this approximation is valid.

In the available literature, both the study and calculation of the stability index has been carried out from a theoretical approach through different way: with the application of contractive operators, see for example [6] [7] [8]; with the use of certain ergodicity conditions, see [9] [10] [11]; and with the application of the use of different probabilistic metrics, see [12] for definitions of the different kinds of probabilistic metrics, so for example, in [9] the total variation metric is used, in [6] the Kantorovich metric is used, in [7] and [8] the Prokhorov metric is used.

The results obtained in all the papers mentioned above are an upper bound for the stability index, which is a function of certain parameters and some probabilistic metric, that is

$$\Delta \leq \mathcal{K}\mu(\cdot, \cdot), \tag{1}$$

where \mathcal{K} is an explicit constant and μ is a certain probability metric.

Clearly, the discount factor (α) involved in optimization criteria is also involved in the explicit constant of \mathcal{K} in inequality (1): Our goal is to determine the behavior of the stability index as a function of $(1-\alpha)$ when the discount factor tends to 1 ($\alpha \uparrow 1$).

Unlike the theoretical study of the stability index as presented in inequality (1), in this work, the stability index will be studied with a more applied perspective.

In this work, a Markov control process about consumption-investment is presented (with expected total discounted reward), for which the stability index is explicitly obtained and later we study its asymptotic behavior when the discount factor tends to 1. These asymptotic evaluations for the stability index will be carried out using techniques statistics; as mentioned above, our goal is to measure the sensitivity of the stability index as a function of $(1-\alpha)$ when ($\alpha \uparrow 1$).

To achieve the above, instead of using inequality (1), we will use statistical techniques to estimate the following model:

$$\Delta(1-\alpha) = \frac{\mathcal{W}}{(1-\alpha)^\kappa}, \quad \mathcal{W} \in \mathbb{R}, \quad \kappa \geq 1, \tag{2}$$

where \mathcal{W} and κ are the (unknown) parameters of the model, but estimable from simulated data of the discount factor α and using the simple linear regression analysis technique. From Equation (2), we will say that the stability index is of order $-\kappa$ with respect to $(1-\alpha)$ and we will express this as

$$\Delta(1-\alpha) \sim \mathcal{M}(1-\alpha)^{-\kappa}.$$

Now, if $\alpha \uparrow 1$ clearly for high values of κ the stability index given in Equ-

tion (2) will tend to increase rapidly, which will be indicative that it will not be optimal to use policy \tilde{f}_* to control the original process \mathcal{M} .

The numerical experiments carried out in this work have the goal of estimating the sensitivity κ of the stability index given in Equation (2) when $\alpha \uparrow 1$. These asymptotic evaluations will give us information to answer the question posed above. In the rest of this document, we will refer to this sensitivity (κ) as the order of Δ , indistinctly.

As far as we go in the literature review, no numerical studies, no simulations, etc. were found that use statistical techniques to evaluate the order of the stability index with respect to the discount factor.

The results obtained in this work using the simple linear regression model technique depend on the value of a parameter involved in the discounted reward function used, however, it is clear that results show that when $\alpha \uparrow 1$, then the stability index as a function of $(1-\alpha)$ tends to increase rapidly, so it is not recommended to use an approximate optimal policy (\tilde{f}_*) to control the original model \mathcal{M} . The results also suggest that the selection of the value of the parameter used in the reward function as well as the value of the discount factor are very important to validate the use of the optimal policy \tilde{f}_* to control \mathcal{M} . From the results obtained in the estimates of κ , the largest value was -1.75 , *i.e.*, $\Delta(1-\alpha) \sim \mathcal{M}(1-\alpha)^{-1.75}$, although clearly from the Equation (2), it would seem natural that the best possible order should be at most $\mathcal{M}(1-\alpha)^{-1}$.

Finally, we would like to comment on the reasons why we propose the model given in (2) for the asymptotic study of the stability index.

In [6] [7] and [8], the stability index is studied under expected total discounted cost criterion and the results found are stability inequalities such as the one given in (1). Furthermore, the constant \mathbb{K} involved in inequality (1) is an explicit and inversely proportional function to the term $(1-\alpha)$ in all cases; for example, in [6] it is found that $\Delta \sim \mathcal{M}(1-\alpha)^{-2}$ using the Kantorovich metric, while in [7] it is obtained that $\Delta \sim \mathcal{M}(1-\alpha)^{-2}$ with the total variation metric, and [8] shows a result in which $\Delta \sim \mathcal{M}(1-\alpha)^{-3}$ using the Prokhorv metric. So, given that in this work a control process is studied with the expected total discounted reward criterion and based on the aforementioned results, it seems natural to propose the use of the model given in Equation (2) for the study of the asymptotic evaluations of the stability index. In [9] [10] and [11] there are also stability inequalities like the one given in inequality (1), but using the average-cost criterion, however in these papers the stability index presented an order of $\mathcal{M}(1-\delta)^{-\gamma}$, where δ is the ergodicity parameter and $\gamma \in \mathbb{R}$.

This work is organized as follows. Section 2, a brief description of Markov control models (also called Markov decision processes) is presented as well as some well-known results for discounted optimal control problem with bounded reward; in Section 2.1, we present the problem of estimating the stability index as well as the assumptions that guarantee the existence of the optimal solution for the original (\mathcal{M}) and approximate processes ($\tilde{\mathcal{M}}$), respectively. In Section 3,

the control process with which it will work (consumption-investment) is presented, while in Section 3.1 its stability index is explicitly obtained; in Section 3.2, the results obtained regarding the asymptotic evaluation of the stability index are presented. Finally in Section 4, the conclusions of this work are presented as well as some proposed future researches.

2. The Discounted Reward Criterion

For a topological space (\mathfrak{X}, τ) , $\mathfrak{B}(\mathfrak{X})$ denotes the Borel σ -algebra generated by the topology τ and measurability will always mean Borel measurability. Moreover, $M(\mathfrak{X})$ is the class of measurable functions on \mathfrak{X} whereas $M_b(\mathfrak{X})$ is the subspace of bounded measurable functions endowed with the supremum norm given as $\|u\|_\infty = \sup_{x \in \mathfrak{X}} |u(x)|$, $u \in M_b(\mathfrak{X})$. The subspace of bounded continuous functions is denoted by $C_b(\mathfrak{X})$. For a subset $\mathcal{B} \subseteq \mathfrak{X}$, $\mathbb{I}_{\mathcal{B}}$ stands for the indicator function of \mathcal{B} , i.e., $\mathbb{I}_{\mathcal{B}}(x) = 1$ for $x \in \mathfrak{X}$ and $\mathbb{I}_{\mathcal{B}}(x) = 0$ for $x \notin \mathfrak{X}$. A Borel space \mathcal{Y} is a measurable subset of a complete separable metric space endowed with inherited metric.

Let

$$\mathcal{M} = (\mathcal{X}, \mathcal{A}, \{\mathcal{A}(x) : x \in \mathcal{X}\}, \mathfrak{r}, \mathcal{Q}) \tag{3}$$

be the standard Markov control model (see [1] [13], for definitions). That is thought as a model of a controlled stochastic process $\{(x_n, a_n)\}$, where the state process $\{x_n\}$ takes values in the Borel space \mathcal{X} and the control process $\{a_n\}$ takes values in the Borel space \mathcal{A} . The controlled process involves as follows: at each time $n \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$, the controlled observes the system in some state $x_n = x$ and choose a control $a_n = a$ from the admissible control subset $\mathcal{A}(x)$, which is assumed to be a Borel subset of \mathcal{A} . It is also assumed that the admissible pairs set $\mathbb{K} := \{(x, a) : x \in \mathcal{X}, a \in \mathcal{A}(x)\}$ belongs to $\mathfrak{B}(\mathcal{X} \times \mathcal{A})$. Then, the controller receives a reward $\mathfrak{r}(x, a)$ where \mathfrak{r} is a real-valued Borel measurable function defined on $\mathbb{K} := \{(x, a) : x \in \mathcal{X}, a \in \mathcal{A}(x)\}$. Moreover, the controlled system moves to a new state $x_n = x'$ according to the distribution measure $\mathcal{Q}(\cdot | x, a)$, where \mathcal{Q} is a stochastic kernel on \mathcal{X} given \mathbb{K} , that is, $\mathcal{Q}(\cdot | x, a)$ is a probability measure on \mathbb{K} for each pair $(x, a) \in \mathbb{K}$, and $\mathcal{Q}(\cdot | x, a)$ is a Borel measurable function on \mathbb{K} for each Borel subset \mathcal{B} of \mathcal{X} . Then, the controlled choose a new control $a_n = a' \in \mathcal{A}(x')$ receiving a reward $\mathfrak{r}(x', a')$ and so on.

Let $\mathbb{H}_n := \mathbb{K}^n \times \mathcal{X}$ for $n \in \mathbb{N}_0$ and $\mathbb{H}_0 = \mathcal{X}$. Observe that a generic element of \mathbb{H}_n has the form $h_n = (x_0, a_0, x_1, a_1, \dots, x_{n-1}, a_{n-1}, x_n)$ where $(x_k, a_k) \in \mathbb{K}$ for $k = 0, 1, \dots, n-1$ and $x_n \in \mathcal{X}$. A control policy is a sequence $\pi = \{\pi_n\}$ where $\pi_n(\cdot, \cdot)$ is a stochastic kernel on \mathcal{A} given \mathbb{H}_n satisfying the constraint $\pi_n(\mathcal{A}(x_n) | h_n) = 1$ for all $h_n \in \mathbb{H}_n$, $n \in \mathbb{N}_0$. Now, let \mathbb{F} be the class of all measurable functions $f : \mathcal{X} \rightarrow \mathcal{A}$ such that $f(x) \in \mathcal{A}(x)$ for each $x \in \mathcal{X}$. A control policy $\pi = \{\pi_n\}$ is said to be (deterministic) stationary if there exists $f \in \mathbb{F}$ such that the measure $\pi_n(\cdot | x)$ is concentrated at $f(x)$ for each

$x \in \mathcal{X}$ and $n \in \mathbb{N}_0$. Following a standard convention, the stationary policy π is identified with the selector f . The class of all policies is denoted by Π and the class of all stationary policies is identified with the class \mathbb{F} .

Let $\Omega := (\mathcal{X} \times \mathcal{A})^\infty$ be the canonical sample space and \mathfrak{F} the product σ -algebra. For each policy $\pi = \{\pi_n\} \in \Pi$ and “initial” state $x_0 \in \mathcal{X}$ there exists a probability measure \mathcal{P}_x^π on the measurable space (Ω, \mathfrak{F}) that governs the evolution of the controlled process $\{(x_n, a_n)\}$.

The expected total discounted reward criterion is given as

$$\mathcal{R}_\alpha(x, \pi) := \mathbb{E}_x^\pi \sum_{t=0}^\infty \alpha^t r(x_t, a_t), \tag{4}$$

where the discount factor $\alpha \in (0, 1)$ is fixed and \mathbb{E}_x^π denotes the expectation operator with respect to the probability measure \mathcal{P}_x^π .

The optimal control problem is to find a control policy $\pi^* \in \Pi$ (if exists) such that

$$\mathcal{R}_\alpha^*(x) := \mathcal{R}_\alpha(x, \pi^*) := \sup_{\pi \in \Pi} \mathcal{R}_\alpha(x, \pi), \tag{5}$$

for all $x \in \mathcal{X}$.

The policy π^* is called discounted optimal policy, while \mathcal{R}_α^* is called the discounted optimal value function. Later we will impose conditions that guarantee the finiteness of the value function \mathcal{R}_α^* and the existence of an optimal policy π^* .

2.1. The Stability Index and the Problem of Its Estimation

The problem of (quantitative) estimation stability (“continuity” or “robustness”) arises when there is an uncertainty about the stochastic kernel $Q(\cdot|x, a)$ defined in the standard Markov control model \mathcal{M} (see model (3)). The “original” task of the controller consists in the search for the optimal policy π^* (see Equation (5)). In many applications this task cannot be fulfilled directly due to any of the following causes:

1) Frequently $Q(\cdot|x, a)$ or some of its parameters are unknown to the controller, and this stochastic kernel is estimated using some statistical procedures. With the results of these estimates, another stochastic kernel $\tilde{Q}(\cdot|x, a)$ is generated that is interpreted as an accessible approximation to the unknown $Q(\cdot|x, a)$.

2) There are situations where $Q(\cdot|x, a)$ is known but too complicated to have any hope of solving the control policy optimization problem. In such cases, $Q(\cdot|x, a)$ is sometimes replaced by a “theoretical approximation” $\tilde{Q}(\cdot|x, a)$, which results in a controllable process with a simpler structure.

We assume that $Q(\cdot|x, a)$ is not available to the controller and it is substituted by a given approximating stochastic kernel $\tilde{Q}(\mathcal{B}|x, a)$, $x \in \mathcal{X}$, $\mathcal{A}(x)$ and $\mathcal{B} \in \mathfrak{B}(\mathcal{X})$. The “approximating” Markov process governed by \tilde{Q} will be denoted by $\{\tilde{x}_t\} \equiv \{\tilde{x}_t, t = 0, 1, \dots\}$, *i.e.*, let

$$\tilde{\mathcal{M}} = (\mathcal{X}, \mathcal{A}, \{\mathcal{A}(\tilde{x}) : \tilde{x} \in \mathcal{X}\}, r, \tilde{Q}), \tag{6}$$

be the “approximate” for the Markov control model given in model (3).

Changing x_i for \tilde{x}_i in Equation (4), we get $\tilde{\mathcal{R}}_\alpha(x, \pi)$ the discounted reward criterion for the approximate process $\tilde{\mathcal{M}}$. Now, suppose that it is possible (at least theoretically) to find an optimal policy $\tilde{\pi}^*$ for process $\tilde{\mathcal{M}}$, *i.e.*,

$$\tilde{\mathcal{R}}_\alpha^*(\tilde{x}) := \tilde{\mathcal{R}}_\alpha(\tilde{x}, \tilde{\pi}^*) := \sup_{\pi \in \Pi} \tilde{\mathcal{R}}_\alpha(\tilde{x}, \pi). \tag{7}$$

The control policy $\tilde{\pi}^*$ defined in Equation (7) is used as the approximation to the optimal non-accessible policy π^* (assuming it exists). In other words, policy $\tilde{\pi}^*$ is used to control the original process \mathcal{M} instead of policy π^* .

The reduction in reward for such an approximation, is estimated by the following stability index (see [3] [4] [5]):

$$\Delta_{\mathcal{R}_\alpha}(x) := \mathcal{R}_\alpha(x, \pi^*) - \mathcal{R}_\alpha(x, \tilde{\pi}^*) \geq 0, \quad x \in \mathcal{X}. \tag{8}$$

The stability estimation problem consists of searching for inequalities of the following type:

$$\Delta_{\mathcal{R}_\alpha}(x) \leq \mathcal{K}(x) \psi \left[\mu(\mathcal{Q}, \tilde{\mathcal{Q}}) \right], \quad x \in \mathcal{X}. \tag{9}$$

where $\mathcal{K}(x)$ is a function with explicitly calculated values; $\psi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is a real continuous function such that $\psi(s) \rightarrow 0$ as $s \rightarrow 0$ and μ is a metric probabilistic on the space of probability measures.

The results obtained in [6]-[11] provide inequalities as given in inequality (9).

In this paper, we consider a particular example of a Markov control process for which optimal stationary policies can be explicitly calculated. The explicit form of these stationary policies π^* (for the “original” process \mathcal{M}) and $\tilde{\pi}^*$ (for the “approximate” process $\tilde{\mathcal{M}}$) makes it possible to explicitly calculate the stability index $\Delta_{\mathcal{R}_\alpha}$. The goal of this work is to study the asymptotic behavior of $\Delta_{\mathcal{R}_\alpha}$ when $\alpha \uparrow 1$. Using direct calculations and numerical approximations, we will show that the stability index (see Equation (8)) can be expressed as a function that depends on $(1 - \alpha)$ and has an order of κ , *i.e.*,

$$\Delta_{\mathcal{R}_\alpha} \equiv \Delta_{\mathcal{R}_\alpha}(1 - \alpha) = \frac{\mathcal{W}}{(1 - \alpha)^\kappa}, \quad \mathcal{W} \in \mathbb{R}, \quad \kappa \geq 1, \tag{10}$$

where the (unknown) parameters \mathcal{W} and κ will be estimated using statistical techniques, see the analogy with Equation (2).

To finish this section, the assumptions that guarantee the existence of the stationary optimal control policy (π^* and $\tilde{\pi}^*$) for the optimal control problems given in equations (5) and (7) respectively, are shown below:

Assumption 2.1. (Existence)

- 1) The function $r(\cdot, \cdot)$ is bounded by a constant $b > 0$;
- 2) $\mathcal{A}(x)$ is non-empty compact subset of \mathcal{A} for each $x \in \mathcal{X}$ and the mapping $x \rightarrow \mathcal{A}(x)$ is continuous;
- 3) $r(\cdot, \cdot)$ is a continuous function on \mathbb{K} ;
- 4) $\mathcal{Q}(\cdot | \cdot, \cdot)$ is weakly continuous on \mathbb{K} , that is, the mapping

$$(x, a) \rightarrow \int_{\mathcal{X}} u(y) \mathcal{Q}(dy | x, a), \tag{11}$$

is continuous for each function $u \in C_b(\mathcal{X})$.

The second set of assumptions guarantees the discounted reward criterion is both well defined and finiteness.

Assumption 2.2. (Finiteness)

The following holds for each $x \in \mathcal{X}$:

- 1) The function $r(\cdot, \cdot)$ is bounded by a constant $b > 0$;
- 2) $\mathcal{A}(x)$ is a non-empty compact subset of \mathcal{X} ;
- 3) $r(x, \cdot)$ is a continuous function on $\mathcal{A}(x)$;
- 4) $\mathcal{Q}(\cdot | x, \cdot)$ is strongly continuous on $\mathcal{A}(x)$, that is, the mapping

$$a \rightarrow \int_{\mathcal{X}} u(y) \mathcal{Q}(dy | x, a), \tag{12}$$

is continuous for each function $u \in M_b(\mathcal{X})$.

For more information see [2] [14] [15]. Now, if $\mathcal{C}(\mathcal{X})$ denote either $C_b(\mathcal{X})$ or $M_b(\mathcal{X})$ depending on whether Assumption 2.1 or 2.2 is being used, respectively; then, under either one of Assumption 2.1 or 2.2, the dynamic programming operator:

$$Tu(x) := \sup_{a \in \mathcal{A}(x)} \left[r(x, a) + \alpha \int_{\mathcal{X}} u(y) \mathcal{Q}(dy | x, a) \right], \tag{13}$$

$x \in \mathcal{X}$, is a contraction operator from Banach space $(\mathcal{C}(\mathcal{X}), \|\cdot\|_\infty)$ into itself with contraction factor α (see [2]).

Remark 2.3. Under Assumptions 2.1 and 2.2, there is a solution to the optimal control problem given in Equation (5); which is unique and the value function does not depend on the initial state of the process. For a proof, see [2] or [13].

3. A Markov Control Consumption-Investment Process and Its Approximation

This example is presented in [1]. Consider the following Markov control process:

Let $\mathcal{X} = [0, \infty)$; $\mathcal{A} = [0, \infty)$; $\mathcal{A}(x) = [0, \infty)$, $x \in \mathcal{X}$. The dynamics of the “original” process (\mathcal{M}) is given by:

$$x_t = a_t \xi_t, \text{ for } t = 1, \dots; \tag{14}$$

and for the “approximate” process ($\tilde{\mathcal{M}}$)

$$\tilde{x}_t = \tilde{a}_t \tilde{\xi}_t, \text{ for } t = 1, \dots; \tag{15}$$

where $\{\xi_t, t \geq 1\}$ and $\{\tilde{\xi}_t, t \geq 1\}$ are two sequences of independent and identically distributed non-negative random variables (*i.i.d*), which have distributions \mathcal{F}_ξ and $\mathcal{F}_{\tilde{\xi}}$ respectively. Clearly, \mathcal{F}_ξ and $\mathcal{F}_{\tilde{\xi}}$ are in the space of all distributions in $(\mathcal{X}, \mathfrak{B}(\mathcal{X}))$.

In this model, x_{t-1} is interpreted as current capital. Amount $a_t \in [0, x_{t-1}]$ represents what is invested in assets (such as stocks, bonds, etc.), which generate a profit/loss given by $a_t \xi_t$. The rest of the capital $x_{t-1} - a_t$ is dedicated to consumption and the satisfaction (or benefit) of this consumption is estimated by the utility function given by $(x_{t-1} - a_t)^p$, where $0 < p < 1$ is a given parameter.

The reward function per unit of time is given by

$$r(x_{t-1}, a_t) = (x_{t-1} - a_t)^p \text{ for } t = 1, \dots; 0 < p < 1. \tag{16}$$

Assumption 3.1. (Only for this example)

The *i.i.d* random variables $\{\xi_t\}_{t \geq 1}$, $\{\tilde{\xi}_t\}_{t \geq 1}$ given in Equations (14) and (15) respectively, satisfy the following (for details, see [1]):

$$\lambda := \mathbb{E}\xi^p < \frac{1}{\alpha}; \quad \tilde{\lambda} := \mathbb{E}\tilde{\xi}^p < \frac{1}{\alpha}. \tag{17}$$

Now, for an “initial” state $x \in [0, \infty)$ the optimal control problem (see Equation (5)) for this Markov control consumption-investment process is

$$\mathcal{R}_\alpha(x, \pi^*) := \sup_{\pi \in \Pi} \mathbb{E}_x^\pi \sum_{t=1}^\infty \alpha^{t-1} r(x_{t-1}, a_t)^p, \tag{18}$$

analogously for the “approximate” process, we have

$$\tilde{\mathcal{R}}_\alpha(\tilde{x}, \tilde{\pi}^*) := \sup_{\pi \in \Pi} \mathbb{E}_{\tilde{x}}^\pi \sum_{t=1}^\infty \alpha^{t-1} r(\tilde{x}_{t-1}, \tilde{a}_t)^p, \tag{19}$$

where $\tilde{x} \in [0, \infty)$ is an “initial” state for the “approximate” process.

Under these conditions, in [1] it is shown that the processes are given in equations (14) and (15) satisfies both Assumptions 2.1 and 2.2 and that fulfill the following:

- 1) The optimal stationary policy for Equation (18) is the following selector

$$f_* = (\alpha\lambda)^{\frac{1}{1-p}} x, \quad x \in [0, \infty). \tag{20}$$

- 2) The value function given in Equation (18) is

$$\mathcal{R}_\alpha(x, f_*) = \frac{1}{\left[1 - (\alpha\lambda)^{\frac{1}{1-p}}\right]^{1-p}} x^p, \quad x \in [0, \infty). \tag{21}$$

- 3) The optimal stationary policy for Equation (19) is the following selector

$$\tilde{f}_* = (\alpha\tilde{\lambda})^{\frac{1}{1-p}} \tilde{x}, \quad \tilde{x} \in [0, \infty). \tag{22}$$

The next thing is that we explicitly calculate the stability index for this control process, which we will use to perform the asymptotic evaluations. In the next section, we show the development we did to obtain this calculation.

3.1. Explicit Calculation of the Stability Index for the Markov Control Consumption-Investment Process

In this section, the stability index ($\Delta_{\mathcal{R}_\alpha}$) is explicitly calculated for the control consumption-investment process which was presented in the previous section. As was mentioned in the introduction section, the expression that we find for the stability index is a function of the parameters p and ϵ , where ϵ is the measure of the approximation between the probability distributions \mathcal{F}_ξ and $\mathcal{F}_{\tilde{\xi}}$ (see Equations (14) and (15)), while p is the parameter involved in the reward function (see Equation (16)).

In economics, this parameter p is associated with elasticity, that is, elasticity measures the percentage change in the consumer’s utility in response to percentage changes in the consumer’s money supply (for more details, see [16] or [17]). For this reason, it is important to measure its effect on the asymptotic behavior of the stability index.

From Equation (16), the possible values for the parameter p lie in interval $0 < p < 1$.

Our goal is to calculate asymptotic evaluations of the stability index when $\epsilon \rightarrow 0$ (which would imply that \mathcal{F}_{ξ} is closer to $\mathcal{F}_{\tilde{\xi}}$) and for extreme values of the range of p , that is, we are interested in values of $p \approx 0$, $p \approx \frac{1}{2}$ and $p \approx 1$.

Now, we will proceed to calculate the stability index and for this, we will take an “initial” state $x = \tilde{x} = 1$ as well as the following distribution functions to measure the effect of the shock on the processes:

Assumption 3.2. (Only for this example)

We consider the random vectors given in processes (14) and (15) respectively, have an exponential distribution with parameters θ and $\tilde{\theta}$ respectively, *i.e.*, $\xi \sim \mathcal{F}_{\xi} \equiv \exp(\theta)$ and $\tilde{\xi} \sim \mathcal{F}_{\tilde{\xi}} \equiv \exp(\tilde{\theta})$ with $\tilde{\theta} = \theta(1 - \epsilon)$, where the values of ϵ measure the approximation between both distributions, $0 < \epsilon < 1$.

Under Assumption 3.1 and 3.2, we have

$$\lambda := \mathbb{E}_{\xi_1} \xi_1^p = \int_0^{\infty} \xi^p \left[\frac{e^{-\frac{\xi}{\theta}}}{\theta} \right] d\xi,$$

and after some direct calculations,

$$\lambda = \theta^p \Gamma(p). \tag{23}$$

Similarly, for the perturbed random vector, we have

$$\tilde{\lambda} = \tilde{\theta}^p \Gamma(p),$$

and since $\tilde{\theta} = \theta(1 - \epsilon)$, then from the above equality it follows that

$$\tilde{\lambda} = \lambda(1 - \epsilon)^p. \tag{24}$$

Next, the stability index is calculated.

From Equation (8) we have

$$\Delta_{\mathcal{R}_{\alpha}}(1) := \mathcal{R}_{\alpha}(1, f_*) - \mathcal{R}_{\alpha}(1, \tilde{f}_*) \geq 0. \tag{25}$$

The first term on the right side of Equation (25) is given in Equation (21) with $x = 1$. The next thing is that we will calculate the second term on the right side of Equation (25): To do this, we substitute the approximate policy of optimal control with $\tilde{x} = 1$, given in Equation (22), in the reward function of the “original” model given in Equation (18), and we have

$$\mathcal{R}_{\alpha}(1, \tilde{f}_*) = \mathbb{E}_1^{\tilde{f}_*} \sum_{t=1}^{\infty} \alpha^{t-1} (\tilde{x}_{t-1} - \tilde{a}_{t-1})^p.$$

The above equation represents the discounted reward obtained when the tra-

jectory of the “original” process given in Equation (14) is controlled by the optimal policy obtained from the “approximate” process given in Equation (15) and the “initial” state is $x = 1$.

Now, since $a_t \in f(x_{t-1})$ (see [1] for details), we have

$$\begin{aligned} \mathcal{R}_\alpha(1, \tilde{f}_*) &= \mathbb{E}_1^{\tilde{f}_*} \sum_{t=1}^\infty \alpha^{t-1} (\tilde{x}_{t-1} - \tilde{a}_{t-1})^p, \\ \mathcal{R}_\alpha(1, \tilde{f}_*) &= \sum_{t=1}^\infty \alpha^{t-1} \mathbb{E}_1^{\tilde{f}_*} \left(\tilde{x}_{t-1} - (\alpha \tilde{\lambda})^{\frac{1}{1-p}} \tilde{x}_{t-1} \right)^p, \end{aligned}$$

finally, we have

$$\mathcal{R}_\alpha(1, \tilde{f}_*) = \left[1 - (\alpha \tilde{\lambda})^{\frac{1}{1-p}} \right]^p \sum_{t=0}^\infty \alpha^t \mathbb{E}_1^{\tilde{f}_*} \tilde{x}_t^p. \tag{26}$$

Now, the evolution of the approximate process (see Equation (15)) is represented as follows

$$\tilde{x}_t = \tilde{a}_t \tilde{\xi}_t,$$

So,

$$\begin{aligned} \tilde{x}_1 &= \tilde{a}_1 \tilde{\xi}_1 = (\alpha \tilde{\lambda})^{\frac{1}{1-p}} \tilde{x}_1 \tilde{\xi}_1 = (\alpha \tilde{\lambda})^{\frac{1}{1-p}} \tilde{\xi}_1, \\ \tilde{x}_2 &= \tilde{a}_2 \tilde{\xi}_2 = (\alpha \tilde{\lambda})^{\frac{1}{1-p}} \tilde{x}_1 \tilde{\xi}_2 = (\alpha \tilde{\lambda})^{\frac{2}{1-p}} \tilde{\xi}_1 \tilde{\xi}_2, \\ &\vdots \\ \tilde{x}_t &= \tilde{a}_t \tilde{\xi}_t = (\alpha \tilde{\lambda})^{\frac{1}{1-p}} \tilde{x}_{t-1} \tilde{\xi}_t = (\alpha \tilde{\lambda})^{\frac{t}{1-p}} \tilde{\xi}_1 \tilde{\xi}_2 \dots \tilde{\xi}_t. \end{aligned}$$

If we raise the last equality to the power p , we have

$$\tilde{x}_t^p = (\alpha \tilde{\lambda})^{\frac{pt}{1-p}} \tilde{\xi}_1^p \tilde{\xi}_2^p \dots \tilde{\xi}_t^p.$$

Now if we take the expected value on both sides of the above equality and since the random elements are *i.i.d.*,

$$\begin{aligned} \mathbb{E}_1^{\tilde{f}_*} \tilde{x}_t^p &= (\alpha \tilde{\lambda})^{\frac{pt}{1-p}} \mathbb{E}_1^{\tilde{f}_*} \left[\tilde{\xi}_1^p \tilde{\xi}_2^p \dots \tilde{\xi}_t^p \right], \\ \mathbb{E}_1^{\tilde{f}_*} \tilde{x}_t^p &= (\alpha \tilde{\lambda})^{\frac{pt}{1-p}} \mathbb{E}_1^{\tilde{f}_*} \left(\tilde{\xi}_1^p \right) \mathbb{E}_1^{\tilde{f}_*} \left(\tilde{\xi}_2^p \right) \dots \mathbb{E}_1^{\tilde{f}_*} \left(\tilde{\xi}_t^p \right). \end{aligned}$$

Now, by inequality (17),

$$\begin{aligned} \mathbb{E}_1^{\tilde{f}_*} \tilde{x}_t^p &= (\alpha \tilde{\lambda})^{\frac{pt}{1-p}} (\tilde{\lambda})^t, \\ \mathbb{E}_1^{\tilde{f}_*} \tilde{x}_t^p &= (\alpha)^{\frac{p}{1-p}t} (\tilde{\lambda})^{\frac{1}{1-p}t}. \end{aligned} \tag{27}$$

Substituting Equation (27) in Equation (26) and after performing some direct calculations, we have

$$\mathcal{R}_\alpha(1, \tilde{f}_*) = \left[1 - (\alpha \tilde{\lambda})^{\frac{1}{1-p}} \right]^p \sum_{t=0}^\infty \alpha^t (\alpha)^{\frac{p}{1-p}t} (\tilde{\lambda})^{\frac{1}{1-p}t},$$

$$\mathcal{R}_\alpha(1, \tilde{f}_*) = \left[1 - (\alpha\tilde{\lambda})^{\frac{1}{1-p}} \right]^p \sum_{t=0}^{\infty} \left[(\alpha\tilde{\lambda})^{\frac{1}{1-p}} \right]^t. \tag{28}$$

Inequalities (17) guarantees $\alpha\tilde{\lambda} < 1$; furthermore, since $0 < p < 1$, it is guaranteed that $1 < \frac{1}{1-p}$. Finally, the two reasons above guarantee that $(\alpha\tilde{\lambda})^{\frac{1}{1-p}} < 1$.

Therefore, calculating the sum of geometric serie involved in Equation (28), this one can be expressed as

$$\begin{aligned} \mathcal{R}_\alpha(1, \tilde{f}_*) &= \left[1 - (\alpha\tilde{\lambda})^{\frac{1}{1-p}} \right]^p \frac{1}{1 - (\alpha\tilde{\lambda})^{\frac{1}{1-p}}}, \\ \mathcal{R}_\alpha(1, \tilde{f}_*) &= \frac{1}{\left[1 - (\alpha\tilde{\lambda})^{\frac{1}{1-p}} \right]}. \end{aligned} \tag{29}$$

Then, to obtain the stability index, Equation (21) with $x=1$ and Equation (29) are substituted in Equation (25) and we obtain

$$\Delta_{\mathcal{R}_\alpha}(1) = \frac{1}{\left[1 - (\alpha\lambda)^{\frac{1}{1-p}} \right]^{1-p}} - \frac{1}{\left[1 - (\alpha\tilde{\lambda})^{\frac{1}{1-p}} \right]^{1-p}}. \tag{30}$$

Now, substituting Equation (24) in Equation (30), we have

$$\Delta_{\mathcal{R}_\alpha}(1) = \frac{1}{\left[1 - (\alpha\lambda)^{\frac{1}{1-p}} \right]^{1-p}} - \frac{1}{\left[1 - (\alpha\lambda(1-\epsilon)^p)^{\frac{1}{1-p}} \right]^{1-p}}. \tag{31}$$

For each fixed p , a θ value in Equation (23) can be selected such that $\lambda = 1$, so Equation (31) can be written as

$$\Delta_{\mathcal{R}_\alpha}(1) = \frac{1}{\left[1 - \alpha^{\frac{1}{1-p}} \right]^{1-p}} - \frac{1}{\left[1 - \alpha^{\frac{1}{1-p}} (1-\epsilon)^{\frac{p}{1-p}} \right]^{1-p}}. \tag{32}$$

The stability index given in Equation (32) remains a function that depends on the discount factor (α), the parameter p of the reward function, see Equation (16), and the level of approximation ϵ of the distributions \mathcal{F}_ξ and $\mathcal{F}_{\tilde{\xi}}$, see Assumption 3.2.

3.2. Study of the Asymptotic Evaluations of the Stability Index

The goal of this work is to perform asymptotic numerical estimations of the stability index as a function of $(1-\alpha)$, that is, find its order (κ) when $\alpha \uparrow 1$, see Equation (10). For this, we will use the result obtained in the previous section of the explicit calculation of the stability index, see Equation (32).

Equation (32) shows that the stability index is a function of p and as mentioned in the previous section, this parameter of the utility function is important

in economics since it is related to elasticity. So, to estimate the effect that this parameter has on the stability index, we will select arbitrary values of this parameter in such way that 1) values close to zero (it would imply consumers insensitive to monetary change); 2) values close to $\frac{1}{2}$ (average consumers); and 3) values close to 1 (sensitive consumers). However for our goal, these values of p would give us information about the conditions in which the approximate policy \tilde{f} can be used to control the original process \mathcal{M} , that is, we want to study if values given of p close to zero (to $\frac{1}{2}$ and to 1) in the reward function allows us to use this approach.

Methodology and results obtained. For a fixed value of p in Equation (32) and given a value of ϵ , we will generate 100 values of α , starting at $\alpha = 0.5$ with increments of 0.005. Then, for each of the 100 generated values of $\alpha = 0.5, 0.505, \dots, 0.995$, the value of $(1 - \alpha)$ is substituted in Equation (32) and we would have 100 values of the stability index (as function of $(1 - \alpha)$). With these 100 values of $(1 - \alpha)$ and the stability index $\Delta_{\mathcal{R}_\alpha}(1 - \alpha)$, a simple linear regression model is performed to estimate the κ parameter involved in Equation (10) and this value would be the estimation of the order of the stability index with respect to $(1 - \alpha)$, *i.e.*, $\Delta_{\mathcal{R}_\alpha} \sim (1 - \alpha)^{-\kappa}$. We are interested in the behavior of the k estimate when $\alpha \uparrow 1$ and $\epsilon \rightarrow 0$.

For example, if $p = \frac{1}{100}$ then of Equation (32) we have that the stability index is expressed as

$$\Delta_{\mathcal{R}_\alpha}(1) = \frac{1}{\left[1 - \alpha^{\frac{100}{99}}\right]^{\frac{99}{100}}} - \frac{1}{\left[1 - \alpha^{\frac{100}{99}}(1 - \epsilon)^{\frac{1}{99}}\right]^{\frac{99}{100}}}, \tag{33}$$

Now, remembering that ϵ values represent the measure of the approximation between the distributions \mathcal{F}_ξ and $\mathcal{F}_{\tilde{\xi}}$ (see Assumption 3.2), so let's assume $\epsilon = 0.2$ and we substitute it in Equation (33), we have

$$\Delta_{\mathcal{R}_\alpha}(1) = \frac{1}{\left[1 - \alpha^{\frac{100}{99}}\right]^{\frac{99}{100}}} - \frac{1}{\left[1 - \alpha^{\frac{100}{99}}(0.8)^{\frac{1}{99}}\right]^{\frac{99}{100}}}, \tag{34}$$

Now, we generate 100 values of $\alpha = 0.5, 0.505, \dots, 0.995$ and later we substitute $(1 - \alpha)$ in Equation (34) and 100 values of the stability index are generated, that is shown in **Figure 1**.

Remark 3.3. In **Figure 1**, the stability index $\Delta_{\mathcal{R}_\alpha}(1)$ given in Equation (34) is represented as delta, this is, $\Delta_{\mathcal{R}_\alpha}(1) \equiv \text{delta}$ and the measure ϵ , we call epsilon.

From **Figure 1**, we can see that when $\alpha \uparrow 1 \Rightarrow \text{delta} \equiv \Delta_{\mathcal{R}_\alpha}(1) \rightarrow \infty$ that is, it is very costly to use the optimal policy of the approximate process given in Equation (22) to control the original process given in Equation (14).

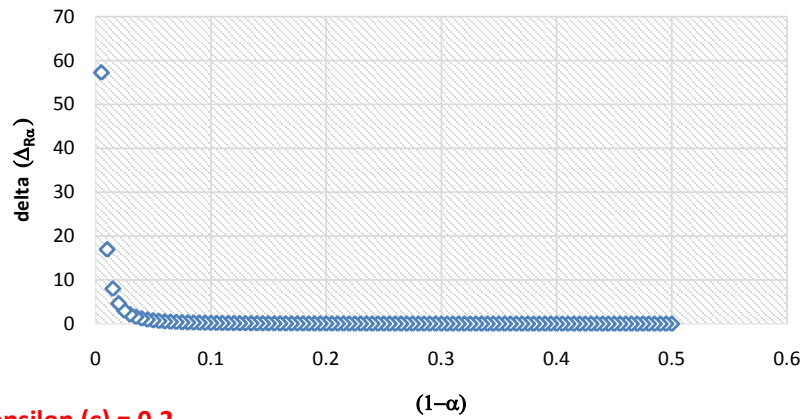


Figure 1. Scatterplot generated by 100 data points of stability index ($\Delta_{\mathcal{R}_\alpha}$) obtained from Equation (34).

On the other hand, to obtain the asymptotic evaluations of the stability index when $\alpha \uparrow 1$, that is, the estimation of the κ parameter that appears in Equation (10):

$$\Delta_{\mathcal{R}_\alpha} \equiv \Delta_{\mathcal{R}_\alpha}(1-\alpha) = \frac{\mathcal{W}}{(1-\alpha)^\kappa}, \quad \mathcal{W} \in \mathbb{R}, \quad \kappa \geq 1,$$

we will proceed to estimate the following simple linear regression model:

$$\ln[\Delta_{\mathcal{R}_\alpha}(1-\alpha)]_i = \ln(\mathcal{W}) + \kappa \ln(1-\alpha)_i + \nu_i, \quad i = 1, \dots, 100. \quad (35)$$

where ν_i is a white noise (see [18] for definition), \mathcal{W} and κ are the parameters to be estimated with the results of 100 data generated and represented in **Figure 1**. The results of the regression estimate given in Equation (35) is shown below:

Regression Analysis: ln(delta) versus ln(1-alpha)

Regression Equation							
ln(delta)	=	-6.7003 - 2.1369 ln(1-alpha)					
Coefficients							
Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF	
Constant	-6.7003	0.0252	(-6.7503; -6.6503)	-265.92	0.000		
ln(1-alpha)	-2.1369	0.0133	(-2.1632; -2.1106)	-161.17	0.000	1.00	
Model Summary							
S	R-sq	R-sq(adj)	PRESS	R-sq(pred)	AICc	BIC	
0.122431	99.62%	99.62%	1.68143	99.57%	-132.02	-124.46	
Analysis of Variance							
Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	1	389.360	99.62%	389.360	389.360	25975.63	0.000
ln(1-alpha)	1	389.360	99.62%	389.360	389.360	25975.63	0.000
Error	98	1.469	0.38%	1.469	0.015		
Total	99	390.829	100.00%				

Therefore, from the above results we have that $\hat{\kappa} = -2.1369$ and from Equation (10) it can be concluded that the asymptotic estimate of the stability index when $\alpha \uparrow 1$, it is $\hat{\kappa} = -2.1369$, that is, the sensitivity of the stability index with respect to $(1-\alpha)$ is

$$\Delta_{\mathcal{R}_\alpha}(1-\alpha) \sim \mathcal{M}(1-\alpha)^{-2.1369}. \tag{36}$$

On the other hand, the estimation of this asymptotic evaluation of κ will be better when the approximation of the \mathcal{F}_ξ distribution is closer to the \mathcal{F}_ξ distribution (see Assumption 3.2), that is

$$\text{if } \epsilon \rightarrow 0, \text{ then } \mathcal{F}_\xi \rightarrow \mathcal{F}_\xi \text{ (and so } \hat{\kappa} \rightarrow \kappa). \tag{37}$$

To see the above, given the fixed value of $p = \frac{1}{100}$, we proceeded to replicate the estimates of κ given in Equation (35) for $\epsilon = 0.1, 0.05, 0.01, 0.001$.

For $p = \frac{1}{100}$ and $\epsilon = 0.1$, from Equation (33) we have the following stability index

$$\Delta_{\mathcal{R}_\alpha}(1) = \frac{1}{\left[1 - \alpha^{\frac{100}{99}}\right]^{\frac{99}{100}}} - \frac{1}{\left[1 - \alpha^{\frac{100}{99}} (0.9)^{\frac{1}{99}}\right]^{\frac{99}{100}}}. \tag{38}$$

For the same 100 values of $(1-\alpha)$ and using the above equation, another 100 values of the stability index were generated, which are presented in **Figure 2**.

From this **Figure 2**, we observe that for $p = \frac{1}{100}$ and $\epsilon = 0.1$, (when $\alpha \uparrow 1$) it remains it is very costly to use the optimal policy of the approximate process given in Equation (22) to control the original process given in Equation (14); however, the stability index $\Delta_{\mathcal{R}_\alpha}(1) \equiv \text{delta}$ is reduced, that is due to the greater precision $\epsilon = 0.1$ of in the approximation of the distribution \mathcal{F}_ξ to the distribution \mathcal{F}_ξ .

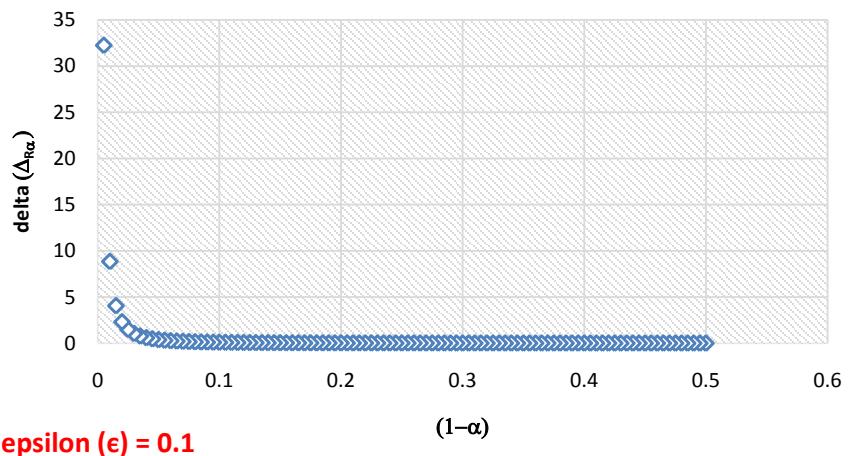


Figure 2. Scatterplot generated by 100 data points of stability index $(\Delta_{\mathcal{R}_\alpha})$ obtained from Equation (38).

Now, with this new 100 data from **Figure 2**, the κ parameter is re-estimated in the simple linear regression model given in Equation (35). The results obtained are the following:

Regression Analysis: ln(delta) versus ln(1-alpha)

Regression Equation							
ln(delta)	=	-7.4722 - 2.1562 ln(1-alpha)					
Coefficients							
Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF	
Constant	-7.4722	0.0231	(-7.5179; -7.4264)	-324.07	0.000		
ln(1-alpha)	-2.1562	0.0121	(-2.1803; -2.1321)	-177.71	0.000	1.00	
Model Summary							
S	R-sq	R-sq(adj)	PRESS	R-sq(pred)	AICc	BIC	
0.112037	99.69%	99.69%	1.38898	99.65%	-149.77	-142.20	
Analysis of Variance							
Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	1	396.422	99.69%	396.422	396.422	31581.39	0.000
ln(1-alpha)	1	396.422	99.69%	396.422	396.422	31581.39	0.000
Error	98	1.230	0.31%	1.230	0.013		
Total	99	397.652	100.00%				

The results show that $\hat{\kappa} = -2.1562$, and we obtain that the stability index has an order -2.156 with respect to $(1-\alpha)$, that is $\Delta_{\mathcal{R}_\alpha}(1-\alpha) \sim \mathcal{M}(1-\alpha)^{-2.1562}$.

Now, to investigate the asymptotic behavior of this sensitivity κ , we will make the approximation between the probability functions better and better, *i.e.*, $\epsilon \rightarrow 0$.

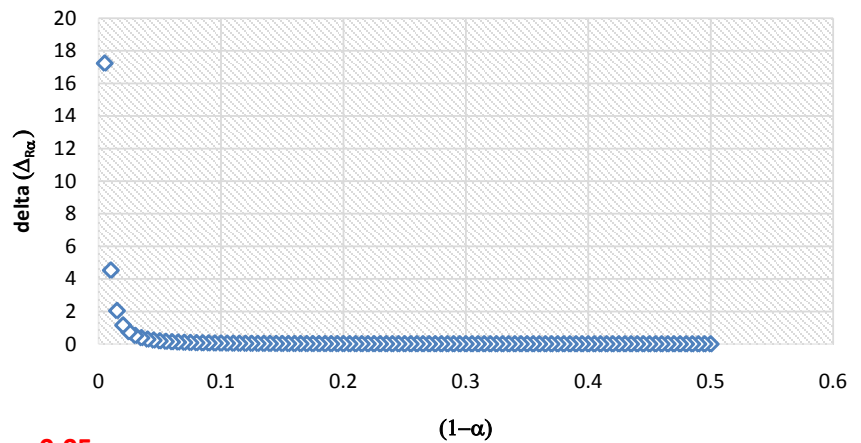
So, analogously to what has already been explained, the results for $p = \frac{1}{100}$ and $\epsilon = 0.05, 0.01, 0.001$ are presented in **Figures 3-5**.

The five figures above show that for fixed $p = \frac{1}{100}$ and when ϵ tends to zero (which implies that \mathcal{F}_ϵ approaches to \mathcal{F}_ϵ), then the stability index has zero (see y-values labels).

The previous interpretation is clearer if we look at **Figure 6** and **Figure 7**, in which we have joined the five previous figures in contour lines of the stability index.

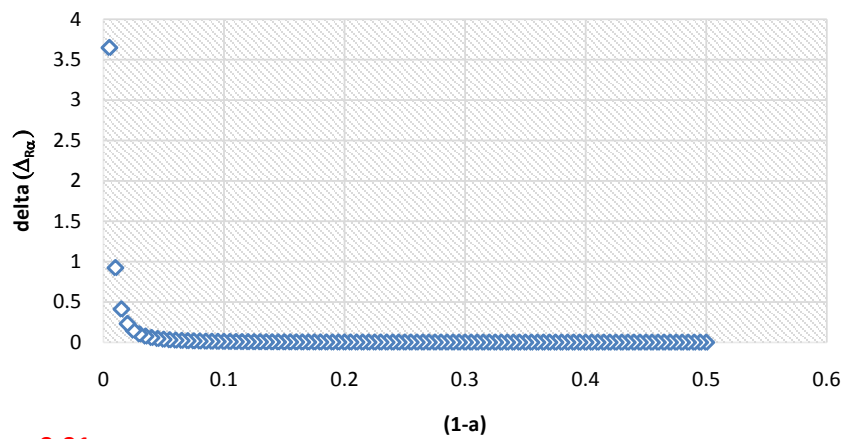
In the last two figures observe y-values labels, it is clear that when epsilon tends to zero, then the stability index also tends to zero. The above implies that, the better the approximation between the distribution functions then the approximate optimal policy can be used to control the original process.

Now, for each group of 100 data generated in each of the five graphs, the κ parameter involved in the simple linear regression model given in Equation (35) was estimated. The results obtained from these estimates are presented in **Table 1**,



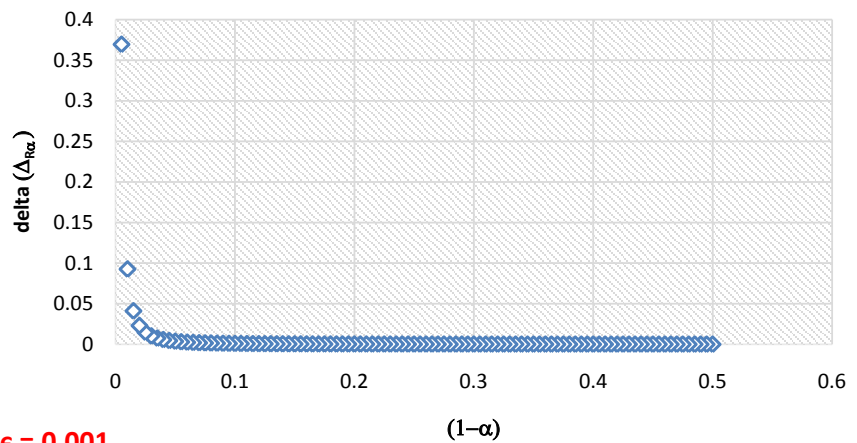
$\epsilon = 0.05$

Figure 3. Scatterplot generated by 100 data points of stability index (Δ_{R_α}) obtained from Equation (33) with $\epsilon = 0.05$.



$\epsilon = 0.01$

Figure 4. Scatterplot generated by 100 data points of stability index (Δ_{R_α}) obtained from Equation (33) with $\epsilon = 0.01$.



$\epsilon = 0.001$

Figure 5. Scatterplot generated by 100 data points of stability index (Δ_{R_α}) obtained from Equation (33) with $\epsilon = 0.001$.

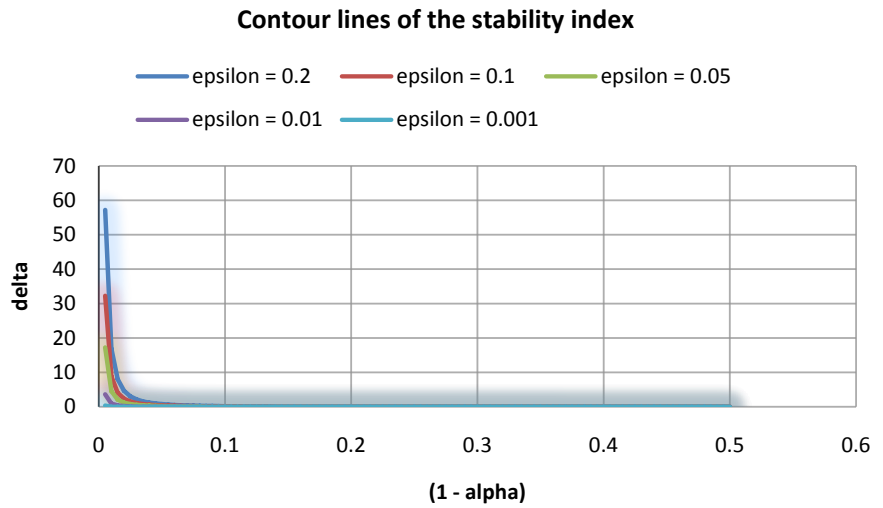


Figure 6. Results of the association of the stability index and the approximation measure in the probability distributions (epsilon).

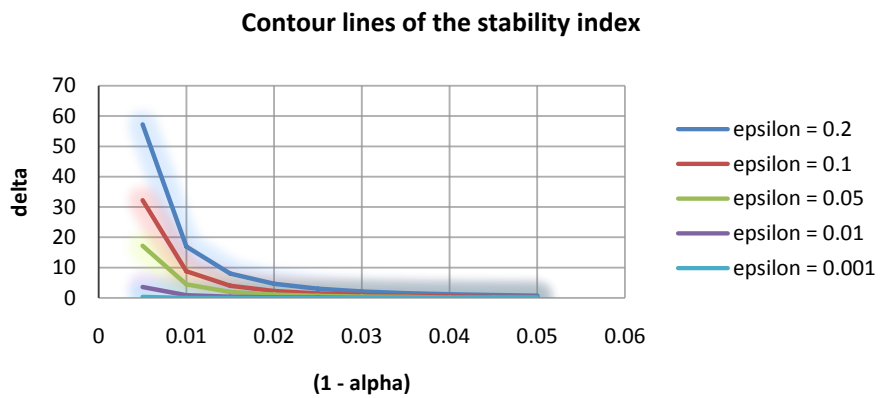


Figure 7. Magnification of **Figure 6**, when alpha approaches to 1.

Table 1. Asymptotic evaluation of the stability index ($\hat{\kappa}$).

	ϵ	$\hat{\kappa}$
$p = 1/100$	0.2	-2.136
	0.1	-2.156
	0.05	-2.160
	0.01	-2.173
	0.001	-2.175

note that the first two cases ($\epsilon = 0.2; \epsilon = 0.1$) correspond to the results that have been explained in previous pages.

In **Table 1**, the green cell shows the best approximation used between the distribution functions (see Assumption 3.2) with which the numerical estimate for the asymptotic evaluation of the stability index was found, which is shown in the skyblue cell.

Based on results of **Table 1**, we can conclude that for $p = \frac{1}{100}$, when $\epsilon \rightarrow 0$

and $\alpha \uparrow 1$, then the asymptotic evaluation of the stability index is $\hat{\kappa} \approx -2.175$, *i.e.*, the stability index has an order

$$\Delta_{\mathcal{R}_\alpha}(1-\alpha) \sim \mathcal{M}(1-\alpha)^{-2.175}.$$

To study the sensitivity of the stability index ($\hat{\kappa}$), numerical experiments were carried out for other values of p . Each of these p values was substituted in Equation (32) and the stability indices ($\Delta_{\mathcal{R}_\alpha}$) were obtained as a function of α and ϵ as shown in **Table 2**.

Then, for each fixed value of p given in **Table 2**, we will use $\epsilon = 0.2, 0.1, 0.05, 0.01, 0.001$; subsequently for each pair of fixed p and ϵ , 100 data of $(1-\alpha)$ were generated and they were substituted in the formulas of **Table 2** obtaining 100 data of the index stability as a function of $(1-\alpha)$; finally, these 100 pairs of $(1-\alpha)$ and $\Delta_{\mathcal{R}_\alpha}(1-\alpha)$ were used for the asymptotic evaluation of the approximate stability index with the estimation of the κ parameter involved in the simple linear regression model given in Equation (35). The results obtained in these numerical estimates are presented in **Table 3**.

Table 2. Explicit expression of the stability index for different values of the reward function parameter (p).

p	Stability index
$\frac{1}{100}$	$\Delta_{\mathcal{R}_\alpha}(1) = \frac{1}{\left(1-\alpha^{\frac{100}{99}}\right)^{\frac{99}{100}}} - \frac{1}{\left[1-\alpha^{\frac{100}{99}}(1-\epsilon)^{\frac{1}{99}}\right]^{\frac{99}{100}}}$
$\frac{1}{50}$	$\Delta_{\mathcal{R}_\alpha}(1) = \frac{1}{\left(1-\alpha^{\frac{50}{49}}\right)^{\frac{49}{50}}} - \frac{1}{\left[1-\alpha^{\frac{50}{49}}(1-\epsilon)^{\frac{1}{49}}\right]^{\frac{49}{50}}}$
$\frac{1}{20}$	$\Delta_{\mathcal{R}_\alpha}(1) = \frac{1}{\left(1-\alpha^{\frac{20}{19}}\right)^{\frac{19}{20}}} - \frac{1}{\left[1-\alpha^{\frac{20}{19}}(1-\epsilon)^{\frac{1}{19}}\right]^{\frac{19}{20}}}$
$\frac{1}{10}$	$\Delta_{\mathcal{R}_\alpha}(1) = \frac{1}{\left(1-\alpha^{\frac{10}{9}}\right)^{\frac{9}{10}}} - \frac{1}{\left[1-\alpha^{\frac{10}{9}}(1-\epsilon)^{\frac{1}{9}}\right]^{\frac{9}{10}}}$
$\frac{1}{9}$	$\Delta_{\mathcal{R}_\alpha}(1) = \frac{1}{\left(1-\alpha^{\frac{9}{8}}\right)^{\frac{8}{9}}} - \frac{1}{\left[1-\alpha^{\frac{9}{8}}(1-\epsilon)^{\frac{1}{8}}\right]^{\frac{8}{9}}}$
$\frac{1}{5}$	$\Delta_{\mathcal{R}_\alpha}(1) = \frac{1}{\left(1-\alpha^{\frac{5}{4}}\right)^{\frac{4}{5}}} - \frac{1}{\left[1-\alpha^{\frac{5}{4}}(1-\epsilon)^{\frac{1}{4}}\right]^{\frac{4}{5}}}$
$\frac{1}{4}$	$\Delta_{\mathcal{R}_\alpha}(1) = \frac{1}{\left(1-\alpha^{\frac{4}{3}}\right)^{\frac{3}{4}}} - \frac{1}{\left[1-\alpha^{\frac{4}{3}}(1-\epsilon)^{\frac{1}{3}}\right]^{\frac{3}{4}}}$
$\frac{3}{10}$	$\Delta_{\mathcal{R}_\alpha}(1) = \frac{1}{\left(1-\alpha^{\frac{10}{7}}\right)^{\frac{7}{10}}} - \frac{1}{\left[1-\alpha^{\frac{10}{7}}(1-\epsilon)^{\frac{3}{7}}\right]^{\frac{7}{10}}}$

Continued

$\frac{2}{5}$	$\Delta_{\kappa_\alpha}(1) = \frac{1}{\left(1 - \alpha^{\frac{5}{3}}\right)^{\frac{3}{5}}} - \frac{1}{\left[1 - \alpha^{\frac{5}{3}}(1 - \epsilon)^{\frac{2}{3}}\right]^{\frac{3}{5}}}$
$\frac{1}{2}$	$\Delta_{\kappa_\alpha}(1) = \frac{1}{(1 - \alpha^2)^{\frac{1}{2}}} - \frac{1}{\left[1 - \alpha^2(1 - \epsilon)\right]^{\frac{1}{2}}}$
$\frac{3}{5}$	$\Delta_{\kappa_\alpha}(1) = \frac{1}{\left(1 - \alpha^{\frac{5}{2}}\right)^{\frac{2}{5}}} - \frac{1}{\left[1 - \alpha^{\frac{5}{2}}(1 - \epsilon)^{\frac{3}{2}}\right]^{\frac{2}{5}}}$
$\frac{7}{10}$	$\Delta_{\kappa_\alpha}(1) = \frac{1}{\left(1 - \alpha^{\frac{10}{3}}\right)^{\frac{3}{10}}} - \frac{1}{\left[1 - \alpha^{\frac{10}{3}}(1 - \epsilon)^{\frac{7}{3}}\right]^{\frac{3}{10}}}$
$\frac{3}{4}$	$\Delta_{\kappa_\alpha}(1) = \frac{1}{(1 - \alpha^4)^{\frac{1}{4}}} - \frac{1}{\left[1 - \alpha^4(1 - \epsilon)^3\right]^{\frac{1}{4}}}$
$\frac{4}{5}$	$\Delta_{\kappa_\alpha}(1) = \frac{1}{(1 - \alpha^5)^{\frac{1}{5}}} - \frac{1}{\left[1 - \alpha^5(1 - \epsilon)^4\right]^{\frac{1}{5}}}$
$\frac{8}{9}$	$\Delta_{\kappa_\alpha}(1) = \frac{1}{(1 - \alpha^9)^{\frac{1}{9}}} - \frac{1}{\left[1 - \alpha^9(1 - \epsilon)^8\right]^{\frac{1}{9}}}$
$\frac{9}{10}$	$\Delta_{\kappa_\alpha}(1) = \frac{1}{(1 - \alpha^{10})^{\frac{1}{10}}} - \frac{1}{\left[1 - \alpha^{10}(1 - \epsilon)^9\right]^{\frac{1}{10}}}$
$\frac{95}{100}$	$\Delta_{\kappa_\alpha}(1) = \frac{1}{(1 - \alpha^{20})^{\frac{5}{100}}} - \frac{1}{\left[1 - \alpha^{20}(1 - \epsilon)^{19}\right]^{\frac{5}{100}}}$
$\frac{99}{100}$	$\Delta_{\kappa_\alpha}(1) = \frac{1}{(1 - \alpha^{100})^{\frac{1}{100}}} - \frac{1}{\left[1 - \alpha^{100}(1 - \epsilon)^{99}\right]^{\frac{1}{100}}}$

Table 3. Explicit expression of the stability index for different values of the reward function parameter (p).

	ϵ	$\hat{\kappa}$
$p = 1/100$	0.2	-2.136
	0.1	-2.156
	0.05	-2.160
	0.01	-2.173
	0.001	-2.175
$p = 1/50$	0.2	-2.090
	0.1	-2.129
	0.05	-2.140
	0.01	-2.160
	0.001	-2.165
$p = 1/20$	0.2	-1.99

Continued

	0.1	-2.06
	0.05	-2.1
	0.01	-2.12
	0.001	-2.135
	0.2	-1.865
<i>p</i> = 1/10	0.1	-1.950
	0.05	-2.012
	0.01	-2.070
	0.001	-2.08
	0.2	-1.840
<i>p</i> = 1/9	0.1	-1.934
	0.05	-1.995
	0.01	-2.050
	0.001	-2.075
	0.2	-1.670
<i>p</i> = 1/5	0.1	-1.789
	0.05	-1.870
	0.01	-1.960
	0.001	-1.99
	0.2	-1.594
<i>p</i> = 1/4	0.1	-1.710
	0.05	-1.805
	0.01	-1.911
	0.001	-1.944
	0.2	-1.525
<i>p</i> = 3/10	0.1	-1.653
	0.05	-1.747
	0.01	-1.861
	0.001	-1.900
	0.2	-1.408
<i>p</i> = 2/5	0.1	-1.539
	0.05	-1.640
	0.01	-1.770
	0.001	-1.816
	0.2	-1.390
<i>p</i> = 1/2	0.1	-1.500
	0.05	-1.600
	0.01	-1.690
	0.001	-1.75

Continued

	0.2	-1.263
	0.1	-1.389
<i>p</i> = 3/5	0.05	-1.492
	0.01	-1.636
	0.001	-1.693
	0.2	-1.265
	0.1	-1.380
<i>p</i> = 7/10	0.05	-1.481
	0.01	-1.626
	0.001	-1.685
	0.2	-1.307
	0.1	-1.416
<i>p</i> = 3/4	0.05	-1.512
	0.01	-1.655
	0.001	-1.716
	0.2	-1.406
	0.1	-1.503
<i>p</i> = 4/5	0.05	-1.594
	0.01	-1.734
	0.001	-1.794
	0.2	-1.969
	0.1	-2.030
<i>p</i> = 8/9	0.05	-2.102
	0.01	-2.226
	0.001	-2.284
	0.2	-2.130
	0.1	-2.184
<i>p</i> = 9/10	0.05	-2.251
	0.01	-2.370
	0.001	-2.420
	0.2	-3.376
	0.1	-3.879
<i>p</i> = 95/100	0.05	-3.913
	0.01	-4.002
	0.001	-4.052
	0.2	-8.513
	0.1	-8.513
<i>p</i> = 99/100	0.05	-8.513
	0.01	-8.532
	0.001	-8.050

Remark 3.4. In **Table 3**, for values of $p \geq \frac{95}{100}$ the speed with which the stability index has to infinity is greater; so it does not allow to obtain $n = 100$ values of $\alpha = 0.5, 0.505, \dots, 0.995$. Thus, for example the results shown in **Table 3** for $p = \frac{95}{100}$ were obtained with $n = 89$ data, while for $p = \frac{99}{100}$, they were performed with $n = 50$ data. The results presented for the rest of the p values in **Table 3** were obtained with 100 data.

Clearly, the results presented for each p value in **Table 3**, must be interpreted in the same way that the results in **Table 1** were previously explained.

Discussion of results. The motivation for studying discounted reward (cost) problems is primarily economic. Capital accumulation processes of an economy, inventory problems, inventory management, portfolio management, are applications of this type of optimization criteria. The reward function used in this work, see Equation (16), is a very used function in economics, it belongs to the family of consumer utility functions, specifically the so-called Cobb-Douglas utility function (see [19] for definitions), so the selection of the parameter p in Equation (16) it must be very careful. The results obtained in this work about the asymptotic evaluations of the stability index (which are presented in **Table 3**) are interpreted as follows:

1) If $p \rightarrow 1: \Delta_{\mathcal{R}_\alpha} \sim \mathcal{M}(1-\alpha)^{-\infty}$. That is, if the parameter p of the reward function approaches 1, then the sensitivity of the stability index grows indefinitely.

Therefore, for values of $p \approx 1$ the use of an approximate policy to control the original process is not recommended, that is because the results show (see **Table 3**) that for $p \geq \frac{8}{9}$ we have $\hat{\kappa} \leq -2.284$, which is why the stability index can be up to $\mathcal{M}(1-\alpha)^{-8.05}$.

2) If $p \rightarrow \frac{1}{2}: \Delta_{\mathcal{R}_\alpha} \sim \mathcal{M}(1-\alpha)^{-1.75}$. In this case, the results obtained (see **Table 3**) suggest that if values of $\frac{2}{5} \leq p \leq \frac{7}{10}$ are selected in the reward function given in Equation (16), then it would seem reasonable to use the approximate policy \tilde{f}_* to control the original process \mathcal{M} .

3) If $p \rightarrow 0: \Delta_{\mathcal{R}_\alpha} \sim \mathcal{M}(1-\alpha)^{-2}$. In this case, the results obtained in this work using statistical techniques are the same as those found in articles [6] and [7], but using upper bounds such as given in Equation (1).

Remember that by definition we have $0 < p < 1$ (see Equation (16)). Now, from the three previous points, the results obtained show that for extreme values of p (close to zero or one) it is not recommended to use an approximate policy to control the original process. The results obtained suggest a selection of the p value close to the average ($p \approx \frac{1}{2}$) in the reward function, to use of such an approximation.

4. Conclusion

Despite the extensive literature that exists on the subject of Markov control process, there are few works developed on the subject of estimating the stability index. The study of stability for control processes represents a challenge, both from a theoretical and an applied point of view. In this application work, it is intended to contribute to the study of stability using statistical techniques instead of probabilistic metrics. The limitations of this work are the use of a simple Markov control process as well as the use of an exponential distribution function to measure the shock effect of the process. However, the numerical estimates found are consistent and show their impact on the sensitivity of the stability index to changes in both the discount factor and the parameter in the reward function; obviously, the results obtained respond favorably to the original question, which was posed in the introduction, so we can conclude that the objective of this work was achieved. Finally, it is recommended to strengthen the results found in this work by carrying out some of the following future investigations: 1) Using more complex Markov control processes; 2) Validate the robustness of the results using another type of distribution function to measure the shock effect of the process; 3) Use another type of reward function; and 4) Use of other statistical techniques for the asymptotic estimation of the stability index.

Acknowledgements

The author wishes to thank referees for valuable suggestions on improvement of the previous version of the paper.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Dynkin, E.B. and Yushkevich, A.A. (1979) *Controlled Markov Processes*. Springer-Verlag, New York.
- [2] Hernandez-Lerma, O. (1989) *Adaptive Markov Control Process*. Vol. 79, Springer-Verlag, New York. <https://doi.org/10.1007/978-1-4419-8714-3>
- [3] Gordienko, E.I. (1992) An Estimate of the Stability of Optimal Control of Certain Stochastic and Deterministic Systems. *Journal of Soviet Mathematics*, **59**, 891-899. <https://doi.org/10.1007/BF01099115>
- [4] Gordienko, E.I. and Salem, F.S. (1998) Robustness Inequalities for Markov Control Processes with Unbounded Cost. *Systems & Control Letters*, **33**, 125-130. [https://doi.org/10.1016/S0167-6911\(97\)00077-7](https://doi.org/10.1016/S0167-6911(97)00077-7)
- [5] Gordienko, E.I. and Yushkevich, A.A. (2003) Stability Estimates in the Problem of Average Optimal Switching of a Markov Chain. *Mathematical Methods of Operations Research*, **57**, 345-365. <https://doi.org/10.1007/s001860200258>
- [6] Gordienko, E.I., Lemus-Rodriguez, E. and Montes-de-Oca, R. (2008) Discounted Cost Optimality Problem: Stability with Respect to Weak Metrics. *Mathematical*

- Methods of Operations Research*, **68**, 77-96.
<https://doi.org/10.1007/s00186-007-0171-z>
- [7] Gordienko, E., Martínez, J. and Ruiz de Chávez, J. (2015) Stability Estimation of Transient Markov Decision Processes. In: Mena, R.H., Pardo, J.C., Rivero, V. and Bravo, G.U., Eds., *XI Symposium on Probability and Stochastic Processes*, Mexico, 18-22 November 2013, 157-176. https://doi.org/10.1007/978-3-319-13984-5_8
- [8] Martínez-Sánchez, J.E. (2020) Stability Estimation for Markov Control Processes with Discounted Cost. *Applied Mathematics*, **11**, 491-509.
<https://doi.org/10.4236/am.2020.116036>
- [9] Gordienko, E.I. and Salem-Silva, F. (2000). Estimates of Stability of Markov Control Processes with Unbounded Costs. *Kybernetika*, **36**, 195-210.
- [10] Montes-de-Oca, R. and Salem-Silva, F. (2005) Estimates for Perturbations of Average Markov Decision Process with a Minimal State and Upper Bounded by Stochastically Ordered Markov Chains. *Kybernetika*, **41**, 757-772.
- [11] Martinez, J. and Zaitzeva, E. (2015) Note on Stability Estimation in Average Markov Control Processes. *Kybernetika*, **51**, 629-638.
<http://doi.org/10.14736/kyb-2015-4-0629>
- [12] Rachev, S.T. (1991) *Probability Metrics and the Stability of Stochastic Models*. Wiley, Chichester.
- [13] Hernandez-Lerma, O. and Lasserre, J. (1996) *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Springer, New York.
<https://doi.org/10.1007/978-1-4612-0729-0>
- [14] Hernandez-Lerma, O. and Lasserre, J.B. (1999) *Further Topics on Discrete-Time Markov Control Processes*. Springer, New York.
<https://doi.org/10.1007/978-1-4612-0561-6>
- [15] Van Nunen, J.A. and Wessels, J. (1978) Note—A Note on Dynamic Programming with Unbounded Rewards. *Management Science*, **24**, 485-586.
<https://doi.org/10.1287/mnsc.24.5.576>
- [16] Carlton, D. and Perloff, J. (2005) *Modern Industrial Organization*. Pearson, Addison Wesley, Boston.
- [17] Viscusi, W., Harrington, J. and Vernon, J. (2005) *Economics of Regulation and Antitrust*. The MIT Press, Cambridge.
- [18] Kmenta, J. (1971) *Elements of Econometrics*. 2nd Edition. Macmillan Publishing Company, New York.
- [19] Cobb, C.W. and Douglas, P.H. (1928) A Theory of Production. *American Economic Review*, **18**, 139-165.