Dyn Games Appl (2011) 1:253–279
DOI 10.1007/s13235-011-0013-8

Stochastic Games with Unbounded Payoffs: Applications to Robust Control in Economics

Anna Jaśkiewicz · Andrzej S. Nowak

Published online: 16 April 2011

Abstract We study a discounted maxmin control problem with general state space. The controller is unsure about his model in the sense that he also considers a class of approximate models as possibly true. The objective is to choose a maxmin strategy that will work under a range of different model specifications. This is done by dynamic programming techniques. Under relatively weak conditions, we show that there is a solution to the optimality equation for the maxmin control problem as well as an optimal strategy for the controller. These results are applied to the theory of optimal growth and the Hansen–Sargent robust control model in macroeconomics. We also study a class of zero-sum discounted stochastic games with unbounded payoffs and simultaneous moves and give a brief overview of recent results on stochastic games with weakly continuous transitions and the limiting average payoffs.

Keywords Zero-sum stochastic games · Robust control · Optimal growth theory · Macroeconomic dynamics

1 Introduction

This paper is devoted to zero-sum discrete-time stochastic games and maxmin control models with Borel state spaces and weakly continuous transition probabilities. We are mainly interested in the discounted payoff criterion. In the classical papers on stochastic dynamic

A. Jaśkiewicz
Institute of Mathematics and Computer Science, Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: anna.jaskiewicz@pwr.wroc.pl

A.S. Nowak (✉)
Faculty of Mathematics, Computer Science and Econometrics, University of Zielona Góra, Podgórna 50, 65-246 Zielona Góra, Poland
e-mail: a.nowak@wmie.uz.zgora.pl

A.S. Nowak
Institute of Finance, PWSZ Nysa, Poland

programming (Markov decision processes), it is assumed that the returns are bounded or bounded from one side; see, for example, [4, 6, 47, 49, 52]. Unbounded returns, on the other hand, are very common in economic models, particularly, in the theory of optimal growth; see [1, 5, 9, 10, 12, 25, 26, 31, 48], and their references. Most of the aforementioned works use the weighted supremum norm approach introduced by Wessels [53, 54]. However, a new idea has been recently presented by Rincón-Zapatero and Rodriguez-Palmero [48] within a deterministic framework. It involves local contraction methods and applies one-sided majorant functions. Related results on stochastic dynamic programming the reader can find in [24, 36, 37]. The weighted supremum norm approach was also applied to zero-sum discounted stochastic games with unbounded payoffs and weakly continuous transitions by Couwenbergh [8], Jaśkiewicz and Nowak [23] and Küenle [29].

In this paper, we first study discounted maxmin control models that can also be called robust dynamic programming problems. The literature on these topics is pretty large. However, for models with Borel state spaces, we should mention such works as [12, 20, 28, 30, 46] or [13, 14]. An important contribution to the deterministic case is the book of Başar and Bernhard [2]. In contrast to the standard control model, in which there is a single decision-maker, in a maxmin problem there are two players, namely, the controller himself and an opponent called nature. The controller's objective is to find a maxmin strategy that guarantees the best performance in the worst possible situation. On the other hand, maxmin control models can be viewed as a special class of two-person dynamic games, in which the opponent receives information about the current state and the actions chosen by the controller in every state. In this paper, we propose a general theory for stochastic systems based on the *one-sided majorant function approach*. The weighted supremum norm approach is based upon the condition that the absolute value of the controller's payoff u is bounded by some function ω. We relax this requirement and assume that the only positive part of u is bounded by ω. This fact, in turn, allows us to deal with the utilities whose values may equal $-\infty$ for certain states and actions. For instance, one can think of logarithmic functions which are commonly used in economics (see Example 4). Moreover, our method in contrast to the weighted supremum norm approach does not usually impose additional constraints on values of discount coefficients (see Example 6). In order to show that the controller possesses an optimal (maxmin) strategy, we apply the dynamic programming techniques. In fact, we prove that there exists a solution V to the optimality equation, which can be obtained as a limit of solutions V_k for the truncated models. More precisely, V_k's are solutions to the optimality equations for models with the utilities $u_k = \max\{-k, u\}$, for $k = 1, 2, \ldots$. Then an optimal strategy for the controller is derived from the optimality equation involving the unbounded solution V.

A common feature of the maxmin control problems is the fact that they help the controller to design a robust strategy. The controller does not know the probability model generating the data. He considers instead a range of models and makes decisions that maximizes utility given in the worst possible case. For instance, in Sect. 4, we consider the usual single-controller system:

$$x_{t+1} = \Psi(x_t, a_t, \xi_t), \tag{1}$$

where x_t is a state, a_t is the action chosen by the controller and ξ_t's are independent random variables called disturbances. Then a nature's strategy is a sequence $\{b_t\}$, where each b_t is a distribution for ξ_t, $t = 0, 1, \ldots$. It turns out that the state evolution for a number of models from classical growth theory can be described with the aid of (1), see Examples 2–5. Another application for linear Gaussian systems is presented in Sect. 5. We focus there on recently developed robust control techniques in macroeconomics by Hansen and Sargent [13, 14]. Namely, the controller considers a set of alternative models which are "close" to the baseline model, where distance between the models is measured by a mutual entropy. We show that our approach can be successfully applied for such systems and our assumptions allow for consideration of a pretty large class of utility functions (not only negative quadratic functions like in [13, 14] and related papers).

The rest of the paper is devoted to zero-sum stochastic games with simultaneous moves (standard games á la Shapley [50]). We allow for unbounded payoffs and consider weakly continuous transitions. It is proved that under some regularity conditions the maximizer has optimal stationary strategy whereas the minimizer has an ϵ-optimal semi-stationary one for any $\epsilon > 0$. The last section describes some recent result for zero-sum games with the expected limiting average payoff criterion under some ergodicity conditions.

2 Preliminaries

We start with some preliminaries. Let R be the set of all real numbers, $\underline{R} = R \cup \{-\infty\}$. Let $N = \{1, 2, \ldots\}$ and $N_0 = N \cup \{0\}$. By a Borel space, we mean a non-empty Borel subset of a complete separable metric space endowed with the Borel σ-algebra $\mathcal{B}(Y)$. For any Borel space Y, by $C(Y)$ we denote the space of all bounded continuous real-valued functions on Y and we use $P(Y)$ to describe the space of all probability measures on Y endowed with the weak topology and the Borel σ-algebra, see [45]. Recall that a sequence $\{p_n\}$ converges weakly to some $p \in P(Y)$ if for any $\phi \in C(Y)$,

$$\int_Y \phi(y) p_n(dy) \to \int_Y \phi(y) p(dy) \quad \text{as } n \to \infty.$$

If Y is a Borel space, then $P(Y)$ is a Borel space, too, and if Y is compact, so is $P(Y)$; see Corollary 7.25.1 and Proposition 7.22 in [4]. Further, let us assume in this section that X and Y are Borel spaces. A transition probability or a stochastic kernel from X to Y is a function $\varphi : \mathcal{B}(Y) \times X \mapsto [0, 1]$ such that $\varphi(D|\cdot)$ is a Borel measurable function on X for every $D \in \mathcal{B}(Y)$ and $\varphi(\cdot|x) \in P(Y)$ for each $x \in X$. It is well known that every Borel measurable mapping $f : X \mapsto P(Y)$ induces a transition probability φ from X to Y. Namely, $\varphi(D|x) = f(x)(D)$, $D \in \mathcal{B}(Y)$, $x \in X$; see Proposition 7.26 in [4]. We shall write $f(dy|x)$ instead of $f(x)(dy)$. Clearly, any Borel measurable mapping $f : X \mapsto Y$ is a special transition probability φ from X to Y such that for each $x \in X$, $\varphi(\cdot|x)$ is the Dirac measure concentrated at the point $f(x)$.

A set-valued mapping $x \mapsto \Phi(x) \subset Y$ is *upper semi-continuous* (*lower semi-continuous*) if the set $\{x \in X : \Phi(x) \cap C \neq \emptyset\}$ is closed (open) for each closed (open) set $C \subset Y$. Φ is *continuous* if it is both lower and upper semi-continuous. Assume that $\Phi(x) \neq \emptyset$ for every $x \in X$. It is well-known that if Φ is compact-valued and upper semi-continuous, then by Brown and Purves [7] Φ admits a measurable selector, that is, there exists a Borel measurable mapping $g : X \mapsto Y$ such that $g(x) \in \Phi(x)$ for each $x \in X$. Moreover, the same holds if Φ is lower semi-continuous and has complete values $\Phi(x)$ for all $x \in X$, see [27].

Consider the set-valued mapping $x \mapsto P(\Phi(x)) \subset P(Y)$. The following result from [19] will be useful.

Lemma 1 *If $x \mapsto \Phi(x)$ is upper (lower) semi-continuous and compact-valued, then so is $x \mapsto P(\Phi(x))$.*

In the sequel, we shall be interested in the set-valued mappings satisfying one of the following assumptions.

(P1) The set-valued mapping $x \mapsto \Phi(x)$ is lower semi-continuous and $\Phi(x)$ is σ-compact for each $x \in X$.
(P2) Y is a complete separable metric space, the set-valued mapping $x \mapsto \Phi(x)$ is lower semicontinuous, and $\Phi(x)$ is closed for each $x \in X$.

Lemma 2 *If either* (P1) *or* (P2) *holds, then there exists a sequence of Borel measurable mappings* $g_n : X \mapsto P(Y)$, $n \in N$, *such that* $\{g_n(x) : n \in N\}$ *is dense in* $P(\Phi(x))$ *for any* $x \in X$.

Proof By condition (P1) there exists a sequence of Borel measurable mappings $\phi_n : X \mapsto Y$, $n \in N$, such that $\{\phi_n(x) : n \in N\}$ is dense in $\Phi(x)$ for any $x \in X$; see Lemma 2 in [44]. The same conclusion can be drawn under (P2) from Theorem 5.6 in [18]. The remaining part of the proof is similar to that of Lemma 2.1 in [43]. □

3 Discounted Stochastic Maxmin Control Problems

A discrete-time discounted *maxmin control model* is defined by the objects: X, A, B, K_A, K, u, q, and β satisfying the following assumptions:

(A1) X is a Borel *state space*;
(A2) A is the *action space* of the *controller* (player 1), B is the action space of the *opponent* (player 2). It is assumed that A and B are Borel spaces;
(A3) $K_A \in \mathcal{B}(X \times A)$ is the *constraint set* for the *controller*. It is assumed that

$$A(x) := \big\{a \in A : (x, a) \in A\big\} \neq \emptyset$$

for each $x \in X$. This is the *set of admissible actions* of the *controller* in the state $x \in X$;
(A4) $K \in \mathcal{B}(X \times A \times B)$ is the *constraint set* for the *opponent*. It is assumed that

$$B(x, a) := \big\{b \in B : (x, a, b) \in B\big\} \neq \emptyset$$

for each $(x, a) \in K_A$. This is the *set of admissible actions* of the *opponent* for each $x \in X$, $a \in A(x)$;
(A5) $u : K \to \underline{R}$ is a Borel measurable *stage payoff function*;
(A6) q is a transition probability from K to X, called the *law of motion* among states. If x_t is a state at the beginning of period t of the process and actions $a_t \in A(x_t)$ and $b_t \in B(x_t, a_t)$ are selected by the players, then $q(\cdot|x_t, a_t, b_t)$ is the probability distribution of the next state x_{t+1};
(A7) $\beta \in (0, 1)$ is called the *discount factor*.

We now make basic assumptions on the sets K_A and K:

(C1) For any $x \in X$, $A(x)$ is compact and the set-valued mapping $x \mapsto A(x)$ is upper semi-continuous;
(C2) The set-valued mapping $(x, a) \mapsto B(x, a)$ is lower semi-continuous;
(C3) There exists a Borel measurable mapping $g : K_A \mapsto B$ such that $g(x, a) \in B(x, a)$ for all $(x, a) \in K_A$.

Remark 1 Condition (C3) holds if $B(x, a)$ is σ-compact for each $(x, a) \in K_A$, see [7], or if B is a complete separable metric space and each set $B(x, a)$ is closed; see [27].

Let $H_0 := X$, $H_n := K^n \times X$ for $n \in N$. Put $H_0^* := K_A$ and $H_n^* := K^n \times K_A$ if $n \in N$. Generic elements of H_n and H_n^* are *histories* of the process of the form $h_0 = x_0$, $h_0^* = (x_0, a_0)$ and for each $n \in N$, $h_n = (x_0, a_0, b_0, \dots, x_{n-1}, a_{n-1}, b_{n-1}, x_n)$, $h_n^* = (h_n, a_n)$, respectively.

A *strategy* for the *controller* is a sequence $\pi = \{\pi_n\}_{n\in N_0}$ of stochastic kernels π_n from H_n to A such that $\pi_n(A(x_n)|h_n) = 1$ for each $h_n \in H_n$. The *class of all strategies* for the *controller* will be denoted by Π. A *strategy* for the *opponent* is a sequence $\gamma = \{\gamma_n\}_{n\in N_0}$ of stochastic kernels γ_n from H_n^* to B such that $\gamma_n(B(x_n, a_n)|h_n^*) = 1$ for all $h_n^* \in H_n^*$. The *class of all strategies* for the *opponent* will be denoted by Γ^*. Let F be the set of Borel measurable mappings f from X to A such that $f(x) \in A(x)$ for each $x \in X$. A *deterministic stationary strategy* for the *controller* is a sequence $\pi = \{f_n\}_{n\in N_0}$ where $f_n = f$ for all $n \in N_0$ and some $f \in F$. Such a strategy can obviously be identified with the mapping $f \in F$. Put

$$u^+(x,a,b) := \max\{u(x,a,b), 0\} \quad \text{and}$$
$$u^-(x,a,b) := \min\{u(x,a,b), 0\}, \quad (x,a,b) \in K.$$

For each initial state $x_1 = x$ and any strategies $\pi \in \Pi$ and $\gamma \in \Gamma^*$, define

$$J^+(x,\pi,\gamma) = E_x^{\pi\gamma}\left(\sum_{t=0}^{\infty} \beta^t u^+(x_t, a_t, b_t)\right) \quad \text{and}$$
$$J^-(x,\pi,\gamma) = E_x^{\pi\gamma}\left(\sum_{t=0}^{\infty} \beta^t u^-(x_t, a_t, b_t)\right).$$

Here, $E_x^{\pi\gamma}$ denotes the expectation operator corresponding to the unique conditional probability measure $P_x^{\pi\gamma}$ defined on the space of histories, starting at state x and endowed with the product σ-algebra, which is induced by strategies π, γ and the transition probability q according to the Ionescu–Tulcea theorem; see Proposition V.1.1 in [41]. In the sequel, we give conditions under which $J^+(x,\pi,\gamma) < \infty$ for any $x \in X$, $\pi \in \Pi$, $\gamma \in \Gamma^*$. They enable us to define the *expected discounted payoff* over an infinite time horizon as follows:

$$J(x,\pi,\gamma) = E_x^{\pi\gamma}\left(\sum_{t=0}^{\infty} \beta^t u(x_t, a_t, b_t)\right). \tag{2}$$

Then, for every $x \in X$ and $\pi \in \Pi$, $\gamma \in \Gamma^*$, $J(x,\pi,\gamma) \in \underline{R}$ and

$$J(x,\pi,\gamma) = J^+(x,\pi,\gamma) + J^-(x,\pi,\gamma) = \sum_{t=0}^{\infty} \beta^t E_x^{\pi\gamma} u(x_t, a_t, b_t).$$

Let

$$v_*(x) := \sup_{\pi\in\Pi} \inf_{\gamma\in\Gamma} J(x,\pi,\gamma), \quad x \in X.$$

This is the *maxmin or lower value* of the game starting at the state $x \in X$. A strategy $\pi^* \in \Pi$ is called *optimal* for the *controller* if

$$\inf_{\gamma\in\Gamma} J(x,\pi^*,\gamma) = v_*(x)$$

for every $x \in X$.

It is worth mentioning that if u is unbounded, then an optimal strategy π^* need not exist even if $0 \le v_*(x) < \infty$ for every $x \in X$ and the available action sets $A(x)$ and $B(x)$ are finite.

Example 1 Let $X = N_0$, $A(0) = \{0\}$, $A(x) = \{0, 1\}$ for $x \in N$ and $B(x) = \{0\}$ for all $x \in X$. Thus, player 2 is dummy. The discounted payoff function is denoted by $J(x, \pi)$. Assume that $0 \in X$ is an *absorbing state* with zero payoffs. If $x \in N$, then $u(x, a) = 0$ for $a = 0$ and $u(x, a) = \frac{x-1}{x\beta^{x-1}}$ for $a = 1$. The transitions are as follows: if $x \in N$, then $q(\{x + 1\}|x, a) = 1$ for $a = 0$ and $q(\{0\}|x, a) = 1$ for $a = 1$. Suppose that $x = 1$ is the initial state and the controller chooses $a = 0$ in states $1, 2, \dots, k$ and $a = 1$ in state $k + 1$. The game moves from $x = k + 1$ to the absorbing state and the discounted payoff is $1 - \frac{1}{k+1}$. Hence, it follows that $v_*(1) = \sup_{\pi\in\Pi} J(1, \pi) = 1$. It is obvious that there does not exist an optimal strategy.[1]

The maxmin control problems with Borel state spaces have been already considered by González-Trejo et al. [12], Iyengar [20], Kurano [30], Küenle [28] and are sometimes referred to as *games against nature* or *robust dynamic programming (Markov decision) models*.

We now describe our regularity assumptions imposed on the payoff and transition probability functions.

(W) The payoff function $u : K \mapsto \underline{R}$ is upper semi-continuous. Moreover, for any $v \in C(X)$,

$$(x, a, b) \mapsto \int_X v(y) q\,(dy|x, a, b)$$

is continuous on K.

(M1) There exist a continuous function $\omega : X \mapsto [1, \infty)$ and a constant $\alpha > 0$ such that

$$\sup_{(x,a,b)\in K} \frac{\int_X \omega(y) q\,(dy|x, a, b)}{\omega(x)} \le \alpha \quad \text{and} \quad \beta\alpha < 1. \tag{3}$$

Moreover, the function $(x, a, b) \mapsto \int_X \omega(y) q\,(dy|x, a, b)$ is continuous.

(M2) There exists a constant $d > 0$ such that

$$\sup_{a\in A(x)} \sup_{b\in B(x,a)} u^+(x, a, b) \le d\omega(x) \tag{4}$$

for all $x \in X$.

Note that under conditions (3) and (4) the discounted payoff function is well defined, since

$$0 \le E_x^{\pi\gamma}\left(\sum_{t=0}^{\infty} \beta^t u^+(x_t, a_t, b_t)\right) \le d \sum_{t=0}^{\infty} \beta^t \alpha^t < \infty.$$

For any function $v : X \mapsto R$ define the ω-norm as:

$$\|v\|_\omega = \sup_{x\in X} \frac{|v(x)|}{\omega(x)},$$

[1] A strategy $\pi^* \in \Pi$ is called ϵ-optimal ($\epsilon > 0$) for the controller if $\inf_{\gamma\in\Gamma^*} J(x, \pi^*, \gamma) \ge v_*(x) - \epsilon$ for each $x \in X$. It is easy to see that in Example 1 we have $w_*(x) = 1/\beta^{x-1}$ for $x \in N$ and the controller does not have any ϵ-optimal stationary strategy.

provided that it is finite. Let $U_\omega(X)$ be the space of all upper semi-continuous functions endowed with the metric induced by the ω-norm. By $\underline{U}_\omega(X)$ we denote the set of all upper semi-continuous functions $v : X \mapsto \underline{R}$ such that $v^+ \in U_\omega(X)$.

The proof of the following lemma is similar to that of Lemma 8.5.5 in [16].

Lemma 3 *Assume* (W) *and* (M1) *and that* $v \in \underline{U}_\omega(X)$. *Then the function* $(x, a, b) \mapsto \int_X v(y)q(dy|x, a, b)$ *is upper semi-continuous on* K.

Define $u_k := \max\{u, -k\}$, $k \in N$. For any $v \in \underline{U}_\omega(X)$, $(x, a, b) \in K$ and $k \in N$, put

$$L_k v(x, a, b) := u_k(x, a, b) + \beta \int_X v(y)q(dy|x, a, b)$$

and

$$Lv(x, a, b) := u(x, a, b) + \beta \int_X v(y)q(dy|x, a, b).$$

By the maximum theorems of Berge [3] (see also Proposition 10.2 in [49]) and Lemma 3, we obtain the following auxiliary result.

Lemma 4 *Assume* (A1)–(A7), (C1)–(C2), (W), *and* (M1)–(M2). *Then for any* $v \in \underline{U}_\omega(X)$ *the functions*

$$\inf_{b\in B(x,a)} L_k v(x, a, b) \quad \text{and} \quad \max_{a\in A(x)} \inf_{b\in B(x,a)} L_k v(x, a, b)$$

are upper semi-continuous on K_A *and* X, *respectively. Similar properties hold if* $L_k v(x, a, b)$ *above is replaced by* $Lv(x, a, b)$.

For any $x \in X$, define

$$T_k^* v(x) := \max_{a\in A(x)} \inf_{b\in B(x,a)} L_k v(x, a, b) \quad \text{and} \quad T^* v(x) := \max_{a\in A(x)} \inf_{b\in B(x,a)} Lv(x, a, b). \tag{5}$$

By Lemma 4, the operators T_k^* and T^* are well defined. Note that

$$T^* v(x) := \max_{a\in A(x)} \inf_{\rho\in P(B(x,a))} \int_B Lv(x, a, b)\rho(db). \tag{6}$$

Using Proposition 10.1 in [49], one can prove the following auxiliary result.

Lemma 5

(i) *Let* $\{w_k\}$ *be a non-increasing sequence of functions* $w_k \in \underline{U}_\omega(X)$, *then* $w_\infty = \lim_{k\to\infty} w_k$ *exists and* $w_\infty \in \underline{U}_\omega(X)$.

(ii) *Let* Y *be a compact metric space,* Z *a non-empty set. Assume that* $\{h_k\}$ *is a non-increasing and bounded from above sequence of functions* $h_k : Y \times Z \mapsto R$ *such that* $h_k(\cdot, z)$ *is upper semi-continuous on* Y *for every* k *and* $z \in Z$. *Then*

$$\max_{a\in Y} \inf_{z\in Z} \lim_{k\to\infty} h_k(a, z) = \lim_{k\to\infty} \max_{a\in Y} \inf_{z\in Z} h_k(a, z).$$

We can now state our first main result.

Theorem 1 *Assume* (A1)–(A7), (C1)–(C2), (W), *and* (M1)–(M2). *Then there exist a function $V \in \underline{U}_\omega$ and a stationary strategy $f^* \in F$ such that*

$$T^*V(x) = V(x) = \inf_{b\in B(x,a)} LV\bigl(x, f^*(x), b\bigr)$$

for $x \in X$. Moreover,

$$V(x) = v_*(x) = \inf_{\gamma\in\Gamma^*} J\bigl(x, f^*, \gamma\bigr) = \sup_{\pi\in\Pi}\inf_{\gamma\in\Gamma^*} J(x, \pi, \gamma)$$

for all $x \in X$, so f^* is an optimal stationary strategy for the controller.

Proof Observe that

$$\bigl|u_k(x, a, b)\bigr| \le (k+d)\omega(x) \quad \text{and} \quad \sup_{a\in A(x)}\sup_{b\in B(x,a)} u_k^+(x, a, b) \le d\omega(x).$$

Let $v \in U_\omega(X)$. By Lemma 4, T_k^* maps the space $U_\omega(X)$ into itself. Clearly, $U_\omega(X)$ is a complete metric space. Using the Banach contraction mapping theorem, we infer that there exists a function $V_k \in U_\omega(X)$ such that $V_k = T_k^* V_k$. Proceeding along the same lines as in the proof of Theorem 4.2 in [12], we infer that V_k is an optimal discounted return in the truncated model, i.e.,

$$V_k(x) = \sup_{\pi\in\Pi}\inf_{\gamma\in\Gamma^*} J_k(x, \pi, \gamma), \quad x \in X. \tag{7}$$

Here,

$$J_k(x, \pi, \gamma) = E_x^{\pi\gamma}\left(\sum_{t=0}^{\infty} \beta^t u_k(x_t, a_t, b_t)\right).$$

Observe next that $V_k^+(x) \le V_1^+(x) \le \|V_1^+\|_\omega \omega(x)$ for all $x \in X$. Thus, by virtue of Lemma 5(i), there exists $V(x) := \lim_{k\to\infty} V_k(x)$ for each $x \in X$ and $V \in \underline{U}_\omega(X)$. Moreover, by the monotone convergence theorem, it follows that

$$\int_X V_k(y)q\,(dy|x, a, b) \searrow \int_X V(y)q(dy|x, a, b) \quad \text{for all } (x, a, b) \in K.$$

Note that for every $a \in A(x)$ and $b \in B(x, a)$ one has

$$L_k V_k(x, a, b) \le \left(d\omega(x) + \beta\int_X V_1^+(y)q\,(dy|x, a)\right) \le \bigl(d + \beta\alpha\bigl\|V_1^+\bigr\|_\omega\bigr)\omega(x)$$

for all $x \in X$. Then, from Lemma 5(ii) and the above discussion, we obtain that

$$\begin{aligned}\lim_{k\to\infty}\max_{a\in A(x)}\inf_{b\in B(x,a)} L_k V_k(x, a, b) &= \max_{a\in A(x)}\inf_{b\in B(x,a)}\lim_{k\to\infty} L_k V_k(x, a, b)\\ &= \max_{a\in A(x)}\inf_{b\in B(x,a)} LV(x, a, b),\end{aligned}$$

for every $x \in X$. Using Lemma 4 and Corollary 1 from [7], we claim that there exists a stationary strategy $f^* \in F$ such that $V(x) = \inf_{b\in B(x,a)} LV(x, f^*(x), b)$, $x \in X$. This and (6) imply that

$$V(x) \le \int_B Lv\bigl(x, f^*(x), b\bigr)\rho\,(db)$$

for any $\rho \in P(B(x, f^*(x)))$. Using this inequality and a simple iteration argument, we obtain

$$V(x) \le E_x^{f^*\gamma}\left(\sum_{t=0}^{n} \beta^t u(x_t, a_t, b_t)\right) + \beta^{n+1} E_x^{f^*\gamma} V(x_{n+1}), \quad x \in X. \tag{8}$$

But

$$\begin{aligned}\beta^{n+1} E_x^{f^*\gamma} V(x_{n+1}) \le \beta^{n+1} E_x^{f^*\gamma} V_1^+(x_{n+1}) &\le \beta^{n+1} \left\| V_1^+ \right\| E_x^{f^*\gamma} \omega(x_{n+1}) \\ &\le \beta^{n+1} \left\| V_1^+ \right\| \alpha^{n+1} \omega(x).\end{aligned}$$

From this and assumption (M1), it follows that

$$\limsup_{n\to\infty} \beta^{n+1} E_x^{f^*\gamma} V(x_{n+1}) \le 0.$$

Hence, letting n tend to infinity in (8) and using (M2) we deduce that

$$V(x) \le \inf_{\gamma\in\Gamma^*} J\big(x, f^*, \gamma\big) \quad \text{for all } x \in X. \tag{9}$$

On the other hand, by (7) and definition of u_k, we obtain that

$$V_k(x) = \sup_{\pi\in\Pi} \inf_{\gamma\in\Gamma^*} J_k(x, \pi, \gamma) \ge \sup_{\pi\in\Pi} \inf_{\gamma\in\Gamma^*} J(x, \pi, \gamma), \quad x \in X \text{ and } k \in N.$$

Therefore,

$$V(x) \ge \sup_{\pi\in\Pi} \inf_{\gamma\in\Gamma^*} J(x, \pi, \gamma). \tag{10}$$

Hence, from (9) and (10), we have that

$$V(x) = \inf_{\gamma\in\Gamma^*} J\big(x, f^*, \gamma\big) = \sup_{\pi\in\Pi} \inf_{\gamma\in\Gamma^*} J(x, \pi, \gamma) = v_*(x)$$

for all $x \in X$. □

Remark 2 The weighted supremum norm approach in Markov decision processes was proposed by Wessels [53, 54] and further developed by other researchers like Durán [9, 10], Hernández-Lerma and Lasserre [16, 17] and others. Then this method has been adopted to zero-sum stochastic games, see [8, 12, 21–23, 29] and references therein. The common feature of the aforementioned works is the fact that the authors use the weighted norm condition instead of assumption (M2). More precisely, in our notation, it means that the following holds:

$$\sup_{a\in A(x)} \sup_{b\in B(x,a)} \big|u(x, a, b)\big| \le d\omega(x), \quad x \in X \tag{11}$$

for some constant $d > 0$. However, this assumption excludes many examples studied in economics where the return function u equals $-\infty$ in some states. Moreover, inequality in (M1) and (11) very often enforce additional constraints on the discount coefficient β in comparison with (M1) and (M2), see Example 3 in [24] and Example 6 in Sect. 4.

4 Models with Unknown Disturbance Distributions

Consider the control system in which

$$x_{t+1} = \Psi(x_t, a_t, \xi_t), \quad t \in N_0. \tag{12}$$

It is assumed that $\{\xi_t\}_{t \in N_0}$ is a sequence of independent random variables with values in a Borel space S having unknown probability distributions that can change from period to period. The set B of all possible distributions is assumed to be a non-empty Borel subset of the space $P(S)$ of all probability measures on S endowed with the weak topology. The mapping $\Psi : X \times A \times S \mapsto X$ is assumed to be *continuous*. Let u_0 be an *upper semi-continuous utility function* defined on $K_A \times S$ such that $u_0^+(x, a, s) \le d\omega(x)$ for some constant $d > 0$ and all $(x, a) \in K_A$, $s \in S$.

Using (12), we can formulate a maxmin control model in the sense of Sect. 3, in which:

(a) $B(x, a) = B \subset P(S)$ for each $(x, a) \in K_A$, $K = K_A \times B$,
(b) $u(x, a, b) = \int_S u_0(x, a, s) b(ds)$, $(x, a, b) \in K$,
(c) for any Borel set $D \subset X$, $q(D|x, a, b) = \int_X 1_D(\Psi(x, a, s)) b(ds)$, $(x, a, b) \in K$.

Then for any bounded continuous function $v : X \mapsto R$,

$$\int_X v(y) q(dy|x, a, b) = \int_X v\big(\Psi(x, a, s)\big) b(ds). \tag{13}$$

From Proposition 7.30 in [4] or Lemma 5.3 in [43] and (13), it follows that q is weakly continuous. Moreover, by virtue of Proposition 7.31 in [4], it is easily seen that u is upper semi-continuous on K.

The following result can be viewed as a corollary to Theorem 1.

Proposition 1 *Let Ψ and u_0 satisfy the above assumptions. If* (A1)–(A3) *and* (M1) *hold, then the controller has an optimal strategy.*

Proposition 1 is a counterpart of the results obtained in Sect. 6 of González-Trejo et al. [12] for discounted models (see Propositions 6.1–6.3 and their consequences). However, our assumptions imposed on the primitive data are weaker than the ones used by González-Trejo et al. [12]. They are satisfied for a pretty large number of systems, in which the disturbances comprise “random noises” that are difficult to observe and often caused by external factors influencing the dynamics. Below we give certain examples which stem from economic growth theory and related topics. Particularly, they are inspired by works of Ljungqvist and Sargent [32], Stokey et al. [51] and Jaśkiewicz and Nowak [24], Matkowski and Nowak [37], in which the set B is a singleton. Moreover, in contrast to works [24, 37, 51], we allow random variables ξ_t, $t \in N_0$ to be unbounded.

Example 2 (A growth model with multiplicative shocks) Let $X = [0, \infty)$ be the set of all possible *capital stocks*. If x_t is a capital stock at the beginning of period t, then utility of consumption of $a_t \in A(x_t) := [0, x_t]$ in this period is $U(a_t) = a_t^\sigma$. Here $\sigma \in (0, 1]$ is a fixed parameter. The evolution of the state process is described by the following equation:

$$x_{t+1} = (x_t - a_t)^\theta \xi_t, \quad t \in N_0,$$

where $\theta \in (0, 1)$ is some constant and ξ_t is a *random shock* in period t. Let $\{\xi_t\}$ be independent and let each ξ_t follow a probability distribution $b \in B$ for some Borel set $B \subset P([0, \infty))$. We assume that b is unknown.

Consider the maxmin control model (game against nature) where $X = [0, \infty)$, $A(x) = [0, x]$, $B(x, a) = B$, and $u(x, a, b) = a^\sigma$ for $(x, a, b) \in K$. Then the transition probability q is of the form

$$q(D|x, a, b) = \int_0^\infty 1_D\big((x-a)^\theta s\big) b(ds),$$

where $D \subset X$ is a Borel set. If $v \in C(X)$, then the integral

$$\int_X v(y) q(dy|x, a, b) = \int_0^\infty v\big((x-a)^\theta s\big) b(ds)$$

is continuous at $(x, a, b) \in K$. We further assume that

$$\bar{s} = \sup_{b \in B} \int_0^\infty s b(ds) < \infty.$$

Define now

$$\omega(x) = (r + x)^\sigma, \quad x \in X, \tag{14}$$

where $r \ge 1$ is a constant. Clearly, $u^+(x, a, b) = a^\sigma \le \omega(x)$ for any $(x, a, b) \in K$. Hence, condition (M2) is satisfied. Moreover, by Jensen's inequality

$$\int_X \omega(y) q(dy|x, a, b) = \int_0^\infty \big(r + (x-a)^\theta s\big)^\sigma b(ds) \le \big(r + x^\theta \bar{s}\big)^\sigma.$$

Thus,

$$\frac{\int_X \omega(y) q(dy|x, a, b)}{\omega(x)} \le \eta^\sigma(x), \tag{15}$$

where

$$\eta(x) = \frac{r + \bar{s} x^\theta}{r + x}, \quad x \in X.$$

If $x \ge \bar{x} := \bar{s}^{1/(1-\theta)}$, then $\eta(x) \le 1$, and consequently, $\eta^\sigma(x) \le 1$. If $x < \bar{x}$, then

$$\eta(x) < \frac{r + \bar{s} x^\theta}{r + x} \le \frac{r + \bar{s} \bar{x}^\theta}{r} = 1 + \frac{\bar{x}}{r},$$

and

$$\eta^\sigma(x) \le \alpha := \left(1 + \frac{\bar{x}}{r}\right)^\sigma. \tag{16}$$

Let $\beta \in (0, 1)$ be any discount factor. Then there exists $r \ge 1$ such that $\alpha\beta < 1$, and from (15) and (16) it follows that assumption (M1) is satisfied.

Example 3 (A growth model with additive shocks) Consider the model from Example 1 with the following state evolution equation:

$$x_{t+1} = (1 + \rho_0)(x_t - a_t) + \xi_t, \quad t \in N_0.$$

Here, $\rho_0 > 0$ is a constant *rate of growth* and ξ_t is an additional *random income* (shock) received in period t. The transition probability q is now of the form

$$q(D|x,a,b) = \int_0^\infty 1_D\big((1+\rho_0)(x-a)+s\big)b(ds),$$

where $D \subset X$ is a Borel set. If $v \in C(X)$, then the integral

$$\int_X v(y)q(dy|x,a) = \int_0^\infty v\big((1+\rho_0)(x-a)+s\big)b(ds)$$

is continuous in $(x,a,b) \in K$. Let the function ω be as in (14). Applying Jensen's inequality, we obtain

$$\begin{aligned}\int_X \omega(y)q(dy|x,a,b) &= \int_0^\infty \omega\big((x-a)(1+\rho_0)+s\big)b(ds) \le \omega\big(x(1+\rho_0)+\bar{s}\big)\\ &= \big(r+x(1+\rho_0)+\bar{s}\big)^\sigma.\end{aligned}$$

Thus,

$$\frac{\int_X \omega(y)q(dy|x,a,b)}{\omega(x)} \le \eta_0^\sigma(x),$$

where

$$\eta_0(x) = \frac{r+x(1+\rho_0)+\bar{s}}{r+x}, \quad x \in X.$$

Take $r > \bar{s}/\rho_0$ and note that

$$\lim_{x\to 0+}\eta_0(x) = 1+\frac{\bar{s}}{r} < \lim_{x\to\infty}\eta_0(x) = 1+\rho_0.$$

Hence,

$$\sup_{(x,a,b)\in K}\frac{\int_X \omega(y)q(dy|x,a,b)}{\omega(x)} \le \sup_{x\in X}\eta_0^\sigma(x) = (1+\rho_0)^\sigma.$$

Therefore, condition (M1) holds for all $\beta \in (0,1)$ satisfying the inequality $\beta(1+\rho_0)^\sigma < 1$.

Example 4 Let us consider again the model from Example 1 but with $u(x,a,b) = \ln a$, $a \in A(x) = [0,x]$. This utility function has a number of applications in economics; see [32, 51]. Nonetheless, the two-sided weighted norm approach cannot be employed, because $\ln(0) = -\infty$. Assume now that the state evolution equation is of the form

$$x_{t+1} = (1+\rho_0)(x_t-a_t)\xi_t$$

where ρ_0 and the sequence $\{\xi_t\}_{t\in N_0}$ are as in the previous examples. Let $\omega(x) = r+\ln(1+x)$ for all $x \in X$ and some $r \ge 1$. Clearly, $u^+(x,a,b) = \max\{0,\ln a\} \le \max\{0,\ln x\} \le \omega(x)$ for all $(x,a,b) \in K$. By Jensen's inequality it follows that:

$$\int_X \omega(y)q(dy|x,a,b) = \int_0^\infty \omega\big((x-a)(1+\rho_0)+s\big)b(ds) \le r+\ln\big(1+x(1+\rho_0)\bar{s}\big)$$

for all $(x, a, b) \in K$. Thus,

$$\frac{\int_X \omega(y)q(dy|x,a)}{\omega(x)} \le \psi(x) := \frac{r + \ln(1 + x(1+\rho_0)\bar{s})}{r + \ln(1+x)}. \tag{17}$$

If we assume that $\bar{s}(1+\rho_0) > 1$, then

$$\psi(x) - 1 = \frac{\ln(\frac{1+(1+\rho_0)\bar{s}x}{1+x})}{r + \ln(1+x)} \le \frac{1}{r}\ln\left(\frac{1+(1+\rho_0)\bar{s}x}{1+x}\right) \le \frac{1}{r}\ln\bigl(\bar{s}(1+\rho_0)\bigr).$$

Hence,

$$\psi(x) \le \alpha := 1 + \frac{1}{r}\ln\bigl(\bar{s}(1+\rho_0)\bigr).$$

Choose now any $\beta \in (0, 1)$. If r is sufficiently large, then $\alpha\beta < 1$, and by (17) condition (M1) holds.

Example 5 (A model involving quadratic cost) Assume that $X = R$, $A(x) = [0, \hat{a}]$ for some $\hat{a} > 0$ and all $x \in X$. Let the state process be described by the recursive equation

$$x_{t+1} = x_t + a_t + \xi_t, \quad t \in N_0,$$

where $x_t \in X$, $a_t \in A(x_t)$ and $\{\xi_t\}$ are independent random variables with probability distributions in a Borel set $B \subset P(R)$. We say that $a_t + \xi_t = x_{t+1} - x_t$ is a *change of the capital stock* from x_t to x_{t+1} where part a_t is chosen by the controller. We assume that the utility function u is of the form

$$u(x, a, b) = x - ca^2$$

where $c > 0$ is a fixed constant, $x \in X$, and $a \in A(x)$. Putting $y = x + a$, we notice that $u(x, a, b) = x - c(y - x)^2$. This utility function consists of two terms: a quadratic cost $c(y - x)^2$ of changing the capital stock from x to y and a value x, called a net revenue of player 1, when capital stock is $x \in X$; see pp. 95–96 in [51]. Assume that

$$\tilde{s} := \sup_{b \in B} \int_{-\infty}^{\infty} |s| b(ds) < \infty$$

and define $\omega(x) = r + |x|$ for some $r \ge 1$, $x \in X$. Clearly, condition (M2) holds, since $u^+(x, a, b) \le |x| < \omega(x)$ for all $(x, a, b) \in K$. Observe next that

$$\frac{\int_X \omega(y)q(dy|x,a,b)}{\omega(x)} \le \frac{r + |x| + \hat{a} + \tilde{s}}{r + |x|} \le \alpha := 1 + \frac{\hat{a} + \tilde{s}}{r}. \tag{18}$$

For any $\beta \in (0, 1)$, there exist some $r \ge 1$ such that $\alpha\beta < 1$, and by (18) condition (M1) is satisfied.

Example 6 (Example 7.2 in [12]) Consider the linear system

$$x_{t+1} = x_t + a_t + \xi_t, \quad t \in N_0,$$

with state space $X = R$ and the set of available actions in state x defined as $A(x) = [-|x|, |x|]$. Let $\hat{s} > 0$ and $S = [-\hat{s}, \hat{s}]$. The set $B \subset P(S)$ is the family of truncated

Gaussian distributions on S, with zero mean and standard deviations $\widehat{\sigma}$ in $[\underline{\sigma}, \overline{\sigma}]$, where $0 < \underline{\sigma} < \overline{\sigma} < \infty$. González-Trejo et al. [12] deal with minimax control problem and a non-negative quadratic *cost function* $\int_S (x^2 + a^2 + s^2) b\,(ds)$. We, in turn, consider maxmin model. Thus, our utility function is

$$u(x, a, b) = -\int_S \left(x^2 + a^2 + s^2\right) b\,(ds) = -x^2 - a^2 - \widehat{\sigma}^2.$$

Since $u^+(x, a, b) = 0$ for all $x \in X$, $a \in A(x)$ and $b \in B$, then assumption (M2) is satisfied for the constant function $\omega \equiv 1$. Moreover,

$$\beta \frac{\int_X \omega(y) q\,(dy|x, a, b)}{\omega(x)} = \beta < 1,$$

for any $(x, a, b) \in K$, which implies that (M1) holds for any $\beta \in (0, 1)$. Observe now that the weighted norm method impose a more severe constraint on β than our approach where ω is a majorant function. Indeed, González-Trejo et al. [12] define

$$\omega(x) = \bar{w} e^{\kappa |x|}, \quad x \in X,$$

where κ and $\bar{w}$ are appropriately chosen constants greater than 2. Then they show that $|u(x, a, b)| \le \omega(x)$ for all $(x, a, b) \in K$. From Lemma 7.1 in [12], it follows that condition (M1) is satisfied if

$$\beta < \exp(-\kappa \hat{s}).$$

This implies that, for instance, for $\hat{s} > 1$ the discount coefficient β must be very small. In other words, the controller has an optimal strategy only when the discount factor is small enough. According to our approach, the controller possesses an optimal stationary strategy for any discount factor $\beta \in (0, 1)$.

5 An Application to the Hansen–Sargent Model in Macroeconomics

In this section, we study maxmin control model (two-player zero-sum dynamic game), in which a minimizing player (*a malevolent nature*) chooses an action to minimize the controller's objective function. The aim of the controller is to design a decision rule that is robust to misspecification of a dynamic approximating model linking controls today to state variables tomorrow. The constraint on nature is represented by a cost based on a reference transition probability q. Nature can deviate away from q, but the larger the deviation, the higher the cost. In particular, this cost is proportional to the relative entropy $R(\hat{q}||q)$ between the chosen probability $\hat{q}$ and the reference probability q, i.e., the cost equals to $\theta_0 R(\hat{q}||q)$, where $\theta_0 > 0$. Such preferences in macroeconomics are called *multiplier preferences*; see [13, 14] and [35].

Let us consider the following scalar system:

$$x_{t+1} = x_t + a_t + \varepsilon_t + b_t, \tag{19}$$

where $x_t \in X := R$, $t \in N_0$, $a_t \in A(x_t) \equiv A := [0, \hat{a}]$ is an action selected by the controller and $b_t \in B(x_t, a_t) \equiv B := (-\infty, 0]$ is a parameter chosen by the *malevolent nature*. The sequence of random variables $\{\varepsilon_t\}_{t \in N_0}$ is i.i.d., where ε_t follows the standard Gaussian distribution with the density denoted by ϕ. At each period, the controller selects a control $a \in A$, which incurs the payoff $u_0(x, a)$. It is assumed that the function u_0 is *upper semi-continuous* on $X \times A$.

The controller has a unique explicitly specified approximating model (when $b_t \equiv 0$ for all t) but concedes that data might actually be generated by a number of models that surround the approximating model; see [13–15].

Let $t \in N_0$ be fixed. By p, we denote the conditional density of variable $Y = x_{t+1}$ implied by (19). Setting $a = a_t$, $x = x_t$, and $b_t = b$ we obtain that

$$p(y|x, a, b) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-x-a-b)^2}{2}}, \quad \text{for } y \in R.$$

Clearly, $p(\cdot|x, a, b)$ defines the probability measure q, where

$$q(D|x, a, b) = \int_D p(y|x, a, b)\, dy \quad \text{for } D \subset \mathcal{B}(R).$$

If $b = 0$, then we deal with the baseline model. Hence, the relative entropy

$$R\big(q(\cdot|x, a, b)||q(\cdot|x, a, 0)\big) = \frac{1}{2}b^2,$$

and consequently, the payoff function in the game is

$$u(x, a, b) = u_0(x, a) + \frac{1}{2}\theta_0 b^2.$$

The term $\frac{1}{2}\theta_0 b^2$ is a penalized cost payed by nature. The parameter θ_0 can be viewed as the degree of robustness. For example, if θ_0 is large, then the penalization becomes so great that only the nominal model remains and the strategy is less robust. Conversely, the lower values of θ_0 allow to design a strategy which is appropriate for a wider set of model misspecifications. Therefore, a lower θ_0 is equivalent to a higher degree of robustness.

In this section, we shall consider the following pure strategies for nature. We say that $\gamma = \{\gamma_t\}_{t \in N_0} \in \Gamma^*$ is an *admissible strategy to nature*, if γ_t is a Borel measurable function from H_t^* to B, i.e., $b_t = \gamma_t(h_t^*)$, $t \in N_0$, and

$$\text{(M3)} \qquad E_x^{\pi\gamma}\left(\sum_{t=0}^{\infty} \beta^t b_t^2\right) < \infty$$

for every $x \in X$ and every $\pi \in \Pi$. The set of all admissible strategies to nature is denoted by Γ_0^*.

The objective of the controller is to find a policy $\pi^* \in \Pi$ such that

$$\inf_{\gamma \in \Gamma_0^*} E_x^{\pi^*\gamma}\left(\sum_{t=0}^{\infty} \beta^t \left\{u_0(x_t, a_t) + \frac{1}{2}\theta_0 b_t^2\right\}\right) = \max_{\pi \in \Pi} \inf_{\gamma \in \Gamma_0^*} E_x^{\pi\gamma}\left(\sum_{t=0}^{\infty} \beta^t \left\{u_0(x_t, a_t) + \frac{1}{2}\theta_0 b_t^2\right\}\right).$$

We solve the problem by proving that there exists a solution to the optimality equation. Therefore, we start with verification of assumption (M1). Note that

$$\int_X v(y) q\,(dy|x, a, b) = \int_R v(x + a + \varepsilon + b)\phi(\varepsilon)\, d\varepsilon$$

for any $v \in \underline{U}_\omega(X)$ and $(x, a, b) \in K$. Define the following weight function:

$$\omega(x) = \max\{x, 0\} + r \quad \text{for some constant } r \geq 1$$

and consider two cases: (i) $x \le 0$, and (ii) $x > 0$.

(i) If $x \le 0$, then $\omega(x) = r$ and

$$\begin{aligned}\int_R \omega(y) q\,(dy|x,a,b) &= \int_R \big[r + \max\{x + a + \varepsilon + b, 0\}\big]\phi(\varepsilon)\,d\varepsilon \\ &\le r + \int_R \max\{\hat{a} + \varepsilon, 0\}\phi(\varepsilon)\,d\varepsilon \\ &= r + \int_{-\hat{a}}^{\infty} [\hat{a} + \varepsilon]\phi(\varepsilon)\,d\varepsilon < r + \hat{a} + 1.\end{aligned}$$

(ii) If $x > 0$, then $\omega(x) = x + r$ and

$$\begin{aligned}\int_R \omega(y) q\,(dy|x,a,b) &= \int_R \big[r + \max\{x + a + \varepsilon + b, 0\}\big]\phi(\varepsilon)\,d\varepsilon \\ &\le r + \int_R \max\{x + \hat{a} + \varepsilon, 0\}\phi(\varepsilon)\,d\varepsilon \\ &= r + \int_{-\hat{a}-x}^{\infty} [x + \hat{a} + \varepsilon]\phi(\varepsilon)\,d\varepsilon < r + x + \hat{a} + 1.\end{aligned}$$

Assumption (M1) is satisfied for $\beta \in (0, 1)$ for which

$$\beta \frac{\int_R \omega(y) q\,(dy|x,a,b)}{\omega(x)} < 1 \quad \text{for all } (x,a,b) \in K.$$

Note that $\frac{r+x+\hat{a}+1}{r+x} \le \frac{r+\hat{a}+1}{r}$. Thus, it suffices to require that

$$\beta\left(1 + \frac{\hat{a}+1}{r}\right) < 1, \tag{20}$$

because

$$\beta \frac{\int_R \omega(y) q\,(dy|x,a,b)}{\omega(x)} < \beta \frac{r + \hat{a} + 1}{r}.$$

Observe that for every $\beta \in (0, 1)$ there exists a sufficiently large constant r such that (20) is satisfied.

Further, we replace condition (M2) with the following:

(M4) There exists a constant $d > 0$ such that $\sup_{a\in A} u_0^+(x,a) \le \omega(x)$ for all $x \in X$.

For any function $v \in \underline{U}_\omega(X)$, define the operator $\mathcal{T}$ as follows:

$$\mathcal{T}v(x) = \max_{a\in A} \inf_{b\in B} \left[u_0(x,a) + \frac{1}{2}\theta_0 b^2 + \beta \int_X v(y) q\,(dy|x,a,b)\right]$$

for all $x \in X$. Note that $\mathcal{T}$ maps the space $\underline{U}_\omega(X)$ into itself. Indeed, by (M1) and (M4), we have

$$\mathcal{T}v(x) \le \max_{a\in A} \left[u_0(x,a) + \beta \int_X v(y) q\,(dy|x,a,b)\right] \le d\omega(x) + \beta\alpha \left\|v^+\right\|_\omega \omega(x)$$

for all $x \in X$. By the maximum theorems of Berge (see Lemma 4) $\mathcal{T}v$ is upper semicontinuous. Hence, $(\mathcal{T}v)^+ \in U_\omega(X)$. Proceeding analogously as in the proof of Theorem 1, we infer that there exist a function $V \in \underline{U}_\omega(X)$ and $f^* \in F$ such that

$$V(x) = \mathcal{T}V(x) = \max_{a\in A}\inf_{b\in B}\left[u_0(x,a) + \frac{1}{2}\theta_0 b^2 + \beta\int_X V(y)q\,(dy|x,a,b)\right]$$
$$= \inf_{b\in B}\left[u_0\big(x, f^*(x)\big) + \frac{1}{2}\theta_0 b^2 + \beta\int_X V(y)q\,\big(dy|x, f^*(x), b\big)\right] \quad (21)$$

for $x \in X$. Moreover,

$$V(x) \le u_0\big(x, f^*(x)\big) + \frac{1}{2}\theta_0 b^2 + \beta\int_X V(y)q\,\big(dy|x, f^*(x), b\big) \quad \text{for any } b \in B \text{ and } x \in X.$$

Iterating this inequality n times, we obtain that

$$V(x) \le E_x^{f^*\gamma}\left(\sum_{t=0}^{n}\beta^t\left\{u_0(x_t,a_t) + \frac{1}{2}\theta_0 b_t^2\right\}\right) + \beta^{n+1}E_x^{f^*\gamma}V(x_{n+1}), \quad (22)$$

for any $\gamma \in \Gamma_0^*$ and $x \in X$. By (M1),

$$\beta^{n+1}E_x^{f^*\gamma}V(x_{n+1}) \le \beta^{n+1}\alpha^{n+1}\big\|V^+\big\|_\omega\omega(x).$$

Letting n tend to infinity in (22), using the above inequality and (M4) we get that

$$V(x) \le E_x^{f^*\gamma}\left(\sum_{t=0}^{\infty}\beta^t\left\{u_0(x_t,a_t) + \frac{1}{2}\theta_0 b_t^2\right\}\right)$$
$$= E_x^{f^*\gamma}\left(\sum_{t=0}^{\infty}\beta^t u_0(x_t,a_t)\right) + \frac{1}{2}\theta_0 E_x^{f^*\gamma}\left(\sum_{t=0}^{\infty} b_t^2\right),$$

for every $\gamma \in \Gamma_0^*$. Hence,

$$V(x) \le \inf_{\gamma\in\Gamma_0^*} J\big(x, f^*, \gamma\big).$$

On the other hand, in a similar way as in the proof of (10), we infer that $V(x) \ge \sup_{\pi\in\Pi}\inf_{\gamma\in\Gamma_0^*} J(x,\pi,\gamma)$. Thus,

$$V(x) = \inf_{\gamma\in\Gamma_0^*} J\big(x, f^*, \gamma\big) = \sup_{\pi\in\Pi}\inf_{\gamma\in\Gamma_0^*} J(x,\pi,\gamma)$$

for $x \in X$.

We have proved the following result.

Proposition 2 *Consider the system given in* (19). *If* (M3)–(M4) *hold, then there exist a function $V \in \underline{U}_\omega(X)$ and a stationary strategy f^* such that* (21) *is satisfied for all $x \in X$. The strategy f^* is optimal for the controller.*

We would like to emphasize that in the existing literature the utility function u_0 for the controller was assumed to be *bounded from above*; see [13, 14]. The above arguments can also be applied (with some minor modifications of calculations) to the situation in which we allow the choices b_t of nature to be bounded from above by some positive constant.

Example 7 Let x_t in (19) be a level of capital stock at a period $t \in N_0$. This level at each t can be changed due to the random shock ε_t as well as can be increased by the controller who selects some action $a_t \in A$. The (unbounded) utility function of the controller is $u_0(x,a) = x - a^2$, where a^2 is interpreted as a cost of changing the capital stock. Note that condition (M4) is satisfied for u_0. Hence, assuming (M3) there exist $V \in \underline{U}_\omega(X)$ and a stationary strategy f^* such that

$$\begin{aligned} V(x) &= \max_{a\in A} \inf_{b\in B}\left[x - a^2 + \frac{1}{2}\theta_0 b^2 + \beta \int_R V(x+a+\varepsilon+b)\phi(\varepsilon)\,d\varepsilon\right] \\ &= \inf_{b\in B}\left[x - \left(f^*(x)\right)^2 + \frac{1}{2}\theta_0 b^2 + \beta \int_R V\left(x + f^*(x) + \varepsilon + b\right)\phi(\varepsilon)\,d\varepsilon\right] \end{aligned}$$

for all $x \in X$. Clearly, f^* is optimal for the controller. A similar result holds if we consider $u_0(x,a) = \ln(1+x) - a^2$ for $x > 0$ and $u_0(x,a) = x - a^2$ for $x \le 0$.

In our setting, the choice of an action $b \in B$ by nature is penalized by the quadratic cost $\frac{1}{2}\theta_0 b^2$. Another approach is presented by Hansen and Sargent [13], Petersen et al. [46], who impose a constraint on the relative entropy, i.e., they assume that there exists a constant $\eta_0 > 0$ such that

$$\frac{1}{2} E_x^{\pi\gamma}\left(\sum_{t=0}^{\infty} b_t^2\right) \le \theta_0, \qquad P_x^{\pi\gamma}\text{-a.s.}$$

for $\pi \in \Pi$ and $\gamma \in \Gamma_0^*$. The mutual relation between these two methods is discussed in [13, 14] and [46].

6 Discounted Stochastic Games with Simultaneous Moves

In this section, we assume that $B(x,a) = B(x)$ is independent of $a \in A(x)$ for each $x \in X$. Thus, for each $t \in N_0$, player 2 does not observe player 1's action $a_t \in A(x_t)$ in state $x_t \in X$ while selecting $b_t \in B(x_t)$. Therefore, one can say that the players act simultaneously and play a standard discounted stochastic game introduced by Shapley [50]. Now a *strategy* for player 2 is a sequence $\gamma = \{\gamma_t\}_{t\in N_0}$ of stochastic kernels γ_t from H_t to B such that $\gamma_t(B(x_t)|h_t) = 1$ for any $h_t \in H_t$. The set of all strategies for player 2 is denoted by Γ. Let G_r be the set of all Borel measurable mappings $g : X \mapsto P(B)$ such that $g(x) \in P(B(x))$ for all $x \in X$. Every $g \in G_r$ induces a transition probability $g(db|x)$ from X to B and is recognized as a *randomized stationary strategy* for player 2. Let F_r be the set of all Borel measurable mappings $f : X \mapsto P(A)$ such that $f(x) \in P(A(x))$ for all $x \in X$. Then F_r can be viewed as the set of all *randomized stationary strategies* for player 1. Recall that the set of all strategies for player 1 is denoted by Π. For any initial state $x \in X$, $\pi \in \Pi$, $\gamma \in \Gamma$, the expected discounted payoff function $J(x,\pi,\gamma)$ is well defined under conditions (M1) and (M2) with the modification that now $B(x,a) = B(x)$ for each $a \in A(x)$. The *lower value* of the game is

$$v_*(x) := \sup_{\pi\in\Pi} \inf_{\gamma\in\Gamma} J(x,\pi,\gamma)$$

and the *upper value* of the game is defined as

$$v^*(x) := \inf_{\gamma \in \Gamma} \sup_{\pi \in \Pi} J(x, \pi, \gamma), \quad x \in X.$$

Suppose that the stochastic game has a *value*, i.e., $\tilde{v}(x) := v_*(x) = v^*(x)$ for each $x \in X$. Then, under our assumptions (M1) and (M2), $\tilde{v}(x) \in \underline{R}$. Let $\underline{X} := \{x \in X : \tilde{v}(x) = -\infty\}$. A strategy $\pi^* \in \Pi$ is *optimal* for player 1 if

$$\inf_{\gamma \in \Gamma} J\big(x, \pi^*, \gamma\big) = \tilde{v}(x), \quad x \in X.$$

Let $\varepsilon > 0$ be fixed. A strategy $\gamma^* \in \Gamma$ is ε-*optimal* for player 2 if

$$\sup_{\pi \in \Pi} J\big(x, \pi, \gamma^*\big) = \tilde{v}(x) \quad \text{for } x \in X \setminus \underline{X} \quad \text{and} \quad \sup_{\pi \in \Pi} J\big(x, \pi, \gamma^*\big) < -1/\varepsilon \quad \text{for } x \in \underline{X}.$$

Recall that $u_k := \max\{u, -k\}$, $k \in N$. By $\mathcal{G}_k$ we mean the game with the payoff function u_k. The discounted expected payoffs in $\mathcal{G}_k$ are denoted by $J_k(x, \pi, \gamma)$. The lower and upper values of the game $\mathcal{G}_k$ and ε-optimal strategies for the players are defined in an obvious way.[2]

Let

$$\bar{K}_A := \big\{(x, \nu) : x \in X, \nu \in P\big(A(x)\big)\big\}, \qquad \bar{K}_B := \big\{(x, \rho) : x \in X, \rho \in P\big(B(x)\big)\big\}$$

and

$$\bar{K} := \big\{(x, \nu, \rho) : x \in X, \nu \in P\big(A(x)\big), \rho \in P\big(B(x)\big)\big\}.$$

For any $(x, \nu, \rho) \in \bar{K}$ and $D \in \mathcal{B}(X)$, define

$$u_k(x, \nu, \rho) := \int_{A(x)} \int_{B(x)} u_k(x, a, b)\nu(da)\rho(db),$$
$$u(x, \nu, \rho) := \int_{A(x)} \int_{B(x)} u(x, a, b)\nu(da)\rho(db)$$

and

$$q(D|x, \nu, \rho) := \int_{A(x)} \int_{B(x)} q(D|x, a, b)\nu(da)\rho(db).$$

If $f \in F_r$ and $g \in G_r$, then

$$u_k(x, f, g) := u_k\big(x, f(x), g(x)\big), \qquad u(x, f, g) := u\big(x, f(x), g(x)\big) \quad \text{and}$$
$$q(D|x, f, g) := q\big(D|x, f(x), g(x)\big).$$

For any $(x, \nu, \rho) \in \bar{K}$ and $v \in \underline{U}_\omega(X)$, let

$$L_k v(x, \nu, \rho) := u_k(x, \nu, \rho) + \beta \int_X v(y)q(dy|x, \nu, \rho)$$

[2] Note that under our assumption (M1) the values of the game $\mathcal{G}_k$ are finite.

and

$$Lv(x,\nu,\rho) := u(x,\nu,\rho) + \beta \int_X v(y) q\,(dy|x,\nu,\rho).$$

From Lemma 7.12 in [4], Lemma 5.4 in [43] and Lemma 3, we obtain the following auxiliary result.

Lemma 6 *Assume* (W), (M1), (M2) *and that* $v \in \underline{U}_\omega(X)$. *Then the functions* $Lv(x,\nu,\rho)$ *and* $L_k v(x,\nu,\rho)$ *are upper semi-continuous on* $\bar{K}$.

Using Lemma 1, our assumption that $x \mapsto B(x)$ is lower semi-continuous, the maximum theorems of Berge [3] and the fact that

$$\inf_{\rho \in P(B(x))} L_k v(x,\nu,\rho) = \inf_{b \in B(x)} L_k v(x,\nu,b), \qquad \inf_{\rho \in P(B(x))} Lv(x,\nu,\rho) = \inf_{b \in B(x)} Lv(x,\nu,b)$$

we obtain the following conclusion.

Lemma 7 *Assume* (C1), (C2), (W), (M1), (M2), *and that* $v \in \underline{U}_\omega(X)$. *Then the functions*

$$(x,\nu) \mapsto \inf_{\rho \in P(B(x))} L_k v(x,\nu,\rho) \quad \text{and} \quad (x,\rho) \mapsto \max_{\nu \in P(A(x))} L_k v(x,\nu,\rho)$$

are upper semi-continuous on $\bar{K}_A$ *and* $\bar{K}_B$, *respectively. Similar properties hold if we replace* $L_k v(x,\nu,\rho)$ *by* $Lv(x,\nu,\rho)$.

For any $x \in X$, $v \in \underline{U}_\omega(X)$ and $k \in N$, define

$$T_k v(x) := \max_{\nu \in P(A(x))} \inf_{\rho \in P(B(x))} L_k v(x,\nu,\rho)$$

and

$$Tv(x) := \max_{\nu \in P(A(x))} \inf_{\rho \in P(B(x))} L_v(x,\nu,\rho).$$

By Lemma 7, the operators T_k and T are well defined. Using the maximum theorem of Berge [3], we conclude the following fact.

Lemma 8 *Under assumptions of Lemma* 7, $T_k v \in U_\omega(X)$ *for each* $v \in U_\omega(X)$ *and* $Tv \in \underline{U}_\omega(X)$ *for any* $v \in \underline{U}_\omega(X)$.

We have arrived at our main result in this section.

Theorem 2 *Assume that* $B(x,a) = B(x)$ *for each* $x \in X$ *and* $a \in A(x)$. *Suppose that* (A1)–(A7), (C1)–(C2), (W) *and* (M1)–(M2) *hold. Then the game has a value* $\tilde{v} \in \underline{U}_\omega(X)$. *Player* 1 *has an optimal stationary strategy* $f^* \in F_r$ *and*

$$T\tilde{v}(x) = \tilde{v}(x) = \max_{\nu \in P(A(x))} \inf_{\rho \in P(B(x))} L\tilde{v}(x,\nu,\rho) = \inf_{\rho \in P(B(x))} L\tilde{v}\big(x, f^*(x), \rho\big) \tag{23}$$

for each $x \in X$. *Moreover, for any* $\epsilon > 0$, *player* 2 *has an* ϵ*-optimal strategy and*

$$\tilde{v}(x) = \inf_{\rho \in P(B(x))} \max_{\nu \in P(A(x))} L\tilde{v}(x,\nu,\rho) \tag{24}$$

for every $x \in X$.

Proof Let $k \in N$. Consider the game $\mathcal{G}_k$. By Lemma 8, T_k maps the space $U_\omega(X)$ into itself. As in the proof of Theorem 1, we conclude from the Banach contraction mapping theorem that there exists a function $V_k \in U_\omega(X)$ such that $V_k = T_k V_k$. The arguments used in the proof of Theorem 1 allow to conclude that V_k is the lower value of the game $\mathcal{G}_k$. Moreover, the sequence $\{V_k\}_{k\in N}$ is non-increasing. Thus, by Lemma 5(i), there exists $V(x) := \lim_{k\to\infty} V_k(x)$ for each $x \in X$ and $V \in \underline{U}_\omega(X)$. In fact, $V(x) = \inf_{k\in N} V_k(x)$, $x \in X$. Using the monotone convergence theorem and Lemmas 5(ii), 6, and 7 we infer that

$$V(x) = TV(x) = \max_{\nu\in P(A(x))} \inf_{\rho\in P(B(x))} LV(x,\nu,\rho) \tag{25}$$

for all $x \in X$. From Corollary 1 in [7] and Lemmas 1 and 7, it follows that there exists some $f^* \in F_r$ such that

$$TV(x) = \inf_{\rho\in P(B(x))} LV\big(x, f^*(x), \rho\big)$$

for all $x \in X$. A simple adaptation of the arguments given in the proof of Theorem 1 enables us to claim that

$$V(x) = \inf_{\gamma\in\Gamma} J\big(x, f^*, \gamma\big) = v_*(x)$$

for all $x \in X$. By the minimax theorem of Fan [11], we have

$$V_k(x) = T_k V_k(x) = \max_{\nu\in P(A(x))} \inf_{\rho\in P(B(x))} L_k V_k(x,\nu,\rho) = \inf_{\rho\in P(B(x))} \max_{\nu\in P(A(x))} L_k V_k(x,\nu,\rho) \tag{26}$$

for every $k \in N$ and $x \in X$. By Lemma 2, there exists a sequence of Borel measurable mappings $g_n : X \mapsto P(B)$ such that the set $\{g_n(x) : n \in N\}$ is dense in $P(B(x))$ for every $x \in X$. By Lemma 7 and (26), we have

$$V_k(x) = T_k V_k(x) = \inf_{\rho\in P(B(x))} \max_{\mu\in P(A(x))} L_k V_k(x,\nu,\rho) = \inf_{n\in N} \max_{\nu\in P(A(x))} L_k V_k\big(x,\nu,g_n(x)\big) \tag{27}$$

for every $x \in X$. Fix any $\epsilon > 0$ and put $\epsilon' = \epsilon(1-\beta)$. Define

$$X_1 := \Big\{x \in X : V_k(x) \ge \max_{\nu\in P(A(x))} L_k V_k\big(x,\mu,g_1(x)\big) - \epsilon'\Big\}$$

and

$$X_n := \Big\{x \in X : V_k(x) \ge \max_{\nu\in P(A(x))} L_k V_k\big(x,\nu,g_n(x)\big) - \epsilon'\Big\} \Big\backslash \bigcup_{i<n} X_i$$

for $n \in N \setminus \{1\}$. Let $I := \{n \in N : X_n \ne \emptyset\}$. Then by (27), it follows that $\{X_i\}_{i\in I}$ is a measurable partition of the state space X. Let $g_0(x) := g_i(x)$ for $x \in X_i$, $i \in I$. Then

$$V_k(x) = T_k V_k(x) \ge \max_{\nu\in P(A(x))} L_k V_k\big(x,\nu,g_0(x)\big) - \epsilon' \ge L_k V_k\big(x,\nu',g_0(x)\big) - \epsilon'$$

for all $x \in X$, $\nu' \in P(A(x))$. By iteration of this inequality n times, for any $\pi \in \Pi$, we obtain

$$V_k(x) \ge E_x^{\pi g_0}\left(\sum_{t=0}^{n} \beta^t u_k(x_t,a_t,b_t)\right) + \beta^{n+1} E_x^{\pi g_0} V_k(x_{n+1}) - \epsilon' \sum_{t=0}^{n} \beta^t, \quad x \in X. \tag{28}$$

By assumption (M1),

$$\lim_{n\to\infty} \beta^{n+1} E_x^{\pi g_0} V_k(x_{n+1}) = 0.$$

Hence, letting n tend to infinity in (28) and using (M1) and (M2) we deduce that

$$V_k(x) \geq \sup_{\pi \in \Pi} J_k(x, \pi, g_0) - \epsilon \tag{29}$$

for all $x \in X$. Since V_k is the lower value of the game $\mathcal{G}_k$, $\epsilon > 0$ is arbitrary, from (29), it follows that V_k is actually the value of this game and for any $\epsilon > 0$ player 2 has an ϵ-optimal stationary strategy.

Recall that $V(x) = v_*(x)$ for each $x \in X$. Since for any $k \in N$, V_k is the upper value of the game $\mathcal{G}_k$, we have $V_k(x) \geq v^*(x)$ for each $x \in X$. Hence, $v_*(x) = V(x) = \inf_{k\in N} V_k(x) \geq v^*(x)$, $x \in X$. Thus, we have shown that the game has a value $\tilde{v}$ and $\tilde{v}(x) = V(x) = \inf_{k\in N} V_k(x)$ for all $x \in X$. Fix an $\epsilon > 0$. Using the games $\mathcal{G}_k$, one can construct measurable partitions $\{X'_i\}_{i\in I_1}$ and $\{X''_j\}_{j\in I_2}$ of the sets $\underline{X}$ and $X \setminus \underline{X}$, respectively, and subsequences $\{n'_i\}$, $\{m'_j\}$ such that

$$V_{n'_i}(x) < -\big(\epsilon^2 + 2\big)/2\epsilon \quad \text{for } x \in X'_i \quad \text{and} \quad V_{m'_j}(x) - \epsilon/2 < V(x) = \tilde{v}(x) \quad \text{for } x \in X''_j. \tag{30}$$

Choose an $(\epsilon/2)$-optimal stationary strategy $g_{n'_i}$ ($g_{m'_j}$) for player 2 in every game $\mathcal{G}_{n'_i}$ ($\mathcal{G}_{m'_j}$). Then

$$\sup_{\pi\in\Pi} J(x, \pi, g_{n'_i}) - \epsilon/2 \leq \sup_{\pi\in\Pi} J_{n'_i}(x, \pi, g_{n'_i}) - \epsilon/2 \leq V_{n'_i}(x) \quad \text{for } x \in X'_i \tag{31}$$

and

$$\sup_{\pi\in\Pi} J(x, \pi, g_{m'_j}) - \epsilon/2 \leq \sup_{\pi\in\Pi} J_{m'_j}(x, \pi, g_{m'_j}) - \epsilon/2 \leq V_{m'_j}(x) \quad \text{for } x \in X''_j. \tag{32}$$

Let $\gamma^* = \{g^*_t\}_{t\in N_0} \in \Gamma$ be defined as follows: $g^*_t = g^*$ for each $t \in N_0$ and $g^*(x_0, x_t) := g_{n'_i}(x_0)$ if $x_0 \in X'_i$, $g^*(x_0, x_t) := g_{m'_j}(x_0)$ if $x_0 \in X''_j$. In other words, using γ^* player 2 plays the stationary strategy $g_{n'_i}$ ($g_{m'_j}$) when the initial state $x_0 \in X'_i$ ($x_0 \in X''_j$). Such a strategy is called semi-stationary and by (30), (31) and (32), γ^* is ϵ-optimal for player 2. Using the fact that $\tilde{v}(x) = V(x)$ for each $x \in X$ and the minimax theorem of Fan [11], we conclude that (23) and (24) follow from (25). □

Remark 3 Zero-sum discounted stochastic games with compact metric state space and weakly continuous transitions were first studied by Maitra and Parthasarathy [33]. Couwenbergh [8] extended their result to unbounded games with a metric state space using the weighted supremum norm approach. He proved that both players possess optimal strategies. In order to obtain such a result additional conditions should be met. Namely, a function u is to be continuous and such that $|u(x, a, b)| \leq l\omega(x)$ for some constant $l > 0$ and all $(x, a, b) \in K$. Moreover, the mappings $x \mapsto A(x)$ and $x \mapsto B(x)$ are assumed to be compact-valued and continuous. However, our condition (M2) allows for much larger class of models and is less restrictive for discount factors compared with the weighted supremum norm approach. It is worth pointing out that zero-sum upper semi-continuous stochastic games with bounded from above non-additive payoff function and weakly continuous transitions were studied by Nowak [43].

Remark 4 A version of Theorem 2 can also be proved if we assume that $u(x, a, b)$ and $q(D|x, a, b)$ are continuous in $a \in A(x)$ for any $D \in \mathcal{B}(X)$, $x \in X$, $b \in B(x)$. Compactness of the sets $A(x)$, $x \in X$, is crucial in the proof, but (C1) and (C2) can be dropped. The proof itself proceeds along similar lines as in [42] and makes use of assumptions (M1) and (M2). However, this type of continuity imposed on the transition probability function is more restrictive and is rarely satisfied in stochastic models.

7 Stochastic Games with Simultaneous Moves and the Limiting Average Payoffs

In this section, we briefly describe some recent results on zero-sum stochastic games with average payoffs and weakly continuous transitions. For each initial state $x_0 = x \in X$ and any strategies $\pi \in \Pi$, $\gamma \in \Gamma$, the *expected average payoff per unit time* to player 1 is defined as follows:

$$J_a(x, \pi, \gamma) = \liminf_{n \to \infty} \frac{E_x^{\pi\gamma}(\sum_{t=0}^{n-1} u(x_t, a_t, b_t))}{n}.$$

Optimal strategies and the lower (upper) value of the game is defined as usual. In order to show that J_a is well defined for any $x \in X$, $\pi \in \Pi$, $\gamma \in \Gamma$, and a stochastic game with the expected average payoff has a value, we need to impose the following ergodicity conditions:

(E1) There exist a continuous function $\omega : X \mapsto [1, \infty)$, a Borel set $C \subset X$ such that, for some $\lambda \in (0, 1)$ and $\eta > 0$, it holds

$$\int_X \omega(y) q\,(dy|x, a, b) \le \lambda\omega(x) + \eta 1_C(x)$$

for every $(x, a, b) \in K$;

(E2) The function ω is bounded on C, that is,

$$v_C := \sup_{x \in C} \omega(x) < \infty;$$

(E3) There exist some $\widetilde{\delta} \in (0, 1)$ and $\mu \in P(C)$ with the property that

$$q(D|x, a, b) \ge \widetilde{\delta}\mu(D)$$

for any $D \in \mathcal{B}(X)$, $x \in C$, $a \in A(x)$ and $b \in B(x)$.

We point out that in this section we do not assume that ω satisfies condition (M1) from Sect. 3. Under (E1)–(E3), for any pair of stationary strategies $(f, g) \in F_r \times G_r$ the embedded state process $\{x_n\}$ is a positive recurrent aperiodic Markov chain and there exists a unique invariant probability measure π_{fg} (see Theorem 11.3.4 and p. 116 in [39]). Moreover, by Theorem 2.3 in [40], $\{x_n\}$ is ω-uniformly ergodic, that is, there exist $\widetilde{\theta} > 0$ and $\widetilde{\alpha} \in (0, 1)$ such that

$$\left| \int_X v(y) q^n\,(dy|x, f, g) - \int_X v(y) \pi_{fg}\,(dy) \right| \le \omega(x) \|v\|_\omega \widetilde{\theta}\,\widetilde{\alpha}^n \tag{33}$$

for every function v with finite ω-norm, $x \in X$ and $n \ge 1$. Here, $q^n(\cdot|x, f, g)$ denotes the n-stage transition probability induced by q and a pair of stationary strategies (f, g). As an immediate consequence of the inequality in (E1), one can easily get

$$J_a(f, g) := J_a(x, f, g) = \int_X u\bigl(x, f(x), g(x)\bigr) \pi_{fg}\,(dx) \tag{34}$$

for every $f \in F_r$ and $g \in G_r$.

Furthermore, we assume that

(B1) There exists an open set $\widetilde{C} \subset C$ such that $\mu(\widetilde{C}) > 0$;
(B2) There exists a constant $d > 0$ such that

$$\left|u(x, a, b)\right| \leq d\omega(x)$$

for all $(x, a, b) \in K$; u is continuous on K;
(B3) $B(x, a) = B(x)$ is compact for each $x \in X$, $a \in A(x)$.
(B4) The set-valued mappings $x \mapsto A(x)$ and $x \mapsto B(x)$ are continuous.

The following result was proved in [21, 22].

Theorem 3 *Assume* (A1)–(A4), (A6) (W), (E1)–(E3), *and* (B1)–(B4). *There exist a constant* $\hat{\xi}$, *a continuous function* $\hat{h}$ *with finite* ω*-norm* (*i.e.*, $\|\hat{h}\|_\omega < \infty$), $f^* \in F_r$ *and* $g^* \in G_r$ *such that following equations hold*:

$$\begin{aligned} \hat{h}(x) + \hat{\xi} &= val\left[u(x, \cdot, \cdot) + \int_X \hat{h}(y) q\,(dy|x, \cdot, \cdot)\right] \\ &= \min_{\rho \in P(B(x))} \left[u\big(x, f^*(x), \rho\big) + \int_X \hat{h}(y) q\,\big(dy|x, f^*(x), \rho\big)\right] \\ &= \max_{\nu \in P(A(x))} \left[u\big(x, \nu, g^*(x)\big) + \int_X \hat{h}(y) q\,\big(dy|x, \nu, g^*(x)\big)\right], \end{aligned} \tag{35}$$

for all $x \in X$. *Moreover*, $\hat{\xi}$ *is the value of the game with the expected average payoffs*, $f^* \in F_r$ *is optimal for player* 1, *whereas* $g^* \in G_r$ *is optimal for player* 2.

The proof of this result is based on the Banach contraction mapping theorem applied to two contractive operators, say T and $\widetilde{T}$, defined in [21, 22]. More precisely, T is an operator corresponding to the game with stage payoff function u, whereas $\widetilde{T}$ is related to the game with payoff function $-u$. It turns out that both of them map the space of all lower semi-continuous functions with finite ω-norm into the same space. The continuity of $\hat{h}$ in the above theorem is a consequence of the fact that the fixed point of operator T, say h, must equal to the fixed point of $\widetilde{T}$, say $\widetilde{h}$, multiplied by -1. This means that the lower semi-continuous function h is equal to the upper semi-continuous function $-\widetilde{h}$, which only happens if $h = -\widetilde{h} =: \hat{h}$ is continuous. It is worth mentioning that works [21, 22] concern a more general setting, namely semi-Markov games, for which the time between successive jumps is a random variable.

Besides articles of Jaśkiewicz [21, 22], the zero-sum Markov games with weakly continuous probabilities were studied by Jaśkiewicz and Nowak [23] and Küenle [29]. In the former work, Jaśkiewicz and Nowak [23] prove that the Markov game has a value and both players have optimal stationary strategies. Their approach is based on applying Fatou's lemma for weakly convergent measures, which in turn allows to get two optimality inequalities instead of one optimality equation. Moreover, the proof employs Michael's theorem on a continuous selection. A completely different approach was presented by Küenle [29]. Under slightly weaker assumptions, he introduced certain contraction operators that lead to a parameterized family of functional equations. Making use of some continuity and monotonicity properties of the solutions to these equations (with respect to the parameter), he obtained a lower semicontinuous solution of the optimality equation. Then he showed that the maximizing player has an ϵ-optimal stationary strategy, whereas the minimizing player has an optimal stationary one. It is worth mentioning that a direct application of a fixed point theorem in the minimax control setting was introduced in [12]. Nevertheless, this method has some disadvantages, because it requires stronger assumptions and excludes many interesting examples. More precisely, the ergodicity assumptions are formulated with the aid of a function $\delta : K \mapsto R_+$ instead of the characteristic function $1_C(\cdot)$ defined on X (see Assumptions 5.1 in [12]). In order to obtain a lower semi-continuous solution to the optimality equation, the authors have to assume that δ is continuous on K (not upper semicontinuous as it is formulated on p. 1,640). Clearly, this requirement is not met in models which satisfy conditions (E1)–(E2) with some Borel set $C \subset X$ and $C \neq X$.

The existence of value for stochastic games with the expected average *bounded* payoff functions, finitely many states and actions, as well as for stochastic games with infinitely many states and actions satisfying some additional conditions, was proved by Mertens and Neyman [38]. A general approach to Borel state space stochastic games was developed by Maitra and Sudderth [34]. They have considered a very large class of payoffs (including average payoff case) and proved the existence of value in the class of universally measurable strategies. Let us note, however, that optimal strategies need not exist in such games. Their assumptions on the transition probability function is different from ours. In fact, q is required to be setwise continuous, but only in the actions of one player, i.e., the function $a \mapsto q(D|x, a, b)$ is continuous on $A(x)$ for any $D \in \mathcal{B}(X)$ and $x \in X$, $b \in B(x)$. Although the results in [34, 38] are very deep, their methods cannot be applied to our setting with *weakly continuous* transition law. This is because the assumption that q is weakly continuous does not imply that the function $(a, b) \mapsto \int_X v(y)q\,(dy|x, a, b)$ is semi-continuous with respect to $a \in A(x)$ or $b \in B(x)$, if v is discontinuous. It is worth emphasizing that a certain type of continuity is essential for application of minimax theorems to auxiliary one shot games.

References

1. Alvarez F, Stokey N (1998) Dynamic programming with homogeneous functions. J Econ Theory 82:167–189
2. Başar T, Bernhard P (1995) H^∞-optimal control and related minimax design problems. Birkhäuser, Boston
3. Berge C (1963) Topological spaces. MacMillan, New York
4. Bertsekas DP, Shreve SE (1978) Stochastic optimal control: the discrete-time case. Academic Press, New York
5. Bhattacharya R, Majumdar M (2007) Random dynamical systems: theory and applications. Cambridge University Press, Cambridge
6. Blackwell D (1965) Discounted dynamic programming. Ann Math Stat 36:226–235
7. Brown LD, Purves R (1973) Measurable selections of extrema. Ann Stat 1:902–912
8. Couwenbergh HAM (1980) Stochastic games with metric state spaces. Int J Game Theory 9:25–36
9. Durán J (2000) On dynamic programming with unbounded returns. Econom Theory 15:339–352
10. Durán J (2003) Discounting long-run average growth in stochastic dynamic programs. Econom Theory 22:395–413
11. Fan K (1953) Minimax theorems. Proc Natl Acad Sci USA 39:42–47
12. González-Trejo JI, Hernández-Lerma O, Hoyos-Reyes LF (2003) Minimax control of discrete-time stochastic systems. SIAM J Control Optim 41:1626–1659
13. Hansen LP, Sargent TJ (2008) Robustness. Princeton University Press, Princeton
14. Hansen LP, Sargent TJ (2010) Wanting robustness in macroeconomics. In: Friedman BM, Woodford M (eds) Handbook of monetary economics, vol 3, pp 1097–1157
15. Hansen LP, Sargent TJ, Tallarini TD (1999) Robust permanent income and pricing. Rev Econ Stud 66:873–907

16. Hernández-Lerma O, Lasserre JB (1996) Discrete-time Markov control processes: basic optimality criteria. Springer, New York
17. Hernández-Lerma O, Lasserre JB (1999) Further topics on discrete-time Markov control processes. Springer, New York
18. Himmelberg CJ (1975) Measurable relations. Fundam Math 87:53–72
19. Himmelberg CJ, Van Vleck FS (1975) Multifunctions with values in a space of probability measures. J Math Anal Appl 50:108–112
20. Iyengar GN (2005) Robust dynamic programming. Math Oper Res 30:257–280
21. Jaśkiewicz A (2009) Zero-sum ergodic semi-Markov games with weakly continuous transition probabilities. J Optim Theory Appl 141:321–348
22. Jaśkiewicz A (2010) On a continuous solution to the Bellman–Poisson equation in stochastic games. J Optim Theory Appl 145:451–458
23. Jaśkiewicz A, Nowak AS (2006) Zero-sum ergodic stochastic games with Feller transition probabilities. SIAM J Control Optim 45:773–789
24. Jaśkiewicz A, Nowak AS (2011) Discounted dynamic programming with unbounded returns: application to economic models. J Math Anal Appl 378:450–462
25. Kamihigashi T (2007) Stochastic optimal growth with bounded and unbounded utility and with bounded or unbounded shocks. J Math Econ 43:477–500
26. Karatzas I, Sudderth WD (2010) Two characterizations of optimality in dynamic programming. Appl Math Optim 61:421–434
27. Kuratowski K, Ryll-Nardzewski C (1965) A general theorem on selectors. Bull Acad Pol Sci (Ser Math) 13:397–403
28. Küenle HU (1986) Stochastische Spiele und Entscheidungsmodelle. BG Teubner, Leipzig
29. Küenle HU (2007) On Markov games with average reward criterion and weakly continuous transition probabilities. SIAM J Control Optim 45:2156–2168
30. Kurano M (1987) Minimax strategies for average cost stochastic games with an application to inventory models. J Oper Res Soc Jpn 30:232–247
31. Le Van C, Morhaim L (2002) Optimal growth models with bounded or unbounded returns: a unifying approach. J Econ Theory 105:158–187
32. Ljungqvist L, Sargent TJ (2000) Recursive macroeconomic theory. MIT Press, Cambridge
33. Maitra A, Parthasarathy T (1970) On stochastic games. J Optim Theory Appl 5:289–300
34. Maitra A, Sudderth W (1993) Borel stochastic games with limsup payoffs. Ann Probab 21:861–885
35. Maccheroni F, Marinacci M, Rustichini A (2006) Dynamic variational preferences. J Econ Theory 128:4–44
36. Martins-da-Rocha VF, Vailakis Y (2010) Existence and uniqueness of fixed-point for local contractions. Econometrica 78:1127–1141
37. Matkowski J, Nowak AS (2011) On discounted dynamic programming with unbounded returns. Econom Theory 46:455–474
38. Mertens JF, Neyman A (1981) Stochastic games. Int J Game Theory 10:53–66
39. Meyn SP, Tweedie RL (1993) Markov chains and stochastic stability. Springer, New York
40. Meyn SP, Tweedie RL (1994) Computable bounds for geometric convergence rates of Markov chains. Ann Appl Probab 4:981–1011
41. Neveu J (1965) Mathematical foundations of the calculus of probability. Holden-Day, San Francisco
42. Nowak AS (1985) Universally measurable strategies in zero-sum stochastic games. Ann Probab 13:269–287
43. Nowak AS (1986) Semicontinuous nonstationary stochastic games. J Math Anal Appl 117:84–99
44. Nowak AS (2010) On measurable minimax selectors. J Math Anal Appl 366:385–388
45. Parthasarathy KR (1967) Probability measures on metric spaces. Academic Press, New York
46. Petersen IR, James MR, Dupuis P (2000) Minimax optimal control of stochastic uncertain systems with relative entropy constraints. IEEE Trans Autom Control 45:398–412
47. Puterman M (1980) Markov decision processes: discrete stochastic dynamic programming. Wiley-Interscience, New York
48. Rincón-Zapatero JP, Rodriguez-Palmero C (2003) Existence and uniqueness of solutions to the Bellman equation in the unbounded case. Econometrica 71:1519–1555. Corrigendum Econometrica 77:317–318
49. Schäl M (1975) Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal. Z Wahrscheinlichkeitstheor Verw Geb 32:179–196
50. Shapley LS (1953) Stochastic games. Proc Natl Acad Sci USA 39:1095–1100

51. Stokey NL, Lucas RE, Prescott E (1989) Recursive methods in economic dynamics. Harvard University Press, Cambridge
52. Strauch R (1966) Negative dynamic programming. Ann Math Stat 37:871–890
53. Wessels J (1977a) Markov programming by successive approximations with respect to weighted supremum norms. J Math Anal Appl 58:326–335
54. Wessels J (1977b) Markov games with unbounded rewards. In: Schäl M (ed) Dynamische Optimierung, Bonn, 1977. Bonner Mathematische Schriften, vol 98, pp 133–147