Markov decision processes with quasi-hyperbolic discounting

Anna Jaśkiewicz[1] · **Andrzej S. Nowak[2]**

Abstract We study Markov decision processes with Borel state spaces under quasi-hyperbolic discounting. This type of discounting nicely models human behaviour, which is time-inconsistent in the long run. The decision maker has preferences changing in time. Therefore, the standard approach based on the Bellman optimality principle fails. Within a dynamic game-theoretic framework, we prove the existence of randomised stationary Markov perfect equilibria for a large class of Markov decision processes with transitions having a density function. We also show that randomisation can be restricted to two actions in every state of the process. Moreover, we prove that under some conditions, this equilibrium can be replaced by a deterministic one. For models with countable state spaces, we establish the existence of deterministic Markov perfect equilibria. Many examples are given to illustrate our results, including a portfolio selection model with quasi-hyperbolic discounting.

✉ A.S. Nowak
a.nowak@wmie.uz.zgora.pl

A. Jaśkiewicz
anna.jaskiewicz@pwr.edu.pl

[1] Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, Wyspiańskiego 27, 50-370 Wrocław, Poland

[2] Faculty of Mathematics, Computer Science and Econometrics, University of Zielona Góra, Licealna 9, 65-417 Zielona Góra, Poland

1 Introduction

The discounted utility approach in dynamic decision making has been used since the beginning of modern economic theory; see e.g. Samuelson [59]. It is based on the assumption that the discount rate is constant over time. In that way, it is possible to compare outcomes occurring at different times by discounting future utility by some constant factor. A decision maker using high discount rates exhibits more impatience than one with low discount rates. It should be noted, however, that there is growing evidence to think that standard (geometric) discounting is not adequate in many real life situations; see e.g. Ainslie [2]. When discounting is non-standard, the decision maker becomes time-inconsistent, that is, a policy chosen as optimal at the beginning of the decision process is no longer optimal if it is considered as a policy in the process from some later point in time onwards. It is said that the decision maker possesses changing time preferences or that his utilities change over time. For example, consider a consumption/saving problem in discrete time. Suppose that the decision maker plans to save a lot tomorrow, but as tomorrow comes, he reconsiders his previous decision and saves little. Consumption is becoming more important and he becomes impatient. In other words, he may redecide on his plans later on. This shows that the consumption/saving problem cannot be solved via the usual dynamic programming methods.

The idea of quasi-hyperbolic discounting used to capture the case of utilities changing in time can also be described as follows. Suppose that u_t is a utility (or reward) to be received in period $t \geq 1$. Then the total utility (reward) collected from period t onwards is

$$U_t := u_t + \alpha(\beta u_{t+1} + \beta^2 u_{t+2} + \cdots), \tag{1.1}$$

where $\alpha > 0$ is called the *short-run discount factor* and $\beta \in (0, 1)$ is called the *long-run discount factor*. If $\alpha = 1$, then (1.1) reduces to the standard discounted utility. If the utility (1.1) was time-consistent, then we would have $U_t = u_t + \alpha\beta U_{t+1}$ for all t. That is the case when $\alpha = 1$. To observe that the Bellman principle may not be a right tool for constructing optimal policies in models with utilities changing over time, the reader is referred to simple examples in e.g. [15, 38] and Sect. 4 in this paper.

Dynamic inconsistent behaviour was first formalised by Strotz [67]. Further works by Pollak [56], Phelps and Pollak [55], Peleg and Yaari [54] and others on this issue suggest that policies optimal in some sense for the decision maker in models with quasi-hyperbolic discounting can be constructed as Nash equilibria in a sequential game played by different temporal *selves*. Each player (*self*) acts only once and takes into account both his instantaneous utility (reward) and a sequence of utilities (received by the players in subsequent periods) discounted by the given coefficient β. Within such a framework, the most commonly used solution concept is that of subgame perfect equilibrium in Markov strategies. Phelps and Pollak [55] considered a deterministic model of economic growth and using discounting with α and β as in (1.1), they introduced a multigenerational game. A generation in their game formulation is a self in the model mentioned above.

Nowadays, dynamic inconsistency plays an increasingly important role in many fields. For instance, we wish to mention the papers of Balbus et al. [6] or Har-

ris and Laibson [27] that deal with consumption/investment problems with a one-dimensional state space. Moreover, Barro [10], Ekeland and Pirvu [24], Haurie [28], Laibson [39] considered interesting applications of time-inconsistency to neoclassical growth theory, portfolio management, global climate change problems and macroeconomic theory, respectively. The reader is also referred to other works studying various related control problems for models with a general state space; see Björk and Murgoci [15], Björk et al. [14], Christensen and Lindensjö [20], Jaśkiewicz and Nowak [35] or Nowak [51].

A seminal paper of Shapley [63] on discounted zero-sum stochastic games is a first study of Markov decision processes over an infinite time horizon. Alj and Haurie [4] extended the finite state space model of Shapley to quasi-hyperbolic discounting. They used an intergenerational dynamic game formulation of Phelps and Pollak [55] and proved that any finite horizon game has an equilibrium in Markovian strategies and each infinite horizon game has a stationary Markov perfect equilibrium. The former result is based on a dynamic-programming-like algorithm and the latter is proved using a fixed point argument. The stochastic variants of the intergenerational game related to that of Alj and Haurie [4] with a Borel state space and compact metric action spaces were further examined in Jaśkiewicz and Nowak [35] and Nowak [51]. For instance, Jaśkiewicz and Nowak [35] studied a model in which generations are risk-averse and obtained a stationary Markov perfect equilibrium in pure strategies making use of the Dvoretzky–Wald–Wolfowitz theorem. This result, however, is valid for transitions which are convex combinations of finitely many atomless measures on the state space with coefficients that depend on the state–action pairs. Although as already mentioned time-inconsistent preferences in various control models were recently studied by Björk and Murgoci [15], Björk et al. [14], Christensen and Lindensjö [20], these papers, in contrast to our present work and works of Alj and Haurie [4], Jaśkiewicz and Nowak [35], Nowak [51], examine neither stationary Markov perfect equilibria nor fixed points of best-response mappings.

Markov decision processes have many applications to economic dynamics, finance, insurance or monetary economics. The reader is referred to the books of Bäuerle and Rieder [11, Chap. 9], Stachurski [65, Chaps. 10–12], Stokey et al. [66, Chaps. 10 and 13], where prominent and representative examples are given. In the present paper, we consider Markov decision processes with a Borel state space and quasi-hyperbolic discounting and the Markov perfect equilibrium as a basic solution concept. Our contribution is four-fold. First, we show that there exists a stationary Markov perfect equilibrium if the transition probability is norm-continuous in actions and has a density function. This result (Theorem 3.2) can be regarded as an improvement and an extension of the basic theorem in Nowak [51], where an additional condition on the transition probability density is imposed. Furthermore, it turns out that the obtained stationary Markov perfect equilibrium can be supported in every state on at most two points from the action set (Theorem 3.4). Second, assuming in addition that the transitions are atomless and the Borel σ-field on the state space has no so-called conditional atoms with respect to the σ-field generated by the transition density functions, we apply Theorem 3.4 and a result of Dynkin and Evstigneev [23] to prove the existence of a deterministic stationary Markov perfect equilibrium (Theorem 3.5). Our third contribution establishes the existence of a *deterministic*

Markov perfect equilibrium in decision processes with countably many states (Theorem 5.2). This result is subsequently used for Markov decision processes with Borel state spaces to obtain ϵ-equilibria by an approximation technique (Theorem 6.2). In Sect. 5, we provide an example of a Markov decision process with two states for which a deterministic stationary Markov perfect equilibrium does not exist, but there exists a deterministic non-stationary one. It is interesting to see that a randomised stationary Markov perfect equilibrium in this example can be dominated in terms of expected utilities (rewards) by a more sophisticated (in some sense) deterministic equilibrium.

Our main results for Markov decision processes with a continuum of states have certain implications for consumption/investment models with i.i.d. shocks. They complete the results obtained by Balbus et al. [6] and Harris and Laibson [27] for such models with atomless transitions. In Sect. 4, we discuss many examples arising from economic theory, macroeconomics or monetary economics. They highlight many issues in the area of quasi-hyperbolic discounting in dynamic decision processes, including some open problems. We also present a closed-form solution to a portfolio selection model originally studied with geometric discounting by Samuelson [60] (see Example 4.1 and Remark 4.2).

The paper is organised as follows. In Sect. 2, we describe our model and define the notion of a stationary Markov perfect equilibrium. In Sect. 3, we state our main results with many comments on the main ideas behind them. Their formal proofs are postponed to Sect. 7. Section 4 is devoted to examples of stationary Markov perfect equilibria, some comments on the literature and open problems. In Sect. 5, we study deterministic Markov perfect equilibria in decision models with countably many states and show that a deterministic Markov perfect equilibrium need not exist even if the state space is finite. In Sect. 6, making use of an approximation of the original Markov decision process by models with countably many states, we establish the existence of ϵ-equilibria. Finally, Sect. 8 contains some concluding remarks.

2 The model and main solution concept

First we give some basic definitions and facts used in the description of our model. Let $\mathbb{N}$ be the set of all positive integers, $\mathbb{R}$ be the set of all real numbers and $\mathbb{R}_+ = [0, \infty)$. A *Borel space*, say X, is a nonempty Borel subset of a complete separable metric space. Let $\mathcal{B}(X)$ denote the σ-field of all Borel subsets of X and $\Pr(X)$ the space of all probability measures on $\mathcal{B}(X)$, endowed with the topology of weak convergence. This is the coarsest topology for which the functionals $p \mapsto \int_X \eta dp$ are continuous for every bounded continuous function $\eta : X \to \mathbb{R}$.

A *Borel transition probability* from X to a Borel space Z is by definition a function $\gamma : \mathcal{B}(Z) \times X \to [0, 1]$ such that $\gamma(B, \cdot)$ is Borel-measurable on X for each $B \in \mathcal{B}(Z)$ and $\gamma(\cdot, x) \in \Pr(Z)$ for each $x \in X$. We write $\gamma(B|x)$ for $\gamma(B, x)$. It is well known that any Borel transition probability from X to Z can be viewed as a Borel mapping from X to $\Pr(Z)$; see Bertsekas and Shreve [13, Chap. 7].

Let S and A be Borel spaces and $\mathbb{K}$ a Borel subset of $S \times A$. Moreover, assume that for each $s \in S$, the section

$$A(s) := \{a \in A : (s, a) \in \mathbb{K}\} \tag{2.1}$$

is σ-compact, i.e., a countable union of nonempty compact sets. Then by Brown and Purves [17, Theorem 1], the correspondence $s \mapsto A(s)$ has a Borel-measurable selector, i.e., there exists a Borel mapping $f : S \to A$ such that $f(s) \in A(s)$ for all $s \in S$.

We consider a *Markov decision process* characterised by the following objects:

(i) S is a Borel *state space*.

(ii) A is a Borel *action space* and $\mathbb{K} \subseteq \mathcal{B}(S \times A)$ is the constraint set for the decision maker. The set $A(s)$ defined in (2.1) is a nonempty σ-compact set of actions available in state $s \in S$. (Our main results are stated for models with compact action spaces, but in some examples, we only assume that the sets $A(s)$ are σ-compact.)

(iii) $u : \mathbb{K} \to \mathbb{R}$ is a bounded from above Borel *instantaneous utility or reward function*.

(iv) q is a Borel transition probability from $\mathbb{K}$ to S.

(v) $\beta \in (0, 1)$ is a *long-run discount factor* and $\alpha > 0$ is a *short-run discount factor*.

Let Φ be the set of all Borel transition probabilities ϕ from S to A such that $\phi(A(s)|s) = 1$ for each $s \in S$. Every $\phi \in \Phi$ can be viewed as a Borel mapping from S to $\Pr(A)$ (denoted also by ϕ) by setting $\phi(s)(\cdot)$ for $\phi(\cdot|s)$. Let F be the set of all Borel selectors of the correspondence $s \mapsto A(s)$. By Brown and Purves [17, Theorem 1], $F \neq \emptyset$. Clearly, $F \subseteq \Phi$.

In the *decision model with quasi-hyperbolic discounting,* we envision an individual decision maker as a sequence of autonomous temporal *selves*. The selves are indexed by period numbers $t \in T := \mathbb{N}$. For each state $s_t \in S$ at the beginning of the tth period, self t chooses an action $a_t \in A(s_t)$ according to some probability distribution over the set $A(s_t)$. A strategy for self t is a Borel transition probability $\varphi_t \in \Phi$. Then $\bar{\varphi} = (\varphi_1, \varphi_2, \dots)$ with $\varphi_t \in \Phi$ for every $t \in T$ is called a *strategy profile* of the selves or a *Markov strategy* of the decision maker. If $\varphi_t = \phi$ for some $\phi \in \Phi$ and all $t \in T$, then such a strategy of the decision maker is called *stationary* or a *stationary strategy profile* of the selves. We often identify a constant sequence $\bar{\phi} = (\phi, \phi, \dots)$ with $\phi \in \Phi$.

By the Ionescu-Tulcea theorem (Neveu [48, Proposition V.1.1]), for any $s_t \in S$ and $\bar{\phi} = (\phi_1, \phi_2, \dots) \in \Phi^\infty$, there exists a unique probability measure $P_{s_t}^{\bar{\phi}}$ on the space $(S \times A)^\infty$ of all sequences of state–action pairs (starting at s_t) endowed with the product Borel σ-field. The symbol $E_{s_t}^{\bar{\phi}}$ denotes the corresponding expectation operator.

The *expected reward* for self t is defined as

$$R_t(\bar{\phi})(s_t) := E_{s_t}^{\bar{\phi}}\left[u(s_t, a_t) + \alpha\beta \sum_{\tau=t+1}^{\infty} \beta^{\tau-t-1} u(s_\tau, a_\tau) \right]. \tag{2.2}$$

This definition explains why $\beta \in (0, 1)$ is a *long-run discount factor* and $\alpha > 0$ is a *short-run discount factor*. Notice that the discount factor applied between periods t and $t + 1$ is $\alpha\beta$, and for $\alpha < 1$, it is lower than the one used between consecutive dates in the future. This fact leads to time-inconsistency in preferences. Such preferences have been observed in many cases and individuals' behaviours; see for instance Krusell et al. [37], Krusell and Smith [38], Laibson [39], Phelps and Pollak [55] or Strotz [67]. Their axiomatic characterisation can be found in Montiel Olea and Strzalecki [47].

In Sects. 2–4, we consider solutions for Markov decision processes in stationary strategies. Therefore, we simplify the notation as follows. Suppose that the selves are going to use a stationary strategy profile $\bar{\phi} = (\phi, \phi, \dots)$ identified with $\phi \in \Phi$. We use $E_{s_t}^{\phi}$ for $E_{s_t}^{\bar{\phi}}$. Moreover, $R_t(\bar{\phi})(s_t)$ defined in (2.2) is equal to $R(\phi)(s)$ given by

$$R(\phi)(s) := E_s^{\phi}\left[u(s, a) + \alpha\beta \sum_{\tau=1}^{\infty} \beta^{\tau-1} u(s_\tau, a_\tau)\right] \tag{2.3}$$

with $s = s_t$. In order to write this reward in a more friendly way, we define

$$J^{\beta}(\phi)(s') := E_{s'}^{\phi}\left[\sum_{\tau=1}^{\infty} \beta^{\tau-1} u(s_\tau, a_\tau)\right]. \tag{2.4}$$

Let $v \in \Pr(A(s))$. We introduce the notations

$$u(s, v) := \int_{A(s)} u(s, a)v(da), \qquad q(ds'|s, v) := \int_{A(s)} q(ds'|s, a)v(da).$$

Moreover, for any $s \in S$ and $\phi \in \Phi$, we put

$$u(s, \phi(s)) := \int_{A(s)} u(s, a)\phi(da|s), \qquad q(ds'|s, \phi(s)) := \int_{A(s)} q(ds'|s, a)\phi(da|s).$$

Observe that now (2.3) takes the form

$$R(\phi)(s) = u(s, \phi(s)) + \alpha\beta \int_S J^{\beta}(\phi)(s')q(ds'|s, \phi(s)).$$

Furthermore, for any $s \in S$ and $v \in \Pr(A(s))$, we define

$$P(s, v, \phi) := u(s, v) + \alpha\beta \int_S J^{\beta}(\phi)(s')q(ds'|s, v). \tag{2.5}$$

For any $a \in A(s)$, let us define

$$P(s, a, \phi) = P(s, \delta_a, \phi),$$

where δ_a is the Dirac measure at the point a.

Assume that self t chooses a randomised action $v \in \Pr(A(s))$ in state $s = s_t$. If all following selves are going to employ a strategy ϕ, then $P(s, v, \phi)$ is the expected utility of self t is state $s = s_t$.

Definition 2.1 A *stationary Markov perfect equilibrium* is a $\phi^* \in \Phi$ such that for every $s \in S$, we have

$$\sup_{v \in \Pr(A(s))} P(s, v, \phi^*) = P(s, \phi^*(s), \phi^*) = R(\phi^*)(s).$$

A stationary Markov perfect equilibrium ϕ^* is called *deterministic* if $\phi^* \in F$.

One can imagine that every self is a short-lived player in a non-cooperative game and acts only once. Such an interpretation is given by Alj and Haurie [4], Phelps and Pollak [55]. The payoff function of self $t \in T$ is given by (2.3). Then a stationary Markov perfect equilibrium is a constant sequence $(\phi^*, \phi^*, \dots)$ (identified with $\phi^* \in \Phi$) being a symmetric Nash equilibrium in this game. From Definition 2.1, it follows that this equilibrium is subgame perfect; see Osborne [53, Chap. 5.4]. The term Markov perfect equilibrium was introduced by Maskin and Tirole [45]. The strategies in a Markov perfect equilibrium have the Markov property of the lack of memory, meaning that each player's mixed action can be conditioned only on the state of the game. Moreover, the state can only encode payoff-relevant information. The strategy for the decision maker built from a Markov perfect equilibrium in the game is *time-consistent*, that is, no self (as time goes on) has an incentive to change his best response to equilibrium strategies of the following selves.

3 Existence of stationary Markov perfect equilibria

In this section, we state our main results on stationary equilibria and give many comments. The proofs are postponed to Sect. 7.

3.1 Basic assumptions and three equilibrium theorems

In order to formulate our results, we need the following additional assumptions.

(C3.1) The function $u(s, \cdot)$ is bounded and continuous on $A(s)$ for each $s \in S$. The set $A(s)$ is compact for every $s \in S$.

(C3.2) There exist a nonnegative Borel function $\rho : \mathbb{K} \times S \to \mathbb{R}$ and a probability measure $p \in \Pr(S)$ such that for all $(s, a) \in \mathbb{K}$ and $B \in \mathcal{B}(S)$,

$$q(B|s, a) = \int_B \rho(s, a, s') p(ds'),$$

and if $a_n \to a_0$ in $A(s)$ as $n \to \infty$, then

$$\lim_{n \to \infty} \int_S |\rho(s, a_n, s') - \rho(s, a_0, s')| p(ds') = 0.$$

Remark 3.1 By Scheffé's lemma [62], condition (C3.2) is equivalent to the norm continuity of $a \mapsto q(\cdot|s, a)$ on $A(s)$. As noted by Schäl [61, Remark 5.1], (C3.2) holds if the function $a \mapsto \rho(s, a, s')$ is lower semicontinuous on $A(s)$ for all $s, s' \in S$.

Now we can state our main results on stationary equilibria.

Theorem 3.2 *Assume that* (C3.1) *and* (C3.2) *hold. Then there exists a stationary Markov perfect equilibrium.*

Let us introduce the following assumption.

(C3.3) The state space S is countable and for all $s, s' \in S$, the function $a \mapsto q(s'|s, a)$ is continuous on $A(s)$.

Let p be any probability distribution on S such that $p(s) > 0$ for all $s \in S$. If we take into account that $\rho(s, a, s') = q(s'|s, a)/p(s')$, then by Scheffé's lemma, (C3.3) implies (C3.2) and we conclude the following fact. A related result for intergenerational games with finite state and action spaces was given in Alj and Haurie [4, Theorem 5.1].

Corollary 3.3 *If* (C3.1) *and* (C3.3) *hold*, *then there exists a stationary Markov perfect equilibrium.*

Our second main result allows to simplify the form of the equilibrium above.

Theorem 3.4 *Assume* (C3.1) *and* (C3.2). *Then there exists a stationary Markov perfect equilibrium* $\phi_* \in \Phi$ *such that for each* $s \in S$, *the support of the probability measure* $\phi_*(\cdot|s)$ *consists of at most two points in* $A(s)$.

For the existence of a deterministic equilibrium, we need some additional assumptions.

Let μ be an atomless probability measure on $\mathcal{B}(S)$. Let $\mathcal{G}$ be a sub-σ-field of $\mathcal{B}(S)$. Following He and Sun [29], we say that $D \in \mathcal{B}(S)$ is a $\mathcal{G}$-atom or a conditional atom under μ if $\mu(D) > 0$ and for any $D_1 \in \mathcal{B}(S)$, there exists a set $D_2 \in \mathcal{G}$ such that

$$\mu\big((D \cap D_1)\triangle(D \cap D_2)\big) = 0.$$

Intuitively, this means that given the realisation of an event D, the σ-fields $\mathcal{G}$ and $\mathcal{B}(S)$ carry essentially the same information. The definition of a $\mathcal{G}$-atom was used by Dynkin and Evstigneev [23] in their studies of the conditional expectation of correspondences. As noted by He and Sun [29], the definitions of a $\mathcal{G}$-atom in their paper and the works of Dynkin and Evstigneev [23] as well as Jacobs [34, Chap. XIV] are equivalent.

We provide here two examples of σ-fields that have no conditional atoms. Let $S := [0, 1] \times [0, 1]$ and $\mathcal{G} := \mathcal{B}([0, 1]) \otimes \{\emptyset, [0, 1]\}$. We define $\mu := \kappa \otimes \nu$, where κ and ν are probability measures on $[0, 1]$ and ν is atomless. Then μ is atomless and $\mathcal{B}(S)$ has no $\mathcal{G}$-atoms under μ. For a formal proof, the reader is referred to [30, Example 1]. In the second case, let us consider a Borel-measurable partition $(B_j)_{j \in \mathbb{N}}$ of S, i.e., $B_j \in \mathcal{B}(S)$ for each $j \in \mathbb{N}$, $S = \bigcup_{j \in \mathbb{N}} B_j$ and $B_i \cap B_j = \emptyset$ for $i \neq j$. By $\mathcal{G}$ we denote the σ-field generated by this partition. Let $\mu \in \text{Pr}(S)$ be atomless. Then $\mathcal{B}(S)$ has no $\mathcal{G}$-atoms under μ on $\mathcal{B}(S)$. Usually, the notion of conditional atoms is applied to stochastic dynamic decision models or games with the product state space; see for instance Duggan [21] or He and Sun [29, 30].

Theorem 3.5 *Assume that conditions* (C3.1) *and* (C3.2) *are satisfied. Let* $\mathcal{G}$ *be the smallest* σ-field *on* S *such that the action correspondence* $s \mapsto A(s)$ *and the family* $\{\rho(s, a, \cdot) : (s, a) \in \mathbb{K}\}$ *of all density functions are measurable. Assume that the Borel* σ-field $\mathcal{B}(S)$ *has no* $\mathcal{G}$-atoms *under the probability measure* p *and* p *is atomless. Then there exists a deterministic stationary Markov perfect equilibrium.*

Remark 3.6 A version of Theorem 3.2 is Nowak [51, Theorem 4.1], where the following additional restrictive condition is imposed:

(C3.4) The integral $\int_S \max_{a \in A(s)} \rho(s, a, s') p(ds')$ is finite for every $s \in S$.

The following example shows that (C3.4) is not implied by (C3.2).

Example 3.7 Let $S = [-1, 1]$ and $A(s) = A = [0, 1]$. Let p be the uniform distribution on S. We define a density $\rho(s, a, \cdot) = \rho(a, \cdot)$ that is independent of $s \in S$ and does not satisfy condition (C3.4). Let $\rho(0, z) := 0$ for $z \in \{-1\} \cup (0, 1]$ and $\rho(0, z) := 2$ for $z \in (-1, 0]$. Put $\rho(1, z) := 0$ for $z \in [-1, 0]$ and $\rho(1, z) := 2$ for $z \in (0, 1]$. If $a \in (0, 1)$, then we define

$$\rho(a, z) := \begin{cases} 0, & \text{if } z \in [-1, -1 + a], \\ 2, & \text{if } z \in (-1 + a, 0], \\ 0, & \text{if } z \in (0, a - a^2], \\ \frac{2}{a}, & \text{if } z \in (a - a^2, a], \\ 0, & \text{if } z \in (a, 1]. \end{cases}$$

Condition (C3.2) holds since the function $a \mapsto \rho(a, z)$ is continuous for each $z \in S$. Note that $m(z) := \max_{a \in A} \rho(a, z) = 2$ for $z \in (-1, 0]$, $m(-1) = 0$ and $m(z) = 2/z$ for $z \in (0, 1]$. Thus $\int_S m(z) p(dz) = \infty$ and (C3.4) is not satisfied.

Remark 3.8 The assertion of Theorem 3.4 cannot be strengthened, that is, a deterministic stationary Markov perfect equilibrium need not exist. This is shown in Example 5.6 in Sect. 5. Let $\phi^* \in \Phi$ be a stationary Markov perfect equilibrium. Assume that the support of $\phi^*(\cdot|s)$ is a connected subset of $A(s)$ for each $s \in S$. The function $a \mapsto P(s, a, \phi^*)$ is continuous and hence has the Darboux property on the support of $\phi^*(\cdot|s)$. (Recall that the Darboux theorem says that every continuous function $f : X \to \mathbb{R}$ on a compact connected space X has the property that for any $x_1, x_2 \in X$ with $f(x_1) \neq f(x_2)$ and any y between $f(x_1)$ and $f(x_2)$, there exists $x \in X$ such that $f(x) = y$.) This implies that for each $s \in S$, there exists some $a_s \in A(s)$ such that

$$P(s, \phi^*(s), \phi^*) = \int_{A(s)} P(s, a, \phi^*) \phi^*(da|s) = P(s, a_s, \phi^*).$$

A simple modification of the proof of Theorem 3.4, using the above conclusion from Darboux's theorem, yields the existence of a deterministic stationary Markov perfect equilibrium. However, to check the mentioned connectedness condition, one has to know ϕ^*.

Remark 3.9 The existence of deterministic stationary Markov perfect equilibria was proved for some classes of models with one-dimensional state and action spaces and atomless transitions by Harris and Laibson [27] and Balbus et al. [6]. The methods of proof used there do not work in models with many commodities (more general state space). Theorem 3.5 provides some sufficient conditions for the existence of deterministic equilibria in models with a general state space. It is inspired by the approach of He and Sun [29], who dealt with Nash equilibria in standard non-zero-sum discounted stochastic games with general state spaces. However, we emphasise

that He and Sun [29] do not prove the existence of deterministic Markov perfect equilibria. Therefore, Theorem 3.5 is new.

Remark 3.10 Uniqueness of a stationary Markov perfect equilibrium can be proved only for specific models. Namely, Balbus et al. [8] showed that the stochastic optimal growth model with quasi-hyperbolic discounting with the state space $S = [0, \infty)$ and concave transition and reward functions admits a unique solution. Within our framework, we may have to deal with multiple equilibria; see Example 5.6 in Sect. 5. It is rather well known that even in simple economic models, we may encounter multiple deterministic equilibria; see Krusell and Smith [38], Vieille and Weibull [69].

Remark 3.11 Björk and Murgoci [15] analyse time-inconsistent stochastic Markov models in discrete time with finite and infinite horizons. Their approach embraces quasi-hyperbolic discounting as a special case. However, their objective is to provide an extension of the standard Bellman equation in the form of a system of nonlinear equations. In other words, assuming that an equilibrium point exists, they show that the corresponding equilibrium function must satisfy a system of nonlinear equations. They do not provide a proof of the existence of an equilibrium point. In Sect. 6, they even stress that "these issues [existence and uniqueness of equilibrium] are in fact quite complicated".

3.2 Some comments on the proofs and possible extensions

In the proof of Theorem 3.2, we consider a best-response correspondence defined on the quotient space Φ_p of p-equivalence classes of functions in Φ endowed with the weak-star topology. The space Φ_p is compact and convex, and the correspondence has closed and convex values. The existence of a stationary Markov perfect equilibrium then follows from a standard fixed point argument. Note that F is not a convex subset of Φ. Therefore, a similar method does not work for determining deterministic equilibria.

If $\hat{\phi} \in \Phi$ is an equilibrium established in Theorem 3.2, then it is easy to see that for any bounded Borel function $\eta : A(s) \to \mathbb{R}$, there exist a_1, a_2 in the support of the probability measure $\hat{\phi}(\cdot|s)$ and some $\vartheta \in [0, 1]$ such that

$$\int_{A(s)} \eta(a)\hat{\phi}(da|s) = \vartheta \eta(a_1) + (1 - \vartheta)\eta(a_2).$$

Using this observation, we can get Borel mappings f, $g \in F$ and a Borel function $\lambda : S \to [0, 1]$ such that $\{f(s), g(s)\}$ is contained in the support of $\hat{\phi}(\cdot|s)$ for each $s \in S$, and $\phi_*(\cdot|s) = \lambda(s)\delta_{f(s)}(\cdot) + (1 - \lambda(s))\delta_{g(s)}(\cdot)$ satisfies $J^\beta(\hat{\phi})(s) = J^\beta(\phi_*)(s)$ for all $s \in S$. These facts imply that ϕ_* is a stationary Markov perfect equilibrium from the assertion of Theorem 3.4.

If the transition probability q is a convex combination of finitely many atomless probability measures on S with coefficients depending on $(s, a) \in \mathbb{K}$, then a deterministic stationary equilibrium can be obtained by applying Theorem 3.4 and the elimination of randomisation method based on a version of Lyapunov's theorem [41] given by Dvoretzky et al. [22]. (This method was used by Jaśkiewicz and Nowak [35] in the

study of some special cases of the model from the present paper.) In our case, we deal with an infinite family of atomless measures $q(ds'|s, a) = \rho(s, a, s')p(ds')$ indexed by $(s, a) \in \mathbb{K}$, and Lyapunov's theorem is not true for infinitely many measures; see Lyapunov [42]. Therefore, the existence of a deterministic stationary Markov perfect equilibrium under assumptions (C3.1) and (C3.2) with an atomless measure p is problematic. The additional assumption in Theorem 3.5 on the lack of $\mathcal{G}$-atoms allows the elimination of randomisation in an equilibrium obtained in Theorem 3.4 (purification of ϕ_*) thanks to an extension of Lyapunov's theorem given by Dynkin and Evstigneev [23].

Remark 3.12 Discounted dynamic programming problems with an unbounded reward function u are usually studied using a so-called "bounding" or "weighted" Borel-measurable function $\omega : S \rightarrow [1, \infty)$ satisfying the following conditions: for all $(s, a) \in \mathbb{K}$, $|u(s, a)| \leq \omega(s)$ and $\int_S \omega(s')q(ds'|s, a) \leq \hat{\beta}\omega(s)$ for some $\hat{\beta} > 0$ with $\hat{\beta}\beta < 1$. For details, the reader is referred to Hernández-Lerma and Lasserre [32, Sect. 8.3] or to Wessels [71]. It is quite easy to see that the n-stage expected discounted utility

$$J_n^\beta(\phi)(s) := E_s^\phi \left[\sum_{\tau=1}^n \beta^{\tau-1} u(s_\tau, a_\tau) \right], \qquad s_1 = s,$$

converges as $n \rightarrow \infty$ to $J^\beta(\phi)$ uniformly in $\phi \in \Phi$. If we assume in addition that for any $s \in S$ and $a_n \rightarrow a_0$ in $A(s)$ as $n \rightarrow \infty$, we have

$$\lim_{n \rightarrow \infty} \int_S |\rho(s, a_n, s') - \rho(s, a_0, s')|\omega(s')p(ds') = 0,$$

then the main results given in this section can be extended to the more general case with an unbounded utility u. The proofs given in Sect. 7 need only some very simple adaptation. In this way, we can apply our results to examples where having u unbounded is very natural.

4 Examples and an overview of selected literature

In this section, we give a number of examples with several comments. Some of them lie in our theoretical framework from Sects. 2 and 3. In other examples (taken from the literature), our assumptions (e.g. compactness of the action spaces) are not satisfied. However, the examples have solutions in closed form that show the difference between Markov perfect equilibria in models with quasi-hyperbolic discounting and solutions obtained in models with standard discounting via a dynamic programming principle.

In 1969, Samuelson [60] published a seminal paper on portfolio selection and stochastic dynamic programming. He considered a finite-horizon model with a power utility function. His paper inspired many researchers to develop a modern theory of portfolio selection in discrete and continuous time; see Bäuerle and Rieder [11,

Chap. 4], Bobryk and Stettner [16], Merton [46], Shreve and Soner [64] and the references cited therein. For instance, Bobryk and Stettner [16] extended the optimal portfolio selection model of Samuelson [60] to an infinite horizon with standard discounting and completely solved the cases with power and logarithmic utilities (see [16, Proposition 1]). Below we provide a solution for power utility with quasi-hyperbolic discounting.

Example 4.1 We start with the portfolio selection problem of Samuelson [60] that can be viewed as a Markov decision process with $S = \mathbb{R}_+$ and $A = \mathbb{R}_+ \times [0, 1]$, $A(s) = [0, s] \times [0, 1]$. Consider a financial market consisting of a risky asset and a risk-free asset. Assume that there are two investment possibilities: a stock with a random rate of return ε_t in period $t \in T$ and a bank account with a constant (riskless) rate of return r. Assume that (ε_t) is a sequence of i.i.d. random variables taking values in the interval $[-1, \infty)$ and $1 + \varepsilon_t$ has a probability distribution $\mu_\varepsilon \in \mathrm{Pr}(S)$. Moreover, $r \leq E[\varepsilon_t] < \infty$ for every $t \in T$. Denote by s_t and a_t the capital (wealth) and the consumption, respectively, in period $t \in T$. Clearly, $a_t \in [0, s_t]$. The remaining value $s_t - a_t$ is invested. Let $w_t \in [0, 1]$ be the portfolio weight on the risky asset in period $t \in T$. We obviously have $(a_t, w_t) \in A(s_t)$. Starting with an initial wealth $s_1 \in \mathbb{R}_+$, we have the recursive formula

$$s_{t+1} = (s_t - a_t)\big((1 - w_t)(1 + r) + w_t(1 + \varepsilon_t)\big).$$

Then the transition probability q is given by

$$q\big(D\big|s, (a, w)\big) = \int_0^\infty 1_D\Big(\big((1 - w)(1 + r) + wx\big)(s - a)\Big)\mu_\varepsilon(dx), \qquad D \in \mathcal{B}(S),$$

with 1_D the indicator function of the set D. The consumer has a reward (utility) function u that measures his satisfaction of consumption, that is, $u(s, (a, w)) = u(a)$ for all $(s, (a, w)) \in \mathbb{K}$. We assume that $u(a) = a^\sigma$ with $\sigma \in (0, 1)$. Define

$$G(w) = \int_0^\infty \big((1 - w)(1 + r) + wx\big)^\sigma \mu_\varepsilon(dx).$$

Let $w_* \in [0, 1]$ be such that

$$G(w_*) = \max_{w \in [0,1]} G(w).$$

Assume that

$$\gamma_* := \beta G(w_*) < 1.$$

We are going to show that a deterministic stationary equilibrium can be found in the simple subclass F_0 of functions f in F where $f(s) = (cs, w)$ for all $s \in S$, $c \in [0, 1]$ and $w \in [0, 1]$ (linear consumption functions and constant portfolio weights). If $f \in F_0$, then $J^\beta(f)(s)$ is of the form $J^\beta(f)(s) = R(c)s^\sigma$ with $R(c) \geq 0$. The constant $R(c)$ can be found using the well-known equation in discounted dynamic programming, namely (see Hernández-Lerma and Lasserre [31, Sect. 4.2])

$$J^\beta(f)(s) = u(cs) + \beta \int_S J^\beta(f)(s')q(ds'|s, cs) \tag{4.1}$$

for all $s \in S$. Substituting $J^{\beta}(f)(s) = R(c)s^{\sigma}$ into (4.1), we infer that

$$R(c) = c^{\sigma} + \beta G(w)R(c)(1-c)^{\sigma}.$$

Hence,

$$R(c) = \frac{c^{\sigma}}{1 - (1-c)^{\sigma}\beta G(w)}. \tag{4.2}$$

Suppose that all future generations are going to use $f_*(s) = (cs, w_*)$. Then the current self t faces the optimisation problem

$$\sup_{(a,w)\in A(s)} P\big(s,(a,w),f_*\big) = \sup_{(a,w)\in A(s)} \left(a^{\sigma} + \alpha\beta \int_S J^{\beta}(f_*)(s')q\big(ds'\big|s,(a,w)\big)\right).$$

Note that

$$\sup_{(a,w)\in A(s)} P\big(s,(a,w),f_*\big) = \max_{a\in[0,s]} \big(a^{\sigma} + \alpha\gamma_* R_*(c)(s-a)^{\sigma}\big),$$

where $R_*(c) = R(c)$ is given in (4.2) with $w = w_*$. Now note that the function $a \mapsto a^{\sigma} + \alpha\gamma_* R_*(c)(s-a)^{\sigma}$ is strictly concave for each $s > 0$. Therefore, $a(s) \in [0,s]$ that attains the maximum in the above optimisation problem is unique and has the form $a(s) = bs$ with

$$b = \frac{1}{1 + (\alpha\gamma_* R_*(c))^{\frac{1}{1-\sigma}}}, \qquad s \in S. \tag{4.3}$$

From Definition 2.1, it follows that $f_* \in F_0$ is a deterministic Markov perfect equilibrium if (4.3) with $b = c$ has a solution for c in $c = c_{\alpha} \in [0,1]$. Let

$$\Delta_{\alpha}(c) = c - \frac{1}{1 + (\alpha\gamma_* R_*(c))^{\frac{1}{1-\sigma}}}, \qquad s \in S, \alpha \in [0,1].$$

Note that $\Delta_{\alpha}(1) > 0$ and $\Delta_1(c_1) = 0$ for $c_1 = 1 - \gamma_*^{\frac{1}{1-\sigma}}$. Thus for any $\alpha \in (0,1)$, $\Delta_{\alpha}(c_1) < 0$ and there exists $c_{\alpha} \in (c_1, 1)$ such that $\Delta_{\alpha}(c_{\alpha}) = 0$. Obviously, c_{α} is a solution for c to (4.3) with $b = c$. It can be proved that $c_{\alpha} \downarrow c_1$ as $\alpha \uparrow 1$. Clearly, $f_*(s) = (c_{\alpha}s, w_*)$ is a deterministic stationary Markov perfect equilibrium and the expected payoff to each self t is

$$P\big(s, f_*(s), f_*\big) = (c_{\alpha}s)^{\sigma} + \alpha\gamma_* R_*(c_{\alpha})(1-c_{\alpha})^{\sigma} s^{\sigma}.$$

If $\alpha = 1$, then $(c_1 s, w_*)$ is a solution to the dynamic programming portfolio problem with standard discounting; see Bobryk and Stettner [16, Proposition 1]. Clearly, we conclude from the above discussion that a decision maker who faces standard discounting consumes less than one with quasi-hyperbolic discounting. At the same time, he is willing to increase his investments.

The reason that the portfolio selection problem with quasi-hyperbolic discounting in the above example is solvable is that the instantaneous utility function u has the property $u(a_1 a_2) = u(a_1)u(a_2)$ for all $a_1, a_2 \geq 0$. Example 4.1 has also a simple analytical solution for the logarithmic utility case, i.e., when $u(a) = \ln a$. Then $u(a_1 a_2) = u(a_1) + u(a_2)$ for $a_1, \; a_2 > 0$. For details, see Björk and Murgoci [15, Proposition 8.3].

Remark 4.2 The existence of a stationary Markov perfect equilibrium in the above portfolio selection model with quasi-hyperbolic discounting is an open problem if u is a general continuous and concave function. The methods used here and in Björk and Murgoci [15] are not adequate. The class $F_0 \subseteq F$ is too small, because both consumption and portfolio weights may depend on the state variable. If u is bounded or the conditions described in Remark 3.12 are satisfied and (C3.2) holds, then Theorem 3.4 implies the existence of a simple (randomised) stationary Markov perfect equilibrium. However, to ensure that (C3.2) holds, we assume that the probability distribution μ_ε of the random variables $1 + \varepsilon_t$ has a continuous density function g with respect to some $p \in \mathrm{Pr}(S)$ and $A(s) = [0, \overline{a}(s)] \times [\underline{w}(s), 1]$ for all $s \in S$, where $\overline{a}$ and $\underline{w}$ are Borel functions on S such that $0 \leq \overline{a}(s) < s$ and $0 < \underline{w}(s) \leq 1$ for each $s > 0$. The portfolio weight w must be in $[\underline{w}(s), 1]$ for any $s > 0$. This definition of $A(s)$ says that there is an upper limit for consumption for any positive stock capital s. Note that taking $a = s$ by any self would stop the process forever. On the other hand, choosing $w = 0$ in some period would remove the risky asset from the portfolio. We now show how to verify condition (C3.2).

Let $s, \; s' \in S$ and

$$\tilde{A} := \{(a, w) \in A(s) : s' - (1 + r)(1 - w)(s - a) \leq 0\}.$$

Since g is continuous and $(a, w) \in [0, \overline{a}(s)] \times [\underline{w}(s), 1]$, we obtain for the density function the formula

$$\rho\big(s, (a, w), s'\big) = \begin{cases} \dfrac{g\left(\frac{s' - (1+r)(1-w)(s-a)}{(s-a)w}\right)}{(s-a)w}, & \text{if } (a, w) \in A(s) \setminus \tilde{A}, \\ 0, & \text{if } (a, w) \in \tilde{A}. \end{cases}$$

The set $\tilde{A}$ is closed in $A(s)$. The function $(a, w) \mapsto \rho(s, (a, w), s')$ is continuous on $\tilde{A}$ and also on $A(s) \setminus \tilde{A}$. Suppose now that $(a_n, w_n) \in A(s) \setminus \tilde{A}$ for all $n \in \mathbb{N}$ and $(a_n, w_n) \to (a_0, w_0) \in \tilde{A}$ as $n \to \infty$. Then $s' - (1 + r)(1 - w_0)(s - a_0) = 0$. By the continuity of the function g, it follows that

$$\lim_{n \to \infty} \rho\big(s, (a_n, w_n), s'\big) = \frac{g(0)}{(s - a_0)w_0} \geq \rho\big(s, (a_0, w_0), s'\big) = 0.$$

This is sufficient to conclude that the function $(a, w) \mapsto \rho(s, (a, w), s')$ is lower semicontinuous on $A(s)$. By Remark 3.1, condition (C3.2) is satisfied.

Below we give two examples from the literature where a deterministic stationary Markov perfect equilibrium can be calculated analytically. We also discuss modifications which look difficult to solve and present some possible applications of our main results on randomised and deterministic equilibria. Other examples of deterministic

stationary Markov perfect equilibria in closed form can be found in Barro [10], Chatterjee and Eyigungor [19], Krusell et al. [37], Krusell and Smith [38], Laibson [39] and Luttmer and Mariotti [40].

Example 4.3 We consider a stochastic optimal growth model known also as consumption/saving model with the state space $S = \mathbb{R}_+ \times \mathbb{R}_+$. An element of S is denoted by (k, z), where k is capital and z is an exogenous variable. If $(k_t, z_t) \in S$ is the state at date t and a_t is the amount consumed by the decision maker (consumer), then the next state evolves according to the equations

$$k_{t+1} = (1 - \hat{d})k_t + z_t \hat{p}(k_t) - a_t,$$

$$z_{t+1} = \pi(z_t, \varepsilon_{t+1}), \qquad \text{where } a_t \in A(k_t, z_t) := [0, (1 - \hat{d})k_t + z_t \hat{p}(k_t)],$$

for every $t \in T$ and

- $\hat{d} \in [0, 1]$ is the depreciation rate,
- $\hat{p} : \mathbb{R}_+ \to \mathbb{R}_+$ is a concave and increasing production function,
- $\pi : \mathbb{R}_+ \times [\kappa_1, \kappa_2] \to \mathbb{R}_+$ is the law of motion of an exogenous variable with $0 < \kappa_1 < \kappa_2$,
- (ε_t) is a sequence of i.i.d. random variables which have the common distribution $m \in \Pr([\kappa_1, \kappa_2])$,
- $(k_1, z_1) \in S$ is a given initial state.

The satisfaction of the consumer is measured by his utility function $u : \mathbb{R}_+ \to \mathbb{R}$ and depends only on the consumed part, i.e., $u(s, a) = u(a)$ for every $(s, a) \in \mathbb{K}$. Observe that the transition probability q takes the form

$$q\big(D\big|(k, z), a\big) = \int_{\kappa_1}^{\kappa_2} 1_D\big((1 - \hat{d})k + z\hat{p}(k) - a, \pi(z, \xi)\big)m(d\xi), \qquad D \in \mathcal{B}(S),$$

for every $(k, z) \in S$ and $a \in A(k, z)$. Assume that all future selves are going to use a stationary strategy $\phi \in \Phi$. Then the current self faces the optimisation problem, which is independent of period t,

$$\sup_{\nu \in \Pr(A(k,z))} P\big((k, z), \nu, \phi\big)$$

$$= \sup_{\nu \in \Pr(A(k,z))} \int_{A(k,z)} \bigg(u(a)$$

$$+ \alpha\beta \int_{\mathbb{R}_+ \times \mathbb{R}_+} J^\beta(\phi)(k', z')q\big(dk' \times dz'\big|(k, z), a\big)\bigg)\nu(da)$$

$$= \sup_{\nu \in \Pr(A(k,z))} \int_{A(k,z)} \bigg(u(a) + \alpha\beta \int_{\kappa_1}^{\kappa_2} J^\beta(\phi)\big((1 - \hat{d})k + z\hat{p}(k)$$

$$- a, \pi(z, \xi)\big)m(d\xi)\bigg)\nu(da)$$

for $(k, z) \in S$. The function $J^{\beta}(\phi)$ satisfies the equation

$$J^{\beta}(\phi)(k, z)$$
$$= u\big(\phi(k, z)\big) + \beta \int_{\kappa_1}^{\kappa_2} J^{\beta}(\phi)\big((1 - \hat{d})k + z\hat{p}(k) - \phi(k, z), \pi(z, \xi)\big)m(d\xi)$$
$$= \int_{A(k,z)} \Bigg(u(a)$$
$$\qquad\qquad + \beta \int_{\kappa_1}^{\kappa_2} J^{\beta}(\phi)\big((1 - \hat{d})k + z\hat{p}(k) - a, \pi(z, \xi)\big)m(d\xi)\Bigg)\phi(da|k, z)$$

for every $(k, z) \in S$.

The existence of a stationary Markov perfect equilibrium in the above example is an open problem. However, as noted by Maliar and Maliar [44], such a model with specific utility and probability functions admits a closed-form solution. Assume now that $u(a) = \ln a$, $\hat{d} = 1$, $\hat{p}(k) = k^{\sigma}$ with $\sigma \in (0, 1)$ and $z' = \pi(z, \xi) = \xi$. This model is similar to the one studied in Stokey et al. [66, Chap. 10.1] or Acemoglu [1, Chap. 17.1]. It is sufficient to consider the class F of deterministic strategies. Assume that all future selves are going to use a stationary deterministic strategy $f \in F$. Then the optimisation problem for the current self t is of the form

$$\sup_{a \in A(k,z)} P\big((k, z), a, f\big) = \sup_{a \in A(k,z)} \Bigg(\ln a + \alpha\beta \int_{\kappa_1}^{\kappa_2} J^{\beta}(f)(zk^{\sigma} - a, \xi)m(d\xi)\Bigg),$$

where $(k, z) \in S$ and the function $J^{\beta}(f)$ satisfies the equation

$$J^{\beta}(f)(k, z) = \ln f(k, z) + \beta \int_{\kappa_1}^{\kappa_2} J^{\beta}(f)\big(zk^{\sigma} - f(k, z), \xi\big)m(d\xi). \qquad (4.4)$$

For a justification of (4.4), see for instance Hernández-Lerma and Lasserre [31, Sect. 4.2]. It is not difficult to find that the consumption strategy

$$f_{\alpha}^{*}(k, z) = \frac{1 - \beta\sigma}{1 - \beta\sigma + \alpha\beta\sigma}zk^{\sigma}$$

is a deterministic stationary Markov perfect equilibrium. The form of $f_{\alpha}^{*}(k, z)$ does not depend on the probability distribution m. The phenomenon that the optimal consumption strategy is identical for stochastic and deterministic transitions in the above model was also discovered for standard discounting; see Acemoglu [1, Example 17.1]. More precisely, for geometric (standard) discounting, the optimal stationary strategy is

$$f_{1}^{*}(k, z) = (1 - \beta\sigma)zk^{\sigma} \qquad \text{for } (k, z) \in S.$$

Clearly, $f_{1}^{*} = f_{\alpha}^{*}$ for $\alpha = 1$. If $\alpha < 1$, then $f_{\alpha}^{*} > f_{1}^{*}$. This means that the decision maker who uses quasi-hyperbolic discounting plans to save less for the future at every stage when compared to the model with standard discounting. This follows from the fact that such a decision maker in period t is represented by self t, who pays less attention to all future selves by taking into account the discount factor $\alpha\beta$.

The function $J^\beta(f^*)$ is complicated and depends on the logarithmic moments of the random variable ε_t. However, if $m = \delta_1$ (the deterministic case) and $z_1 = 1$, then as noted by Maliar and Maliar [44], we obtain for $k \in \mathbb{R}_+$ that

$$J^\beta(f^*)(k, 1) = \frac{\sigma}{1 - \beta\sigma} \ln k$$
$$+ \frac{1}{1 - \beta} \left(\ln \frac{1 - \beta\sigma}{1 - \beta\sigma + \alpha\beta\sigma} + \frac{\beta\sigma}{1 - \beta\sigma} \ln \frac{\alpha\beta\sigma}{1 - \beta\sigma + \alpha\beta\sigma} \right).$$

Example 4.4 We now present another consumption/saving model with quasi-hyperbolic discounting that can be solved analytically. The decision maker (consumer) observes the state $s = (b, \ell) \in S := \mathbb{R} \times \mathbb{R}$ and chooses an amount $a \in A(s) := A = \mathbb{R}$ for consumption. Here, b is an asset (or a capital) level, ℓ is the labour endowment and $w\ell$ is the labour income. The state evolves according to the equations

$$b_{t+1} = w\ell_t + (1 + r)b_t - a_t,$$
$$\ell_{t+1} = \upsilon\ell_t + \varepsilon_{t+1}, \qquad \text{where } a_t \in A,$$

for every $t \in T$ and

- w is the wage per unit of labour,
- r is the riskless rate of return on asset holdings,
- $\upsilon \in [0, 1]$ and (ε_t) is a sequence of i.i.d. random variables with the normal distribution $\mathcal{N}(0, \sigma^2)$.

Caballero [18] considered the above model with standard discounting in the framework of monetary economics and noticed that it can be solved analytically if the utility function of the consumer is exponential. Next, Maliar and Maliar [43] provided a closed-form solution for quasi-hyperbolic discounting. In fact, Caballero [18] analysed the more general case in which (ℓ_t) is an ARMA process. As in Maliar and Maliar [43], we assume that

$$u(s, a) = u(a) = -\frac{1}{\theta}\exp(-\theta a), \qquad \theta > 0, a \in A, s \in S.$$

The parameter θ is the individual's risk coefficient and reflects his risk attitude. Observe that the transition probability q has the form

$$q\big(D|(b, \ell), a\big) = \int_{-\infty}^{\infty} 1_D\big(w\ell + (1 + r)b - a, \upsilon\ell + \xi\big)g(\xi)d\xi, \qquad D \in \mathcal{B}(S),$$

for $(b, \ell) \in S$ and $a \in A$. Here, g is the density function of the normal distribution $\mathcal{N}(0, \sigma^2)$. Assume that all future selves apply $f \in F$. Then the current self faces the maximisation problem

$$\sup_{a \in A} P\big((b, \ell), a, f\big)$$
$$= \sup_{a \in A} \left(u(a) + \alpha\beta \int_{-\infty}^{\infty} J^\beta(f)\big(w\ell + (1 + r)b - a, \upsilon\ell + \xi\big)g(\xi)d\xi \right)$$

for every $(b, \ell) \in S$. The function $J^\beta(f)$ is a solution of the equation

$$J^\beta(f)(b, \ell) = u\big(f(b, \ell)\big) + \beta \int_{-\infty}^{\infty} J^\beta(f)\big(w\ell + (1+r)b - f(b, \ell), \upsilon\ell + \xi\big)g(\xi)d\xi,$$

where $(b, \ell) \in S$. According to Maliar and Maliar [43, Proposition 1], the deterministic stationary Markov perfect equilibrium is of the form

$$\tilde{f}^*(b, \ell) = rb + \frac{rw}{1+r-\upsilon}\ell - \frac{1}{\theta r}\ln\big(\beta(1+\alpha r)\big) - \frac{\theta r w^2 \sigma^2}{2(1+r-\upsilon)^2}.$$

Moreover,

$$J^\beta(\tilde{f}^*)(b, \ell) = -\frac{1+\alpha r}{\theta \alpha r}\exp\big(-\theta\tilde{f}^*(b, \ell)\big).$$

Obviously, the function $\tilde{f}^*$ is affine in the variables b and ℓ. From the above formula for $\tilde{f}^*$, it follows that a decision maker with discount factors $\alpha < 1$ and $\beta \in (0, 1)$ has the identical consumption strategy as a consumer in the standard discounted decision model with short-run discount factor $\tilde{\alpha} = 1$ and long-run discount factor $\tilde{\beta} = \frac{\beta(1+\alpha r)}{1+r} < \beta$. That is because $\tilde{\beta}(1+r) = \beta(1+\alpha r)$. We also observe that larger values of α and/or β imply a lower amount of consumption.

Examples 4.3 and 4.4 require some comments. First, we do not claim that the derived equilibria are unique. It very often occurs that except for a smooth equilibrium, there may exist equilibria with discontinuous strategies. This fact was reported among others by Krusell and Smith [38]. Moreover, Chatterjee and Eyigungor [19] proved that in a dynamic consumer model with constant relative risk aversion preferences, equilibrium strategies must be discontinuous if the decision maker's net wealth cannot fall below a strictly positive value. Second, the numerical computation of a stationary Markov perfect equilibrium is difficult. Certain numerical methods based on the first order condition and the Euler equation are analysed by Maliar and Maliar [44]. Other specific numerical examples are provided in Chatterjee and Eyigungor [19].

In Example 4.4 and also in the model of Caballero [18], the state and action spaces are unbounded. This actually simplifies calculations. Additional assumptions such as nonnegativity or/and boundedness of variables may lead to discontinuous equilibria. For some states, the solutions may lie on the boundary of the constraint sets.

The next example is a modification of the previous one and examines a model with compact action spaces and nonnegative state variables. This modification, although natural, impedes finding a tractable solution. However, within such a framework, Theorem 3.4 is in force.

Example 4.5 Assume in the previous model that the state $s_{t+1} = (b_{t+1}, \ell_{t+1})$ from $S := \mathbb{R}_+ \times \mathbb{R}_+$ evolves according to the equations

$$b_{t+1} = w\ell_t + (1+r)b_t - a_t + \xi_{t+1},$$
$$\ell_{t+1} = \upsilon\ell_t + \zeta_{t+1}, \qquad \text{where } a_t \in [0, w\ell_t + (1+r)b_t] := A(b_t, \ell_t),$$

for every $t \in T$. Moreover, (ξ_{t+1}) and (ζ_{t+1}) are sequences of nonnegative i.i.d. random variables having continuous densities g_1 and g_2 (respectively) with respect to

Lebesgue measure on $\mathbb{R}_+$. It is also assumed that ξ_{t+1} and ζ_{t+1} are independent for each $t \in T$. Recall that the payoff function is $u(s, a) = u(a) = -\frac{1}{\theta}\exp(-\theta a)$ with $a \in A(b, \ell)$, $s = (b, \ell) \in S$ and the individual's risk coefficient $\theta > 0$. The transition probability q has now the form

$$q\big(D\big|(b, \ell), a\big) = \int_0^\infty \int_0^\infty 1_D\big(w\ell + (1 + r)b - a + x, \upsilon\ell + z\big)g_1(x)g_2(z)dxdz$$

for any $D \in \mathcal{B}(S)$, $a \in A(s)$ and $s = (b, \ell) \in S$.

In order to apply Theorem 3.4, it is sufficient to see that for every $s = (b, \ell) \in S$ and $a \in A(s)$, $q(\cdot|s, a)$ is absolutely continuous with respect to some probability measure p on $\mathcal{B}(S)$ and that condition (C3.2) is satisfied. We first determine the density function $x \mapsto \rho_1(s, a, x)$ of the random variable $w\ell + (1 + r)b - a + \xi_{t+1}$. From the continuity of g_1, it follows that

$$\rho_1(s, a, x) = \begin{cases} g_1(x + a - w\ell - (1 + r)b), & \text{if } x + a - w\ell - (1 + r)b > 0, \\ 0, & \text{if } x + a - w\ell - (1 + r)b \le 0. \end{cases}$$

It is obvious that $\int_0^\infty \rho_1(s, a, x)dx = 1$. Let

$$\tilde{A} := \{a \in A(b, \ell): \ x + a - w\ell - (1 + r)b > 0\}.$$

Let $s = (b, \ell)$ and x be fixed. Assume that $a_n \to a_0$ in $A(s)$ as $n \to \infty$. If $\tilde{A} = \emptyset$, then $\rho_1(s, a_n, x) = 0 \to \rho_1(s, a_0, x) = 0$. Suppose now that $\tilde{A} \neq \emptyset$. If $a_0 \notin \tilde{A}$, then $\liminf_{n\to\infty} \rho_1(s, a_n, x) \ge \rho_1(s, a_0, x) = 0$. If $a_0 \in \tilde{A}$, then $a_n \in \tilde{A}$ for all sufficiently large n, and therefore $\lim_{n\to\infty} \rho_1(s, a_n, x) = \rho_1(s, a_0, x)$ by the continuity of g_1. We have shown that the function $a \mapsto \rho_1(s, a, x)$ is lower semicontinuous on $A(s)$ for each $s \in S$ and $x \ge 0$.

By the continuity of the function g_2, it is easy to see that the density function $z \mapsto \rho_2(\ell, z)$ of the random variable $\upsilon\ell + \zeta_{t+1}$ is $\rho_2(\ell, z) = g_2(z - \upsilon\ell)$, if $z > \upsilon\ell$, and $\rho_2(\ell, z) = 0$, if $z \le \upsilon\ell$.

Let p_1 be the probability measure on $\mathbb{R}_+$ with the density $1/(1 + x)^2$. Let us define

$$\rho(s, a, s') = \rho\big((b, \ell), a, (x, z)\big) := \rho_1\big((b, \ell), a, x\big)\rho_2(\ell, z)(1 + x)^2(1 + z)^2.$$

Then for each set $D \in \mathcal{B}(S)$,

$$q\big(D\big|(b, \ell), a\big) = \int_0^\infty \int_0^\infty 1_D(x, z)\rho\big((b, \ell), a, (x, z)\big)p_1(dx)p_1(dz).$$

This implies that $(x, z) \mapsto \rho((b, \ell), a, (x, z))$ is a density for $q(\cdot|(b, \ell), a)$ with respect to the probability measure $p_1 \otimes p_1$ on S. Obviously, $\rho((b, \ell), a, (x, z))$ is lower semicontinuous in $a \in A(b, \ell)$. By Remark 3.1, condition (C3.2) is satisfied. The instantaneous utility function u is bounded and (C3.1) holds as well. As a consequence, Theorem 3.4 applies. The existence of a deterministic stationary Markov perfect equilibrium in this model is an open problem.

Example 4.6 Assume that there is a single good (called also a renewable resource) that can be used in each period for consumption or productive investment. The set of all resource stocks is $S = \mathbb{R}_+$. Self t observes the current stock $s_t \in S$ and chooses $a_t \in A(s_t) := [0, s_t]$ for consumption. The remaining part $y_t = s_t - a_t$ is left as an investment for future selves. The next self's inheritance or endowment is determined by a transition probability q_0 from S to S (stochastic production function) which depends on $y_t \in A(s_t) \subseteq S$, i.e., $q(\cdot|s_t, a_t) = q_0(\cdot|y_t)$. We assume that

$$s_{t+1} = \pi(s_t - a_t) + \varepsilon_t, \qquad t \in T,$$

where $\pi : \mathbb{R}_+ \to \mathbb{R}_+$ is a continuous increasing function with $\pi(0) = 0$ and (ε_t) is a sequence of i.i.d. nonnegative random variables having no atoms. In Harris and Laibson [27], the function π is linear and the probability distribution of ε_t has a twice differentiable density function with respect to Lebesgue measure on S. Moreover, Harris and Laibson [27] and Balbus et al. [6] assume that $u(s, a) = u(a)$ for all $(s, a) \in \mathbb{K}$, i.e., the instantaneous utility function only depends on the consumption in state $s \in S$. Moreover, u is increasing and strictly concave. Using an additional assumption on the relative risk aversion coefficient for u and some other technical conditions, Harris and Laibson [27] establish the existence of a deterministic stationary Markov perfect equilibrium in the class of functions with locally bounded variation. They also study a stochastic version of the Euler equation associated with this model.

Balbus et al. [6] consider a more general case. They only assume that q_0 is atomless and weakly continuous in investment. Under these assumptions, they show that there exists a deterministic stationary Markov perfect equilibrium in some special class $F_0 \subseteq F$. A function $f \in F$ belongs to the class F_0 if and only if $s - f(s)$ is nondecreasing and continuous from the left. The fact that the probability distribution of ε_t is atomless plays a crucial role in the proof. Moreover, some results from the theory of supermodular functions (see Topkis [68, Chap. 2]) and the Schauder fixed point theorem are applied. Condition (C3.2) need not be satisfied.

Remark 4.7 The existence of a stationary Markov perfect equilibrium in stochastic growth models with weakly continuous transitions is a difficult issue. If we deal with a one-dimensional state space and do not impose any additional conditions on the transition probabilities, then the function

$$y \mapsto \int_S J^\beta(f)(s') q_0(ds'|y)$$

is not concave in general, even if all future selves use a Lipschitz-continuous strategy $f \in F$. The best reply (if one exists) is very often a discontinuous function. Therefore, Balbus et al. [6] and Harris and Laibson [27] consider a class of discontinuous strategies. The assumption that the transitions are atomless is very helpful. However, the techniques used in the above papers do not work for consumption/investment models with multidimensional state spaces (many commodities). The problem is not solved either if we allow the transition probabilities to possess some atoms. In particular, the existence of a stationary Markov perfect equilibrium is an open problem if the transitions are deterministic.

If the state and action spaces are one-dimensional and the transition probability is a convex combination of finitely many probability measures on the state space with coefficients depending on the state–action pairs, then under some stochastic dominance condition, one can prove the existence of a deterministic stationary Markov perfect equilibrium in the class of Lipschitz-continuous strategies with Lipschitz constant one; see Balbus et al. [5, Sect. 3.2]. Below, we give an example of such a model for which it is possible to find a solution in closed form.

Example 4.8 Consider the consumption/investment model with $S = [0, 1]$ and a transition probability q_0 of the form

$$q_0(\cdot|s - a) := \left(1 - (s - a)^\sigma\right)\delta_0(\cdot) + (s - a)^\sigma m_L(\cdot), \qquad \sigma \in (0, 1),$$

where m_L is Lebesgue measure on S. As mentioned above, a denotes the amount of consumption and $a \in A(s) = [0, s]$. The utility function for the consumer is $u(a) = a^\sigma$. Observe also that $q_0(\{0\}|0) = 1$, which means that 0 is an absorbing state and $J^\beta(f)(0) = 0$ for any $f \in F$.

Assume that all future selves are going to use a stationary deterministic consumption strategy $f \in F$. Then current self has to solve the optimisation problem

$$
\begin{aligned}
\sup_{a \in A(s)} P(s, a, f) &= \sup_{a \in A(s)} \left(u(a) + \alpha\beta \int_S J^\beta(f)(s')q_0(ds'|s - a) \right) \\
&= \sup_{a \in A(s)} \left(a^\sigma + \alpha\beta\left((1 - (s - a)^\sigma)J^\beta(f)(0) \right.\right. \\
&\qquad\qquad\qquad \left.\left. + (s - a)^\sigma \int_0^1 J^\beta(f)(s')m_L(ds') \right)\right) \\
&= \sup_{a \in A(s)} \left(a^\sigma + \alpha\beta(s - a)^\sigma \int_0^1 J^\beta(f)(s')m_L(ds') \right)
\end{aligned}
$$

for every $s \in S$, where

$$J^\beta(f)(s) = \left(f(s)\right)^\sigma + \beta\left(s - f(s)\right)^\sigma \int_0^1 J^\beta(f)(s')m_L(ds'), \qquad s \in S.$$

It turns out that in this example, there exists a deterministic stationary Markov perfect equilibrium in the class of linear functions. Consider the subclass $F_0 \subseteq F$ such that $f \in F_0$ if $f(s) = cs$ for some constant $c \in [0, 1]$ and all $s \in S$. Using the above equations and assuming that the equilibrium f^* is in F_0 and that $J^\beta(f^*)(s) = Cs^\sigma$ for $s \in S$ and some constant $C > 0$, we obtain for C the equation

$$C = \frac{\tilde{c}^\sigma(\sigma + 1)}{\sigma + 1 - \beta(1 - \tilde{c})^\sigma} \qquad \text{with} \quad \tilde{c} := \frac{(\frac{\sigma+1}{C\alpha\beta})^{\frac{1}{1-\sigma}}}{1 + (\frac{\sigma+1}{C\alpha\beta})^{\frac{1}{1-\sigma}}}.$$

Clearly, a deterministic stationary Markov perfect equilibrium is $f^*(s) = \tilde{c}s$. For instance, if $\beta = 0.9$, $\alpha = \sigma = 0.5$, then $C = 1.17851$ and $\tilde{c} = 0.888889$. If on the other hand $\beta = 0.9$, $\alpha = 1$ and $\sigma = 0.5$, then $C = 1.25$ and $\tilde{c} = 0.64$. Similarly as

in Example 4.3, a decision maker with quasi-hyperbolic discounting saves less for the future and prefers to consume more in each period than a decision maker with standard discounting. Moreover, we note that the greater β, the smaller the amount that is consumed. If y^σ is replaced by a concave increasing function $\eta(y)$ with range $\eta(S) \subseteq [0, 1]$ and if u is strictly concave and increasing, then from Balbus et al. [5, Theorem 2], we know that a Lipschitz equilibrium exists in the example with the transition function

$$q_0(\cdot|s, a) = \big(1 - \eta(s - a)\big)\delta_0(\cdot) + \eta(s - a)m_L(\cdot),$$

but the calculations look very difficult.

We close this section with an application of Theorem 3.5.

Example 4.9 We now consider a model with a transition probability whose conditional density functions with respect to some atomless probability measure generate a σ-field $\mathcal{G}$ such that the original σ-field on the state space has no $\mathcal{G}$-atoms. Usually, for such examples, the state space is represented as $S := Z \times Y$, where Z and Y are complete separable metric spaces with their Borel σ-fields $\mathcal{B}(Z)$ and $\mathcal{B}(Y)$, respectively. The space S is endowed with the product σ-field. Consider a controller (decision maker) of a certain production process, whose state $s \in S$ consists of two coordinates $s = (z, y)$, where $z \in Z$ is a capital stock and $y \in Y$ is a noise component determining for instance specific technological shocks. Assume that in each period, the controller needs to make a decision $a = (a^1, \ldots, a^m) \in A$ on the intensities for m different production processes. Here, A is a compact subset of $\mathbb{R}_+^m$ and $A(s) = A$ for every $s \in S$. Given the current state (z, y) and an action profile a, the transition law q is determined by

$$q(B|s, a) := \int_Z \int_Y 1_B(z', y')\tilde{\lambda}(dy')q_Z(dz'|s, a), \qquad s = (z, y) \in S, a \in A,$$

where $B \in \mathcal{B}(Z) \otimes \mathcal{B}(Y)$. Here,

$-$ $q_Z(\cdot|s, a)$ denotes the marginal of $q(\cdot|s, a)$ on Z; additionally, $q_Z(\cdot|s, a)$ is absolutely continuous with respect to some $\kappa \in \mathrm{Pr}(Z)$ for every $(s, a) \in S \times A$; it is assumed that the corresponding Radon–Nikodým derivative $\rho(s, a, \cdot)$ is such that $a \mapsto \rho(s, a, z')$ is continuous on A for every $s \in S, z' \in Z$;

$-$ $\tilde{\lambda} \in \mathrm{Pr}(Y)$ is atomless.

Hence, $s = (z, y)$, where z is influenced by the action of the controller and y is a technological shock. Define $p := \kappa \otimes \lambda$ and observe that $\mathcal{G} = \mathcal{B}(Z) \otimes \{\emptyset, Y\}$. Since $\tilde{\lambda}$ is atomless, $\mathcal{B}(Z) \otimes \mathcal{B}(Y)$ has no $\mathcal{G}$-atoms under p.

Let $u : S \times A \to \mathbb{R}$ be a one-period bounded reward function. The controller wishes to find an equilibrium for the infinite-horizon problem with quasi-hyperbolic discounting. From Theorem 3.5, we conclude that there exists a deterministic stationary Markov perfect equilibrium for that problem. For further comments and possible structures of the reward functions, we refer the reader to Duggan [21] and references therein.

It should be mentioned that similar sets of states and transition laws were already considered in the area of standard stochastic games in the context of existence of randomised stationary Nash equilibria; see Duggan [21] and He and Sun [29, 30].

5 Markov perfect equilibria in countable state space models

5.1 Existence of deterministic non-stationary Markov perfect equilibria

In this section, we assume that S is a countable set. We shall prove that in a model with countably many states satisfying assumptions (C3.1) and (C3.3) there exists a deterministic Markov perfect equilibrium.

Let $F^\infty = F \times F \times \cdots$ and let $\bar{f} = (f_1, f_2, \dots) \in F^\infty$ be a sequence of deterministic strategies of all selves. For any $t \in T$, we put $\bar{f}^t := (f_t, f_{t+1}, \dots)$. Let $E_{s_t}^{\bar{f}^t}$ denote the expectation operator with respect to the unique probability measure $P_{s_t}^{\bar{f}^t}$ on the space $(S \times A)^\infty$ of all sequences of state–action pairs, when the process starts at s_t and is induced by $\bar{f}^t \in F^\infty$ and the transition probability q (see the Ionescu-Tulcea theorem in Neveu [48, Proposition V.1.1]).

The *expected utility* of self t is defined as

$$R_t(\bar{f}^t)(s_t) := E_{s_t}^{\bar{f}^t} \left[u(s_t, a_t) + \alpha\beta \sum_{\tau=t+1}^{\infty} \beta^{\tau-t-1} u(s_\tau, a_\tau) \right].$$

Introducing the notation

$$J_{t+1}^\beta(\bar{f}^{t+1})(s_{t+1}) := E_{s_{t+1}}^{\bar{f}^{t+1}} \left[\sum_{\tau=t+1}^{\infty} \beta^{\tau-t-1} u(s_\tau, a_\tau) \right],$$

we obtain that

$$R_t(\bar{f}^t)(s_t) = u(s_t, f_t(s_t)) + \alpha\beta \int_S J_{t+1}^\beta(\bar{f}^{t+1})(s_{t+1}) q(ds_{t+1}|s_t, f_t(s_t)).$$

Furthermore, for any $s \in S$ and $a \in A(s)$, we set

$$P_t(s, a, \bar{f}^{t+1}) := u(s, a) + \alpha\beta \int_S J_{t+1}^\beta(\bar{f}^{t+1})(s_{t+1}) q(ds_{t+1}|s, a).$$

Definition 5.1 A *deterministic Markov perfect equilibrium* is defined to be a sequence $\bar{f} = (f_t)_{t \in T} \in F^\infty$ such that for every $s \in S$ and $t \in T$, we have

$$\sup_{a \in A(s)} P_t(s, a, \bar{f}^{t+1}) = P_t(s, f_t(s), \bar{f}^{t+1}) = R_t(\bar{f}^t)(s).$$

From this definition, it follows that a deterministic Markov perfect equilibrium is subgame perfect.

Theorem 5.2 *Under assumptions* (C3.1) *and* (C3.3) *and if S is countable, there exists a deterministic Markov perfect equilibrium.*

For the proof, we need some lemmas. For any $f \in F$ and any bounded function $v : S \to \mathbb{R}$, we define

$$u_f(s) := u(s, f(s)), \qquad q_f(v)(s) := \sum_{s' \in S} v(s') q(s'|s, f(s)), \qquad s \in S.$$

Then we observe that for any $t \in T$, $s = s_{t+1} \in S$, we have

$$J_{t+1}^{\beta}(\bar{f}^{t+1})(s)$$

$$= u_{f_{t+1}}(s) + \beta q_{f_{t+1}}(u_{f_{t+2}})(s) + \sum_{n=t+2}^{\infty} \beta^{n-t} q_{f_{t+1}} \cdots q_{f_n}(u_{f_{n+1}})(s). \qquad (5.1)$$

Lemma 5.3 *Assume that condition* (C3.3) *holds. Let* $(v_n)_{n\in\mathbb{N}}$ *be a sequence of functions on* S *such that* $\sup_{n\in\mathbb{N}, s\in S} |v_n(s)| < \infty$. *Assume that* $v_n(s') \to v(s')$ *for each* $s' \in S$ *and* $(f_n)_{n\in\mathbb{N}}$ *is a sequence in* F *such that* $f_n(s) \to f(s)$ *for each* $s \in S$ *as* $n \to \infty$. *Then the following statements hold*:

(a) $q_{f_n}(v_n)(s) \to q_f(v)(s)$ *for all* $s \in S$ *as* $n \to \infty$,
(b) $\max_{a \in A(s)} \sum_{s' \in S} v_n(s') q(s'|s, a) \to \max_{a \in A(s)} \sum_{s' \in S} v(s') q(s'|s, a)$ *as* $n \to \infty$.

Proof Part (a) follows directly from Royden [58, Proposition 18]. For (b), note that

$$\left| \max_{a \in A(s)} \sum_{s' \in S} v_n(s') q(s'|s, a) - \max_{a \in A(s)} \sum_{s' \in S} v(s') q(s'|s, a) \right|$$

$$\leq \xi_n(s) := \max_{a \in A(s)} \left| \sum_{s' \in S} \big(v_n(s') - v(s')\big) q(s'|s, a) \right|.$$

Since $A(s)$ is compact metric, there exists a sequence (b_n) of elements of $A(s)$ such that

$$\xi_n(s) = \left| \sum_{s' \in S} \big(v_n(s') - v(s')\big) q(s'|s, b_n) \right| \qquad \text{for all } n \in \mathbb{N},$$

and we can assume without loss of generality that (b_n) is convergent to some $b_0 \in A(s)$. Then Royden [58, Proposition 18] yields that $\xi_n(s) \to 0$ as $n \to \infty$. Hence the proof is complete. $\qquad \square$

The set F of all selectors of the correspondence $s \mapsto A(s)$ (see Sect. 2) can be viewed as the product space $\prod_{s \in S} A(s)$. Then F with the product topology is by Tychonoff's theorem a compact metric space, and so is F^{∞}. We point out that $\bar{f}_n = (f_{n1}, f_{n2}, \dots)$ converges to $\bar{f} = (f_1, f_2, \dots)$ in F^{∞} as $n \to \infty$ if $\lim_{n \to \infty} f_{nk}(s) = f_k(s)$ for all $k \in \mathbb{N}$ and $s \in S$.

Lemma 5.4 *Under assumptions* (C3.1) *and* (C3.3), *the mapping*

$$\bar{f} \mapsto J_{t+1}^{\beta}(\bar{f}^{t+1})(s)$$

is continuous on F^{∞} *for all* $s \in S$ *and* $t \in T$.

Proof Since u is bounded, the series in (5.1) converges uniformly in $\bar{f} \in F^{\infty}$ and $s \in S$. It is sufficient to prove that $\bar{f} \mapsto q_{f_{t+1}} \cdots q_{f_n}(r_{f_{n+1}})$ is continuous on F^{∞} for each $t \in T$ and $n > t + 1$. This can be shown by induction with the help of Lemma 5.3. $\qquad \square$

Let $\bar{g} = (g_1, g_2, \dots) \in F^\infty$ and

$$J^\beta(\bar{g})(s') := E_{s'}^{\bar{g}}\left[\sum_{n=1}^{\infty} \beta^{n-1} u(s_n, a_n)\right]$$

$$= u_{g_1}(s') + \beta q_{g_1}(u_{g_2})(s') + \sum_{n=2}^{\infty} \beta^n q_{g_1} \cdots q_{g_n}(u_{g_{n+1}})(s').$$

For the proof of Theorem 5.2, we define two correspondences. Let $B(\bar{g})$ denote the set of all $f \in F$ such that

$$f(s) \in \arg\max_{a \in A(s)} \left(u(s, a) + \alpha\beta \sum_{s' \in S} J^\beta(\bar{g})(s')q(s'|s, a)\right)$$

for all $s \in S$, and for any $t \in T$, let $B_t(\bar{g}^{t+1})$ be the set of all $f \in F$ such that

$$f(s) \in \arg\max_{a \in A(s)} \left(u(s, a) + \alpha\beta \sum_{s_{t+1} \in S} J^\beta_{t+1}(\bar{g}^{t+1})(s_{t+1})q(ds_{t+1}|s, a)\right)$$

for each $s \in S$.

Proof of Theorem 5.2 Fix any $\bar{f} = (f, f, \dots) \in F^\infty$. Choose some $f_1 \in B(\bar{f})$ and define $\bar{f}_1 := (f_1, f, f, \dots)$. Suppose that $\bar{f}_{n-1} = (f_{n-1}, \dots, f_1, f, f, \dots)$ has been defined for some $n \geq 2$. Choose any $f_n \in B(\bar{f}_{n-1})$ and define

$$\bar{f}_n = (f_n, f_{n-1}, \dots, f_1, f, f, \dots).$$

Since F^∞ is a compact metric space, the sequence $(\bar{f}_n)$ has a subsequence $(\bar{f}_{n'})$ converging to some $\bar{f}_0 = (f_{01}, f_{02}, f_{03}, \dots)$ as $n' \to \infty$. Denote this subsequence by $\bar{f}_{n'} = (f_{n'1}, f_{n'2}, f_{n'3}, \dots)$. Take any $t \in T$. Observe that

$$f_{n't} \in B_t(\bar{f}_{n'}^{t+1}) \qquad \text{for each } t \in T, n' > t. \tag{5.2}$$

Since $\bar{f}_{n'} \to \bar{f}_0$, we have $\bar{f}_{n'}^{t+1} \to \bar{f}_0^{t+1}$ as $n' \to \infty$. Using (5.2), Lemmas 5.3 and 5.4, one can easily deduce that $f_{0t} \in B_t(\bar{f}_0^{t+1})$ for each $t \in T$, that is, $\bar{f}_0$ is a deterministic Markov perfect equilibrium. $\qquad\square$

Remark 5.5 (a) Theorem 5.2 is new and we apply in its proof a backward induction method similar to that used in standard dynamic programming (see for instance Hernández-Lerma and Lasserre [31, Sect. 3.2] or Puterman [57, Sect. 4.5]) or finite-horizon models with quasi-hyperbolic discounting (see Alj and Haurie [4], Bernheim and Ray [12] or Goldman [26]). In our setup, this method determines a sequence of strategies in F^∞ which need not be convergent and may have many accumulation points. In Example 5.6 below, we present two accumulation points that give different deterministic Markov perfect equilibria.

(b) It is worth noticing that the existence proof of Theorem 5.2 is not based on any fixed point argument. A similar "iterative method" gives in Balbus et al. [7] a non-stationary deterministic Markov perfect equilibrium in a model with quasi-hyperbolic discounting and one-dimensional state and action spaces satisfying some additional conditions similar to that in Balbus et al. [6] and Harris and Laibson [27]. The deterministic equilibrium obtained in [7] belongs to the intersection of a decreasing family of closed sets of strategies for the decision maker. The methods used in this section and in [7] have many limitations in the sense that they work only under some special conditions on the primitive data of the model.

(c) Theorem 5.2 can be extended (with minor changes in the proof) to the unbounded reward case discussed in Remark 3.12.

5.2 An example with a finite state space

As noted in Sect. 3, a stationary Markov perfect equilibrium is a fixed point of a best-response correspondence defined on a compact convex set. The set F of deterministic strategies is not convex. Thus an argument based on fixed point theorems is difficult to apply. This suggests that a deterministic stationary Markov perfect equilibrium need not exist even in simple models. Below we provide an example of a Markov decision process with finite state and action spaces in which a deterministic stationary Markov perfect equilibrium does not exist. This example also shows two different deterministic Markov perfect equilibria being two accumulation points of a sequence of strategies as in the proof of Theorem 5.2.

Example 5.6 The state space is $S = \{1, 2\}$ and the action sets are $A(1) = \{a, b\}$, $A(2) = \{a\}$. The transition probabilities are defined as $q(2|1, a) = 1 = q(1|2, a)$ and $q(1|1, b) = q(2|1, b) = 0.5$. The set F of all deterministic stationary strategies consists of two elements f and g, where $f(1) = a$ and $g(1) = b$. For simplicity, we apply standard matrix/vector notation. Any function $w : S \to \mathbb{R}$ is written as the column vector $(w(1), w(2))^T$. Transition probabilities are given by stochastic matrices. For any $\phi \in \Phi$, the function $u(\cdot, \phi(\cdot))$ is denoted by the column payoff vector U_ϕ. By Q_ϕ, we denote the transition probability matrix induced by ϕ.

For any stationary strategy profile $(\phi, \phi, \dots)$ with $\phi \in \Phi$, we write $J^\beta(\phi) = J^\beta_\phi$ as a column vector. Assume that $\beta = 0.8$ and $\alpha = 0.5$. To compute J^β_ϕ, we use the well-known formula (see Neyman [49, Lemma 1(b)])

$$J^\beta_\phi = (I - \beta Q_\phi)^{-1} U_\phi,$$

where I is the identity matrix.

Note that we have

$$Q_f = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad (I - \beta Q_f)^{-1} = \begin{pmatrix} \frac{25}{9} & \frac{20}{9} \\ \frac{20}{9} & \frac{25}{9} \end{pmatrix},$$

$$Q_g = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad (I - \beta Q_g)^{-1} = \begin{pmatrix} \frac{25}{7} & \frac{10}{7} \\ \frac{20}{7} & \frac{15}{7} \end{pmatrix}.$$

We now show that there exists a stationary Markov perfect equilibrium $\phi \in \Phi$ and there are no deterministic ones. Let the reward vectors be given by

$$U_f = \begin{pmatrix} 0 \\ 17 \end{pmatrix} \qquad \text{and} \qquad U_g = \begin{pmatrix} 2 \\ 17 \end{pmatrix}.$$

Immediately, we get

$$J_f^\beta = \begin{pmatrix} \frac{340}{9} \\ \frac{425}{9} \end{pmatrix} \qquad \text{and} \qquad J_g^\beta = \begin{pmatrix} \frac{220}{7} \\ \frac{295}{7} \end{pmatrix}.$$

It is easy to check that $B_f = \{g\}$ and $B_g = \{f\}$. Thus there is no deterministic stationary Markov perfect equilibrium. Indeed, it follows that

$$U_f + \alpha\beta Q_f J_f^\beta = \begin{pmatrix} \frac{170}{9} \\ \frac{289}{9} \end{pmatrix} \qquad \text{and} \qquad U_g + \alpha\beta Q_g J_f^\beta = \begin{pmatrix} \frac{171}{9} \\ \frac{289}{9} \end{pmatrix},$$

which implies that in state $s = 1$, the better strategy is g if we assume that all future selves are going to use f. Similarly, we obtain

$$U_f + \alpha\beta Q_f J_g^\beta = \begin{pmatrix} \frac{118}{7} \\ \frac{207}{7} \end{pmatrix} \qquad \text{and} \qquad U_g + \alpha\beta Q_g J_g^\beta = \begin{pmatrix} \frac{117}{7} \\ \frac{207}{7} \end{pmatrix}.$$

Hence the best response of the current self is f if all future selves are going to use the stationary deterministic strategy g.

It is easy to check that $\phi \in \Phi$ with $\phi(1) = \phi(2) = 0.5$ is a stationary Markov perfect equilibrium. We have

$$Q_\phi = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ 1 & 0 \end{pmatrix} \qquad \text{and} \qquad U_\phi = \begin{pmatrix} 1 \\ 17 \end{pmatrix},$$

and

$$J_\phi^\beta = (I - \beta Q_\phi)^{-1} U_\phi = \begin{pmatrix} \frac{25}{8} & \frac{15}{8} \\ \frac{20}{8} & \frac{20}{8} \end{pmatrix} \begin{pmatrix} 1 \\ 17 \end{pmatrix} = \begin{pmatrix} 35 \\ 45 \end{pmatrix}.$$

Next, we have

$$U_f + \alpha\beta Q_f J_\phi^\beta = U_g + \alpha\beta Q_g J_\phi^\beta = \begin{pmatrix} 18 \\ 31 \end{pmatrix}.$$

Hence it follows that $B_\phi = \Phi$, i.e., every $\hat\phi \in \Phi$ is a best reply to ϕ. In particular, $\phi \in B_\phi$, i.e., ϕ is a stationary Markov perfect equilibrium.

Consider now two deterministic "periodic" strategy profiles

$$\overline{fg} = (f, g, f, g, \dots) \qquad \text{and} \qquad \overline{gf} = (g, f, g, f, \dots).$$

The discounted expected reward vector over the infinite horizon under $\overline{fg}$ is

$$
\begin{aligned}
J_{\overline{fg}}^{\beta} &= U_f + \beta Q_f U_g + \beta^2 Q_f Q_g U_f + \beta^3 Q_f Q_g Q_f U_g + \beta^4 Q_f Q_g Q_f Q_g U_f + \cdots \\
&= \left(I + \beta^2 Q_f Q_g U_f + \beta^4 (Q_f Q_g)^2 U_f + \cdots \right)(U_f + \beta Q_f U_g) \\
&= \left(\sum_{n=0}^{\infty} \beta^{2n} (Q_f Q_g)^n \right)(U_f + \beta Q_f U_g) = (I - \beta^2 Q_f Q_g)^{-1}(U_f + \beta Q_f U_g),
\end{aligned}
$$

where $(Q_f Q_g)^0 = I$. Therefore, we have

$$
J_{\overline{fg}}^{\beta} = \begin{pmatrix} \frac{25}{9} & 0 \\ \frac{200}{153} & \frac{25}{17} \end{pmatrix} \left(\begin{pmatrix} 0 \\ 17 \end{pmatrix} + \begin{pmatrix} 0 & \frac{4}{5} \\ \frac{4}{5} & 0 \end{pmatrix} \begin{pmatrix} 2 \\ 17 \end{pmatrix} \right) = \begin{pmatrix} \frac{340}{9} \\ \frac{6905}{153} \end{pmatrix}.
$$

Proceeding in an analogous way, we may calculate the expected discounted reward vector $J_{\overline{gf}}^{\beta}$ when the Markov strategy profile $\overline{gf} = (g, f, g, f, \ldots)$ is applied. We get

$$
J_{\overline{gf}}^{\beta} = (I - \beta^2 Q_g Q_f)^{-1}(U_g + \beta Q_g U_f) = \begin{pmatrix} \frac{5380}{153} \\ \frac{425}{9} \end{pmatrix}.
$$

Suppose that the selves following self t are going to employ the strategy profile $\overline{fg}$. Then the rewards for self t using f or g are

$$
U_f + \alpha \beta Q_f J_{\overline{fg}}^{\beta} = \begin{pmatrix} \frac{2762}{153} \\ \frac{289}{9} \end{pmatrix} \qquad \text{and} \qquad U_g + \alpha \beta Q_g J_{\overline{fg}}^{\beta} = \begin{pmatrix} \frac{2843}{153} \\ \frac{289}{9} \end{pmatrix}.
$$

In state $s = 1$, it is better for self t to use g. In state $s = 2$, the rewards are same. Assuming that the selves following self t are going to apply $\overline{gf}$, we obtain for self t the rewards

$$
U_f + \alpha \beta Q_f J_{\overline{gf}}^{\beta} = \begin{pmatrix} \frac{170}{9} \\ \frac{4753}{153} \end{pmatrix} \qquad \text{and} \qquad U_g + \alpha \beta Q_g J_{\overline{gf}}^{\beta} = \begin{pmatrix} \frac{2827}{153} \\ \frac{4753}{153} \end{pmatrix}.
$$

In this setup, it is better for self t to use in state $s = 1$ the strategy f. From these calculations, we conclude that both profiles $\overline{fg}$ and $\overline{gf}$ are deterministic Markov perfect equilibria.

It is interesting to note that in both cases, the Markov equilibria $\overline{fg}$ and $\overline{gf}$ give higher rewards than in the stationary randomised equilibrium obtained above.

Moreover, if the Markov decision process starts in state $s = 1$, the equilibrium profile $\overline{fg}$ is more advantageous for the decision maker since $170/9 > 2843/153$. For the initial state $s = 2$, the profile $\overline{gf}$ is better since $289/9 > 4753/153$.

6 Approximate deterministic Markov perfect equilibria in Borel state space models

It is well known that if p is atomless, then the set of all p-equivalence classes of mappings in F is dense in Φ_p; see Warga [70, Theorem IV.3.10]. Therefore, the limit in the weak-star topology on Φ_p of a sequence of deterministic strategies may be a randomised strategy; see Elliott et al. [25, Example 3.16]. This implies that the approach taken in the proof of Theorem 5.2 for a countable state space cannot be extended to a model with a Borel set of states. However, similarly as in stochastic games (see Nowak [50] and Whitt [72]), one can think about an approximation of a Markov decision process on a Borel state space by processes with countably many states. In other words, for a Borel state space, deterministic Markov perfect equilibria in the approximating model can be used to obtain deterministic Markov ϵ-equilibria in the original model.

Definition 6.1 Let $\epsilon > 0$. A *deterministic Markov perfect ϵ-equilibrium* is a sequence $\bar{f} = (f_t)_{t \in T} \in F^\infty$ such that for every $s \in S$ and $t \in T$, we have

$$\sup_{a \in A} P_t(s, a, \bar{f}^{t+1}) \le P_t\big(s, f_t(s), \bar{f}^{t+1}\big) + \epsilon.$$

Let $C(A)$ denote the Banach space of all continuous functions on A endowed with the supremum norm $\| \cdot \|_c$. In this section, we study the Borel state space decision model, denoted by $\mathcal{M}$, satisfying the following assumption:

(C6.1) A is a compact metric space, $A(s) = A$ for all $s \in S$, (C3.1) holds and the transition q has a Borel density function $\rho : S \times A \times S \to \mathbb{R}$ with respect to p satisfying (C3.4) and such that $\rho(s, \cdot, s') \in C(A)$ for all $s, s' \in S$.

Theorem 6.2 below is new for decision models with quasi-hyperbolic discounting. It is based on Theorem 5.2 and modified arguments from the works of Nowak [50] and Whitt [72] on stochastic games. The result cannot be obtained by an approximation of the original model by models with finite horizons.

Theorem 6.2 *Assume that* (C6.1) *holds. Then for any $\epsilon > 0$, there exists a deterministic Markov perfect ϵ-equilibrium in the model $\mathcal{M}$.*

Proof As noted in Nowak [50, Lemma 4.2], under condition (C6.1), one can construct for any $\delta > 0$ a measurable partition $(S_j)_{j \in \mathbb{N}_o}$ of the state space, where $\mathbb{N}_o \subseteq \mathbb{N}$ and $S_j \in \mathcal{B}(S)$ for each $j \in \mathbb{N}_o$, and functions $u_j : A \to \mathbb{R}$, $\rho_j : A \times S \to [0, \infty)$ such that $u_j \in C(A)$ and $\rho_j(\cdot, s') \in C(A)$ for all $j \in \mathbb{N}_o$ and

$$\|u(s, \cdot) - u_j(\cdot)\|_c + \int_S \|\rho(s, \cdot, s') - \rho_j(\cdot, s')\|_c p(ds') < \delta \qquad (6.1)$$

for every $s \in S_j$ and all $j \in \mathbb{N}_o$. Moreover, $\rho_j(a, \cdot)$ is a density function, i.e., $\int_S \rho_j(a, s') p(ds') = 1$ for all $j \in \mathbb{N}_o$ and $a \in A$. The transition probability in the approximating model is

$$\tilde{q}(B|s, a) := \int_B \rho_j(a, s') p(ds') \qquad \text{for } B \in \mathcal{B}(S) \text{ and } s \in S_j,$$

and the reward function $\tilde{u}$ is $\tilde{u}(s,a) = u_j(a)$ for $s \in S_j$ and $a \in A$. We denote the Markov decision process with u_j and ρ_j satisfying (6.1) by $\mathcal{M}^\delta$.

Let $\bar{f} = (f_1, f_2, \ldots)$ be an arbitrary sequence in F^∞. We define the corresponding reward functions in $\mathcal{M}^\delta$ as follows. For $s \in S_j$, $j \in \mathbb{N}_o$, and $t \in T$, we put

$$\tilde{R}_t(\bar{f}^t)(s) := u_j\big(f_t(s)\big) + \alpha\beta \int_S \tilde{J}^\beta_{t+1}(\bar{f}^{t+1})(s')\tilde{q}\big(ds'\big|s, f_t(s)\big),$$

where

$$\tilde{J}^\beta_{t+1}(\bar{f}^{t+1})(s') := \tilde{E}^{\bar{f}^{t+1}}_{s'}\left[\sum_{\tau=t+1}^\infty \beta^{\tau-t-1} \sum_{j\in\mathbb{N}_o} 1_{S_j}(s_\tau)u_j(a_\tau)\right].$$

Here, $\tilde{E}^{\bar{f}^{t+1}}_{s'}$ denotes the expectation operator with respect to the unique probability measure on $(S \times A)^\infty$ which is well defined by the Ionescu-Tulcea theorem; see Neveu [48, Proposition V.1.1]. This measure is induced by the transition probability $\tilde{q}$ and $\bar{f}^{t+1} \in F^\infty$ when the state in period $t+1$ is s'.

Let $\|\cdot\|$ be the supremum norm on the space of all bounded Borel functions on S and suppose that $|u(s,a)| \leq C$ for all $(s,a) \in \mathbb{K}$ and some constant $C > 0$. Then by minor modifications of Nowak [50, proofs of Lemmas 4.3 and 4.4], we can deduce that

$$\|J^\beta_{t+1}(\bar{f}^{t+1}) - \tilde{J}^\beta_{t+1}(\bar{f}^{t+1})\| \leq \frac{\delta(1 + \beta(C-1))}{(1-\beta)^2}.$$

This fact and condition (6.1) imply that for every $s \in S$, we have

$$\left|\int_S J^\beta_{t+1}(\bar{f}^{t+1})(s')q\big(ds'\big|s, f_t(s)\big) - \int_S \tilde{J}^\beta_{t+1}(\bar{f}^{t+1})(s')\tilde{q}\big(ds'\big|s, f_t(s)\big)\right|$$

$$\leq \left|\int_S J^\beta_{t+1}(\bar{f}^{t+1})(s')q\big(ds'\big|s, f_t(s)\big) - \int_S J^\beta_{t+1}(\bar{f}^{t+1})(s')\tilde{q}\big(ds'\big|s, f_t(s)\big)\right|$$

$$+ \left|\int_S J^\beta_{t+1}(\bar{f}^{t+1})(s')\tilde{q}\big(ds'\big|s, f_t(s)\big) - \int_S \tilde{J}^\beta_{t+1}(\bar{f}^{t+1})(s')\tilde{q}\big(ds'\big|s, f_t(s)\big)\right|$$

$$\leq \frac{\delta C}{1-\beta} + \frac{\delta(1 + \beta(C-1))}{(1-\beta)^2} = \frac{\delta(1 - \beta + C)}{(1-\beta)^2}.$$

Consequently, for every $t \in T$, we have

$$\|R_t(\bar{f}^t) - \tilde{R}_t(\bar{f}^t)\| \leq \delta\left(1 + \alpha\beta\frac{1 - \beta + C}{(1-\beta)^2}\right),$$

which means that for any $f \in F$,

$$\left\|P_t\big(\cdot, f(\cdot), \bar{f}^{t+1}\big) - \tilde{P}_t\big(\cdot, f(\cdot), \bar{f}^{t+1}\big)\right\| \leq \delta\left(1 + \alpha\beta\frac{1 - \beta + C}{(1-\beta)^2}\right). \tag{6.2}$$

Here,

$$\tilde{P}_t\big(s, f(s), \bar{f}^{t+1}\big) := u_j\big(f(s)\big) + \alpha\beta \int_S \tilde{J}^\beta_{t+1}(\bar{f}^{t+1})(s')\tilde{q}\big(ds'\big|s, f(s)\big),$$

for $s \in S_j$, $j \in \mathbb{N}_o$. Observe that the constant on the right-hand side of (6.2) is independent of t.

Clearly, the approximating model $\mathcal{M}^\delta$ induces a Markov decision process with countable state space $\mathbb{N}_o$ and transition probability $\tilde{q}(k|j,a) = \int_{S_k} \rho_j(a,s') p(ds')$, which is continuous on A for every $j,k \in \mathbb{N}_o$. This countable state space model will also be denoted by $\mathcal{M}^\delta$. Let $\tilde{F}$ be the space of piecewise constant functions $\tilde{f} : S \to A$ defined as follows. A function $\tilde{f}$ belongs to $\tilde{F}$ if for each $j \in \mathbb{N}_o$ there exists $m_j \in A$ such that $\tilde{f}(s) = m_j$ for all $s \in S_j$. Clearly, $\tilde{F} \subseteq F$. Hence, a deterministic Markov strategy for the decision maker in $\mathcal{M}^\delta$ is a sequence $(\tilde{f}_t) \in \tilde{F}^\infty$.

Choose $\delta > 0$ such that

$$\delta \left(1 + \alpha\beta \frac{1 - \beta + C}{(1 - \beta)^2} \right) \leq \frac{\epsilon}{2}.$$

From Theorem 5.2, we conclude that there exists a deterministic Markov perfect equilibrium $\bar{g} = (g_t) \in \tilde{F}^\infty$ in the model $\mathcal{M}^\delta$. We claim that $\bar{g}$ is an ϵ-equilibrium in the model $\mathcal{M}$. Fix an arbitrary function $f \in F$. From (6.2) and the definition of $\bar{g}$, it follows that for every $t \in T$ and $s \in S$

$$
\begin{aligned}
P_t\big(s, f(s), \bar{g}^{t+1}\big) &\leq \frac{\epsilon}{2} + \tilde{P}_t\big(s, f(s), \bar{g}^{t+1}\big) \\
&\leq \frac{\epsilon}{2} + \sup_{a \in A} \tilde{P}_t(s, a, \bar{g}^{t+1}) = \frac{\epsilon}{2} + \tilde{P}_t\big(s, g_t(s), \bar{g}^{t+1}\big) \\
&\leq \epsilon + P_t\big(s, g_t(s), \bar{g}^{t+1}\big).
\end{aligned}
$$

This proves our claim. $\qquad \square$

7 Proofs of Theorems 3.2, 3.4 and 3.5

Let $\varphi \in \Phi$ and $v : S \to \mathbb{R}$ be a bounded Borel function. We define

$$u_\varphi(s) := u\big(s, \varphi(s)\big) = \int_{A(s)} u(s,a)\varphi(da|s),$$

$$q_\varphi(v)(s) := \int_S v(s')q\big(ds'\big|s, \varphi(s)\big), \qquad s \in S.$$

Let q_φ^n be the composition of q_φ with itself n times. Then $J^\beta(\varphi)(s')$ (defined in (2.4) with $\varphi = \phi$) can be expressed as

$$J^\beta(\varphi)(s') = u_\varphi(s') + \sum_{n=1}^{\infty} \beta^n q_\varphi^n(u_\varphi)(s'), \qquad s' \in S. \tag{7.1}$$

Let $L^1(S,p)$ be the Banach space of all absolutely integrable (with respect to p) functions on S and $L^\infty(S,p)$ the space of all p-essentially bounded functions on S. We endow $L^\infty(S,p)$ with the weak-star topology. Recall that $v_n \to^* v_0$ as $n \to \infty$, i.e., a sequence (v_n) converges to v_0 weak-star in $L^\infty(S,p)$, if and only if $\int_S v_n(s)h(s)p(ds) \to \int_S v_0(s)h(s)p(ds)$ for every $h \in L^1(S,p)$.

A function $c : \mathbb{K} \to \mathbb{R}$ is *Carathéodory* (is a *C-function*) if it is Borel on $\mathbb{K}$, $c(s, \cdot)$ is continuous on $A(s)$ for each $s \in S$ and

$$\int_S \max_{a \in A(s)} |c(s, a)| p(ds) < \infty.$$

Let Φ_p be the space of p-equivalence classes of functions in Φ. The elements of Φ_p are called *Young measures*. The space Φ_p is endowed with the weak-star topology. Since $\mathcal{B}(S)$ is countably generated, Φ_p is metrisable. Moreover, since every set $A(s)$ is compact, Φ_p is a *compact convex* subset of a locally convex linear topological space. For a detailed discussion of these issues, see Balder [9, Theorem 1] or Warga [70, Chap. IV]. We recall that $\phi_n \to^* \phi_0$ in Φ_p if and only if for every C-function $c : \mathbb{K} \to \mathbb{R}$, it holds that

$$\lim_{n \to \infty} \int_S \int_{A(s)} c(s, a) \phi_n(da|s) p(ds) = \int_S \int_{A(s)} c(s, a) \phi_0(da|s) p(ds).$$

Lemma 7.1 *Assume that $v_n \to^* v_0$ in $L^\infty(S, p)$ and $\varphi_n \to^* \varphi_0$ in Φ_p as $n \to \infty$. Then under assumption* (C3.2), *it follows that $q_{\varphi_n}(v_n) \to^* q_{\varphi_0}(v_0)$ in $L^\infty(S, p)$ as $n \to \infty$.*

Proof Take any $h \in L^1(S, p)$. We have

$$\left| \int_S q_{\varphi_n}(v_n)(s) h(s) p(ds) - \int_S q_{\varphi_0}(v_0)(s) h(s) p(ds) \right|$$

$$\leq \left| \int_S \big(q_{\varphi_n}(v_n)(s) - q_{\varphi_n}(v_0)(s) \big) h(s) p(ds) \right|$$

$$+ \left| \int_S \big(q_{\varphi_n}(v_0)(s) - q_{\varphi_0}(v_0)(s) \big) h(s) p(ds) \right|. \tag{7.2}$$

The second term on the right-hand side in (7.2) converges to zero since $\varphi_n \to^* \varphi_0$ as $n \to \infty$. Observe that

$$|q_{\varphi_n}(v_n)(s) - q_{\varphi_n}(v_0)(s)|$$

$$\leq \int_{A(s)} \left| \int_S \big(v_n(s') - v_0(s') \big) \rho(s, a, s') p(ds') \right| \varphi_n(da|s)$$

$$\leq M_n(s) := \max_{a \in A(s)} \left| \int_S \big(v_n(s') - v_0(s') \big) \rho(s, a, s') p(ds') \right|. \tag{7.3}$$

The fact that $M_n(s) \to 0$ for every $s \in S$ as $n \to \infty$ follows from Nowak and Raghavan [52, proof of Lemma 7]. For the sake of completeness, we provide a short argument here. For any $n \in \mathbb{N}$, we can find $a_n \in A(s)$ that attains the maximum in (7.3). Without loss of generality, we can assume that $a_n \to a_0 \in A(s)$ as $n \to \infty$. Note that

$$0 \leq M_n(s) \leq \left| \int_S \big(v_n(s') - v_0(s') \big) \rho(s, a_0, s') p(ds') \right|$$

$$+ \left| \int_S \big(v_n(s') - v_0(s') \big) \big(\rho(s, a_n, s') - \rho(s, a_0, s') \big) p(ds') \right|. \tag{7.4}$$

The first term on the right-hand side in (7.4) converges to zero since $v_n \to^* v_0$ in $L^\infty(S, p)$. Clearly, there exists some constant b such that $|v_n(s') - v_0(s')| \le b$ p-a.e. Thus

$$\left| \int_S (v_n(s') - v_0(s'))(\rho(s, a_n, s') - \rho(s, a_0, s')) p(ds') \right|$$

$$\le b \int_S |\rho(s, a_n, s') - \rho(s, a_0, s')| p(ds').$$

Using this inequality and (C3.2) we conclude that $M_n(s) \to 0$ for any $s \in S$ as $n \to \infty$. Obviously, $\int_S M_n(s) h(s) p(ds) \to 0$ as $n \to \infty$. This property together with (7.2) and (7.3) completes the proof. $\square$

Lemma 7.2 *If* (C3.1) *and* (C3.2) *hold and* $\varphi_k \to^* \varphi_0 \in \Phi_p$, *then* $J^\beta(\varphi_k) \to^* J^\beta(\varphi_0)$ *in* $L^\infty(S, p)$. *In particular,* $\int_B J^\beta(\varphi_k)(s) p(ds) \to \int_B J^\beta(\varphi_0)(s) p(ds)$ *as* $k \to \infty$ *for all* $B \in \mathcal{B}(S)$.

Proof Note that the series in (7.1) is convergent uniformly on $\Phi \times S$. Obviously, $r_{\varphi_k} \to^* r_{\varphi_0}$ in $L^\infty(S, p)$ as $k \to \infty$. By induction and Lemma 7.1, for each $n \in \mathbb{N}$, $q_{\varphi_k}^n(u_{\varphi_k}) \to^* q_{\varphi_0}^n(u_{\varphi_0})$ in $L^\infty(S, p)$ as $k \to \infty$. Thus the lemma follows. $\square$

For any $\varphi \in \Phi_p$, we define the correspondence

$$\varphi \mapsto BR_p(\varphi) := \left\{ \psi \in \Phi_p : \psi(s) \in \underset{v \in \mathrm{Pr}(A(s))}{\arg\max} \ P(s, v, \varphi) \ p\text{-a.e.} \right\}.$$

Lemma 7.3 *If* (C3.1) *and* (C3.2) *hold, the correspondence* $\varphi \mapsto BR_p(\varphi)$ *has a closed graph.*

Proof Due to measurable selection theorems (see Brown and Purves [17]), we have that $BR_p(\varphi) \ne \emptyset$ for each $\varphi \in \Phi_p$. Suppose that $\varphi_n \to^* \varphi_0$ in Φ_p as $n \to \infty$. Assume that $\phi_n \in BR_p(\varphi_n)$ for all $n \in \mathbb{N}$. Since Φ_p is compact metric, we can assume without loss of generality that $\phi_n \to^* \phi_0 \in \Phi_p$ as $n \to \infty$. By Lemma 7.2, $J^\beta(\varphi_n) \to^* J^\beta(\varphi_0)$ in $L^\infty(S, p)$ as $n \to \infty$. Using the arguments from the proof of Lemma 7.1, we can show that

$$\max_{a \in A(s)} \left| \int_S J^\beta(\varphi_n)(s') q(ds'|s, a) - \int_S J^\beta(\varphi_0)(s') q(ds'|s, a) \right|$$

$$= \max_{a \in A(s)} \left| \int_S (J^\beta(\varphi_n)(s') - J^\beta(\varphi_0)(s')) \rho(s, a, s') p(ds') \right| \longrightarrow 0 \qquad (7.5)$$

as $n \to \infty$. Recall (2.5) and note that

$$\left| \int_B \max_{v \in \mathrm{Pr}(A(s))} P(s, v, \varphi_n) p(ds) - \int_B \max_{v \in \mathrm{Pr}(A(s))} P(s, v, \varphi_0) p(ds) \right|$$

$$\le \int_B \left| \max_{v \in \mathrm{Pr}(A(s))} P(s, v, \varphi_n) p(ds) - \max_{v \in \mathrm{Pr}(A(s))} P(s, v, \varphi_0) \right| p(ds)$$

$$\le \int_B \max_{a \in A(s)} \left| \int_S (J(\varphi_n)(s') - J(\varphi_0)(s')) \rho(s, a, s') p(ds') \right| p(ds)$$

for any Borel set $B \in \mathcal{B}(S)$. This and (7.5) imply that

$$\lim_{n \to \infty} \int_B \max_{v \in \mathrm{Pr}(A(s))} P(s, v, \varphi_n) p(ds) = \int_B \max_{v \in \mathrm{Pr}(A(s))} P(s, v, \varphi_0) p(ds) \qquad (7.6)$$

for any $B \in \mathcal{B}(S)$. By Lemmas 7.1 and 7.2, we have

$$\lim_{n \to \infty} \int_B P\big(s, \phi_n(s), \varphi_n\big) p(ds) = \int_B P\big(s, \phi_0(s), \varphi_0\big) p(ds), \qquad B \in \mathcal{B}(S). \quad (7.7)$$

Since $\phi_n \in BR_p(\varphi_n)$ for all $n \in \mathbb{N}$, we conclude from (7.6) and (7.7) that

$$\int_B P\big(s, \phi_0(s), \varphi_0\big) p(ds) = \int_B \max_{v \in \mathrm{Pr}(A(s))} P(s, v, \varphi_0) p(ds)$$

for all $B \in \mathcal{B}(S)$. This implies that $\phi_0 \in BR_p(\varphi_0)$, i.e., BR_p has a closed graph. $\qquad \square$

Proof of Theorem 3.2 Clearly, each set $BR_p(\varphi)$ is convex. Moreover, since Φ_p is compact, $\varphi \mapsto BR_p(\varphi)$ is upper semicontinuous. By the Kakutani–Fan–Glicksberg fixed point theorem (see Aliprantis and Border [3, Corollary 17.55]), there exists some $\hat{\phi} \in \Phi_p$ such that $\hat{\phi} \in BR_p(\hat{\phi})$. Thus there exists a Borel set $B_1 \subseteq S$ such that $p(B_1) = 1$, the restriction of $\hat{\phi}$ to B_1 is Borel and

$$\hat{\phi}(s) \in \arg\max_{v \in \mathrm{Pr}(A(s))} P(s, v, \hat{\phi})$$

for all $s \in B_1$. Choose any Borel mapping $\hat{f} : S \to A$ such that

$$\hat{f}(s) \in \arg\max_{a \in A(s)} P(s, a, \hat{\varphi})$$

for all $s \in S \setminus B_1$. The existence of $\hat{f}$ is guaranteed by Brown and Purves [17, Corollary 1] or the Arsenin–Kunugui theorem (see Kechris [36, Theorem 18.18]). Define $\hat{\varphi}(s) = \hat{\phi}(s)$ for $s \in B_1$ and $\hat{\varphi}(s) = \hat{f}(s)$ for $s \in S \setminus B_1$. Since $q(\cdot|s, a) \ll p(\cdot)$ for all $(s, a) \in \mathbb{K}$, we have

$$\hat{\varphi}(s) \in \arg\max_{v \in \mathrm{Pr}(A(s))} P(s, v, \hat{\varphi})$$

for all $s \in S$, which completes the proof. $\qquad \square$

For any bounded Borel function $v : S \to \mathbb{R}$, we define

$$L(v)(s, a) := u(s, a) + \beta \int_S v(s') q(ds'|s, a), \qquad (s, a) \in \mathbb{K},$$

and if $\phi \in \Phi$, then

$$L_\phi(v)(s) := \int_{A(s)} L(v)(s, a) \phi(da|s) = u\big(s, \phi(s)\big) + \beta \int_S v(s') q\big(ds'|s, \phi(s)\big).$$

Let L_ϕ^n denote the composition of L_ϕ with itself n times.

The following fact is well known; see for instance Hernández-Lerma and Lasserre [31, Sect. 4.2].

Lemma 7.4 (a) *The equality* $v(s) = L_\phi(v)(s)$ *holds for all* $s \in S$ *if and only if* $v(s) = J^\beta(\phi)(s)$ *for all* $s \in S$.

(b) *For any bounded Borel function* $v : S \to \mathbb{R}$, *it holds that*

$$\lim_{n \to \infty} L_\phi^n(v)(s) = J^\beta(\phi)(s) \qquad \textit{for every } s \in S.$$

Proof of Theorem 3.4 By Theorem 3.2, there exists a stationary Markov perfect equilibrium $\hat{\varphi} \in \Phi$. For every $s \in S$, we define

$$\Gamma(s) := \mathrm{supp}(\hat{\varphi}(\cdot|s)) = \bigcap \{K : K \subseteq A(s), K \text{ is closed and } \hat{\varphi}(K|s) = 1\}.$$

The closed-valued correspondence Γ is weakly measurable because for any open set $U \in A$, the set of all $s \in S$ with $\Gamma(s) \cap U \neq \emptyset$ is precisely $\{s \in S : \hat{\varphi}(U|s) \neq 0\}$ and belongs to $\mathcal{B}(S)$. Since $A(s)$ is compact for each $s \in S$, also $\Gamma(s)$ is compact for each $s \in S$. Therefore Γ has a Borel graph (see Himmelberg [33]). Let Λ be the set of all $(s, a_1, a_2, \ell) \in S \times A \times A \times [0, 1]$ such that $a_1, a_2 \in \Gamma(s) = \mathrm{supp}(\hat{\varphi}(\cdot|s))$ and

$$\ell L\big(J^\beta(\hat{\varphi})\big)(s, a_1) + (1 - \ell)L\big(J^\beta(\hat{\varphi})\big)(s, a_2) = L_{\hat{\varphi}}\big(J^\beta(\hat{\varphi})\big)(s).$$

Clearly, Λ is a Borel set. Moreover, the set

$$\Lambda(s) = \{(a_1, a_2, \ell) : (s, a_1, a_2, \ell) \in \Lambda\}$$

is nonempty and compact. By the Arsenin–Kunugui theorem (see Kechris [36, Theorem 18.18]), there exist Borel mappings $f : S \to A$, $g : S \to A$ and $\lambda : S \to [0, 1]$ such that $(f(s), g(s), \lambda(s)) \in \Lambda(s)$ for all $s \in S$. Let

$$\phi_*(\cdot|s) := \lambda(s)\delta_{f(s)}(\cdot) + \big(1 - \lambda(s)\big)\delta_{g(s)}(\cdot). \tag{7.8}$$

Clearly, $\phi_* \in \Phi$ and $\phi_*(\{f(s), g(s)\}|s) = 1$ for each $s \in S$. We have

$$L_{\phi_*}\big(J^\beta(\hat{\varphi})\big)(s) = L_{\hat{\varphi}}\big(J^\beta(\hat{\varphi})\big)(s) = J^\beta(\hat{\varphi})(s)$$

for all $s \in S$. From Lemma 7.4 (a), it follows that $J^\beta(\phi_*) = J^\beta(\hat{\varphi})$. Let

$$A_0(s, \hat{\varphi}) = \arg\max_{a \in A(s)} P(s, a, \hat{\varphi}).$$

Since $f(s), g(s) \in \Gamma(s) \subseteq A_0(s, \hat{\varphi})$ for all $s \in S$, we conclude that

$$\max_{\nu \in \mathrm{Pr}(A(s))} \left(u(s, \nu) + \alpha\beta \int_S J^\beta(\phi_*)(s')q(ds'|s, \nu) \right)$$

$$= u\big(s, \phi_*(s)\big) + \alpha\beta \int_S J^\beta(\phi_*)(s')q\big(ds'|s, \phi_*(s)\big)$$

$$= P\big(s, \phi_*(s), \phi_*\big). \tag{7.9}$$

From (7.9), we conclude that ϕ_* is a stationary Markov perfect equilibrium with the required property that the support of every measure $\phi_*(\cdot|s)$ contains at most two points. This completes the proof. $\qquad\square$

Let $\mu \in \mathrm{Pr}(S)$ and $w : \mathbb{K} \to \mathbb{R}$ be a Borel function such that for all $\psi \in F$, the integral $\int_S |w(s, \psi(s))| \mu(ds)$ is finite. We define

$$I_w(f, g, \lambda)(s) := \lambda(s) w(s, f(s)) + (1 - \lambda(s)) w(s, g(s))$$

where $f, g \in F$ and $\lambda : S \to [0, 1]$ is a Borel function. If $\psi \in F$, then

$$I_w(\psi)(s) := w(s, \psi(s)).$$

Given a Borel function $Y : S \to \mathbb{R}$ such that $\int_S |Y(s)| \mu(ds) < \infty$, we denote by $E[Y|\mathcal{G}]$ a version of the conditional expectation of Y with respect to the σ-field $\mathcal{G}$.

The following result is a corollary to Dynkin and Evstigneev [23, Theorem 1.2], which is an extension of the classical Lyapunov theorem.

Lemma 7.5 *Assume that $\mathcal{G}$ is a σ-field contained in $\mathcal{B}(S)$ and $\mathcal{B}(S)$ has no $\mathcal{G}$-atoms under an atomless probability measure μ. Let $f, g \in F$. Then for any Borel function $\lambda : S \to [0, 1]$, there exists some $\psi \in F$ such that $\psi(s) \in \{f(s), g(s)\}$ for all $s \in S$ and*

$$E[I_w(f, g, \lambda)|\mathcal{G}] = E[I_w(\psi)|\mathcal{G}] \qquad \mu\text{-a.e.}$$

Proof of Theorem 3.5 Let ϕ_* be the stationary Markov perfect equilibrium given in (7.8). By Lemma 7.5, there exists some $\phi_0 \in F$ such that $\phi_0(s) \in \{f(s), g(s)\}$ for all $s \in S$ and

$$E\big[L_{\phi_*}\big(J^\beta(\phi_*)\big)\big|\mathcal{G}\big] = E\big[L_{\phi_0}\big(J^\beta(\phi_*)\big)\big|\mathcal{G}\big] \qquad p\text{-a.e.}$$

Since $\rho(s, a, \cdot)$ and $s \mapsto A(s)$ are $\mathcal{G}$-measurable, this implies that for all $(s, a) \in \mathbb{K}$, we have

$$\begin{aligned}
E\big[L_{\phi_*}\big(J^\beta(\phi_*)\big)\rho(s, a, \cdot)\big|\mathcal{G}\big] &= E\big[L_{\phi_*}\big(J^\beta(\phi_*)\big)\big|\mathcal{G}\big]\rho(s, a, \cdot) \\
&= E\big[L_{\phi_0}\big(J^\beta(\phi_*)\big)\big|\mathcal{G}\big]\rho(s, a, \cdot) \\
&= E\big[L_{\phi_0}\big(J^\beta(\phi_*)\big)\rho(s, a, \cdot)\big|\mathcal{G}\big] \qquad p\text{-a.e.} \quad (7.10)
\end{aligned}$$

By taking the expectation on both sides of (7.10) with respect to p, we obtain

$$\int_S L_{\phi_*}\big(J^\beta(\phi_*)\big)(s') q(ds'|s, a) = \int_S L_{\phi_0}\big(J^\beta(\phi_*)\big)(s') q(ds'|s, a) \qquad (7.11)$$

for all $(s, a) \in \mathbb{K}$. Multiplying both sides of (7.11) by β, putting $a = \phi_0(s)$ and adding to both sides $u(s, \phi_0(s))$, we obtain

$$L_{\phi_0}\Big(L_{\phi_*}\big(J^\beta(\phi_*)\big)\Big)(s) = L_{\phi_0}^2\big(J^\beta(\phi_*)\big)(s), \qquad s \in S.$$

Since $J^\beta(\phi_*) = L_{\phi_*}(J^\beta(\phi_*))$, it follows that

$$L_{\phi_0}\big(J^\beta(\phi_*)\big)(s) = L_{\phi_0}^2\big(J^\beta(\phi_*)\big)(s), \qquad s \in S.$$

By iterating this equality, we obtain

$$L_{\phi_0}\big(J^\beta(\phi_*)\big)(s) = L_{\phi_0}^n\big(J^\beta(\phi_*)\big)(s), \qquad s \in S.$$

This equality and Lemma 7.4 (b) imply that

$$L_{\phi_0}\big(J^\beta(\phi_*)\big)(s) = \lim_{n\to\infty} L_{\phi_0}^n\big(J^\beta(\phi_*)\big)(s) = J^\beta(\phi_0)(s), \qquad s \in S. \tag{7.12}$$

Since $J^\beta(\phi_0) = L_{\phi_0}(J^\beta(\phi_0))$, we conclude from (7.12) that

$$\beta \int_S J^\beta(\phi_*)(s')q\big(ds'|s, \phi_0(s)\big) = \beta \int_S J^\beta(\phi_0)(s')q\big(ds'|s, \phi_0(s)\big) \tag{7.13}$$

for all $s \in S$. Multiplying both sides of (7.13) by α and adding to both sides $u(s, \phi_0(s))$, we obtain

$$P\big(s, \phi_0(s), \phi_*\big) = P\big(s.\phi_0(s), \phi_0\big) \tag{7.14}$$

for all $s \in S$. We know from Lemma 7.5 that $\phi_0(s) \in \{f(s), g(s)\}$ for all $s \in S$. From (7.9) and (7.14), we deduce that

$$P\big(s, \phi_0(s), \phi_0\big) = P\big(s, \phi_0(s), \phi_*\big) = \max_{v \in \mathrm{Pr}(A(s))} P(s, v, \phi_*), \qquad s \in S. \tag{7.15}$$

Since $L_{\phi_*}(J^\beta(\phi_*)) = J^\beta(\phi_*)$, we obtain from (7.12) that $L_{\phi_0}(J^\beta(\phi_*)) = J^\beta(\phi_0)$. This fact and (7.11) imply that

$$\int_S J^\beta(\phi_*)(s')q(ds'|s, a) = \int_S J^\beta(\phi_0)(s')q(ds'|s, a) \tag{7.16}$$

for all $(s, a) \in \mathbb{K}$. From (7.16), we easily conclude that

$$P(s, a, \phi_*) = P(s, a, \phi_0) \qquad \text{for all } (s, a) \in \mathbb{K}.$$

This equality in turn implies that

$$\max_{v \in \mathrm{Pr}(A(s))} P(s, v, \phi_*) = \max_{v \in \mathrm{Pr}(A(s))} P(s, v, \phi_0). \tag{7.17}$$

Now, we easily deduce from (7.15) and (7.17) that ϕ_0 is a deterministic stationary Markov perfect equilibrium. □

Remark 7.6 In order to adapt the proofs in this section to the unbounded utility case discussed in Remark 3.12, one has to replace $L^\infty(S, p)$ by the space of classes of functions $v : S \to \mathbb{R}$ such that $s \mapsto \frac{v(s)}{\omega(s)}$ is p-essentially bounded.

8 Concluding remarks

In this paper, we have studied a fairly general class of time-inconsistent Markov decision processes with a Borel state space. Using quasi-hyperbolic discounting and the game-theoretic formulation as for instance in Balbus et al. [6], Harris and Laibson [27], Pelag and Yaari [54], Phelps and Pollak [55] or Pollak [56], we have established the existence of a stationary Markov perfect equilibrium in models with transitions

having a density function. In order to obtain a stationary equilibrium, we have used a fixed point argument. More importantly, we have shown that a stationary Markov equilibrium may be simplified in the sense that all selves can randomise their choices over at most two pure actions in each state. The existence of a deterministic stationary equilibrium requires some additional assumptions on an atomless transition probability. The dynamic-programming-like algorithm used in Sect. 5 for Markov decision processes with countably many states produces a sequence $(\bar{f}_n)$ of strategies having a subsequence converging to a deterministic Markov perfect equilibrium. The sequence $(\bar{f}_n)$ itself need not be convergent (see Example 5.6). This non-stationary equilibrium may have interesting properties. Namely, it can dominate (in the sense of expected utilities) the randomised stationary one. We have also shown by a suitable approximation that ϵ-equilibria in deterministic Markovian strategies exist in some models with a Borel state space.

We should like to emphasise that an analysis of optimality (equilibria) in dynamic decision models under quasi-hyperbolic discounting cannot be done using the Bellman optimality principle. An extensive discussion of this issue can be found in Björk and Murgoci [15], Krusell and Smith [38], Maliar and Maliar [43, 44]. The examples given in this paper also confirm this statement. Therefore, we apply game-theoretic tools and a fixed point theorem. However, as noted by Maliar and Maliar [44], numerical calculations of stationary Markov perfect equilibria are complicated even in simple cases where closed-form (analytical) solutions are already known. The question of existence of deterministic equilibria in different types of models with a general state space remains open. Here, we have solved this problem for some important subclasses of decision processes.

As indicated earlier, studying Markov perfect equilibria in Markov decision processes with quasi-hyperbolic discounting has some relevance to macroeconomics, portfolio management or finance. We wish to point out in conclusion that Theorems 3.4 and 3.5 for Markov decision processes with a continuum of states extend and complete the results obtained by Balbus et al. [6] and Harris and Laibson [27] for consumption/investment models with atomless transitions. In Theorem 3.4, the transitions may have some atoms. Our results can be applied to various Markov decision processes with a multidimensional state space.

References

1. Acemoglu, D.: Introduction to Modern Economic Growth. Princeton University Press, Princeton (2008)
2. Ainslie, G.W.: Picoeconomics. Cambridge University Press, Cambridge (1992)
3. Aliprantis, C.D., Border, K.C.: Infinite Dimensional Analysis, A Hitchhiker's Guide, 3rd edn. Springer, Heidelberg (2006)
4. Alj, A., Haurie, A.: Dynamic equilibria in multigeneration stochastic games. IEEE Trans. Autom. Control **28**, 193–203 (1983)
5. Balbus, Ł., Jaśkiewicz, A., Nowak, A.S.: Robust Markov perfect equilibria in a dynamic choice model with quasi-hyperbolic discounting. In: Haunschmied, J., et al. (eds.) Dynamic Games in Economics. Dynamic Modeling and Econometrics in Economics and Finance, vol. 16, pp. 1–22. Springer, Berlin (2014)
6. Balbus, Ł., Jaśkiewicz, A., Nowak, A.S.: Existence of stationary Markov perfect equilibria in stochastic altruistic growth economies. J. Optim. Theory Appl. **165**, 295–315 (2015)
7. Balbus, Ł., Jaśkiewicz, A., Nowak, A.S.: Markov perfect equilibria in a dynamic decision model with quasi-hyperbolic discounting. Ann. Oper. Res. **287**, 573–591 (2020)
8. Balbus, Ł., Reffett, K., Woźny, Ł.: On uniqueness of time-consistent Markov policies for quasi-hyperbolic consumers under uncertainty. J. Econ. Theory **176**, 293–310 (2018)
9. Balder, E.J.: Lectures on Young measure theory and its applications in economics. Rend. Ist. Mat. Univ. Trieste **31**, 1–69 (2000)
10. Barro, R.: Ramsey meets Laibson in the neoclassical growth model. Q. J. Econ. **114**, 1125–1152 (1999)
11. Bäuerle, N., Rieder, U.: Markov Decision Processes with Applications to Finance. Springer, Heidelberg (2011)
12. Bernheim, B.D., Ray, D.: Markov perfect equilibria in altruistic growth economies with production uncertainty. J. Econ. Theory **47**, 195–202 (1989)
13. Bertsekas, D.P., Shreve, S.E.: Stochastic Optimal Control: The Discrete-Time Case. Academic Press, New York (1978)
14. Björk, T., Khapko, A., Murgoci, A.: On time-inconsistent stochastic control in continuous time. Finance Stoch. **21**, 331–360 (2017)
15. Björk, T., Murgoci, A.: A theory of Markovian time-inconsistent stochastic control in discrete time. Finance Stoch. **18**, 545–592 (2014)
16. Bobryk, R.V., Stettner, Ł.: Discrete-time portfolio selection with proportional transaction costs. Probab. Math. Stat. **19**, 235–248 (1999)
17. Brown, L.D., Purves, R.: Measurable selections of extrema. Ann. Stat. **1**, 902–912 (1973)
18. Caballero, R.: Consumption puzzles and precautionary savings. J. Monet. Econ. **25**, 113–136 (1990)
19. Chatterjee, S., Eyigungor, B.: Continuous Markov equilibria with quasi-geometric discounting. J. Econ. Theory **163**, 467–494 (2016)
20. Christensen, S., Lindensjö, K.: On finding equilibrium stopping times for time-inconsistent Markovian problems. SIAM J. Control Optim. **56**, 4228–4255 (2018)
21. Duggan, J.: Noisy stochastic games. Econometrica **80**, 2017–2045 (2012)
22. Dvoretzky, A., Wald, A., Wolfowitz, J.: Elimination of randomization in certain statistical decision procedures and zero-sum two-person games. Ann. Math. Stat. **22**, 1–21 (1951)
23. Dynkin, E.B., Evstigneev, I.V.: Regular conditional expectations of correspondences. Theory Probab. Appl. **21**, 325–338 (1976)
24. Ekeland, I., Pirvu, T.: Investment and consumption without commitment. Math. Financ. Econ. **2**, 57–86 (2008)
25. Elliott, R.J., Kalton, N.J., Markus, L.: Saddle-points for linear differential games. SIAM J. Control Optim. **11**, 100–112 (1973)
26. Goldman, S.M.: Consistent plans. Rev. Econ. Stud. **47**, 533–537 (1980)
27. Harris, C., Laibson, D.: Dynamic choices of hyperbolic consumers. Econometrica **69**, 935–957 (2001)
28. Haurie, A.: A multigenerational game model to analyze sustainable development. Ann. Oper. Res. **137**, 369–386 (2005)
29. He, W., Sun, Y.: Stationary Markov perfect equilibria in discounted stochastic games. J. Econ. Theory **169**, 35–61 (2017)
30. He, W., Sun, Y.: Conditional expectation of correspondences and economic applications. Econ. Theory **66**, 265–299 (2018)

31. Hernández-Lerma, O., Lasserre, J.B.: Discrete-Time Markov Control Processes: Basic Optimality Criteria. Springer, New York (1996)
32. Hernández-Lerma, O., Lasserre, J.B.: Further Topics on Discrete-Time Markov Control Processes. Springer, New York (1999)
33. Himmelberg, C.J.: Measurable relations. Fundam. Math. **87**, 53–72 (1975)
34. Jacobs, K.: Measure and Integral. Academic Press, New York (1978)
35. Jaśkiewicz, A., Nowak, A.S.: Stationary Markov perfect equilibria in risk sensitive stochastic overlapping generations models. J. Econ. Theory **151**, 411–447 (2014)
36. Kechris, A.S.: Classical Descriptive Set Theory. Springer, New York (1995)
37. Krusell, P., Kuruşçu, B., Smith, A.: Equilibrium welfare and government policy with quasi-geometric discounting. J. Econ. Theory **105**, 42–72 (2002)
38. Krusell, P., Smith, A.: Consumption and savings decisions with quasi-geometric discounting. Econometrica **71**, 366–375 (2003)
39. Laibson, D.: Golden eggs and hyperbolic discounting. Q. J. Econ. **112**, 443–477 (1997)
40. Luttmer, E. Mariotti, T.: Subjective discounting in an exchange economy. J. Polit. Econ. **111**, 959–989 (2003)
41. Lyapunov, A.A.: Sur les fonctions-vecteurs complètement additives. Izv. Akad. Nauk SSSR, Sér. Mat. **4**, 465–478 (1940)
42. Lyapunov, A.A.: Sur les fonctions-vecteurs complètement additives. Izv. Akad. Nauk SSSR, Sér. Mat. **10**, 277–279 (1946)
43. Maliar, L., Maliar, S.: Quasi-geometric discounting: a closed-form solution under the exponential utility function. Bull. Econ. Res. **56**, 201–206 (2004)
44. Maliar, L., Maliar, S.: Ruling out multiplicity of smooth equilibria in dynamic games: a hyperbolic discounting example. Dyn. Games Appl. **6**, 243–261 (2016)
45. Maskin, E., Tirole, J.: A theory of dynamic oligopoly: I & II. Econometrica **56**, 549–600 (1988)
46. Merton, R.C.: Optimum consumption and portfolio rules in a continuous-time model. J. Econ. Theory **3**, 373–413 (1971)
47. Montiel Olea, J.L., Strzalecki, T.: Axiomatization and measurement of quasi-hyperbolic discounting. Q. J. Econ. **129**, 1449–1499 (2014)
48. Neveu, J.: Mathematical Foundations of the Calculus of Probability. Holden-Day, San Francisco (1965)
49. Neyman, A.: From Markov chains to stochastic games. In: Neyman, A., Sorin, S. (eds.) Stochastic Games and Applications. NATO AIS Series, pp. 9–25. Kluwer Academic, Dordrecht (2003)
50. Nowak, A.S.: Existence of equilibrium stationary strategies in discounted noncooperative stochastic games with uncountable state space. J. Optim. Theory Appl. **45**, 591–602 (1985)
51. Nowak, A.S.: On a noncooperative stochastic game played by internally cooperating generations. J. Optim. Theory Appl. **144**, 88–106 (2010)
52. Nowak, A.S., Raghavan, T.E.S.: Existence of stationary correlated equilibria with symmetric information for discounted stochastic games. Math. Oper. Res. **17**, 519–526 (1992)
53. Osborne, M.J.: An Introduction to Game Theory. Oxford University Press, Oxford (2004)
54. Peleg, B., Yaari, M.E.: On the existence of a consistent course of action when tastes are changing. Rev. Econ. Stud. **40**, 391–401 (1973)
55. Phelps, E., Pollak, R.A.: On second best national savings, and game equilibrium growth. Rev. Econ. Stud. **35**, 195–199 (1968)
56. Pollak, R.A.: Consistent planning. Rev. Econ. Stud. **35**, 185–199 (1968)
57. Puterman, M.L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley, New York (2005)
58. Royden, H.: Real Analysis. Macmillan, New York (1968)
59. Samuelson, P.: A note on measurement of utility. Rev. Econ. Stud. **4**, 155–161 (1937)
60. Samuelson, P.: Lifetime portfolio selection by the dynamic stochastic programming. Rev. Econ. Stat. **51**, 239–246 (1969)
61. Schäl, M.: On dynamic programming and statistical decision theory. Ann. Stat. **7**, 432–445 (1979)
62. Scheffé, H.: A useful convergence theorem for probability distributions. Ann. Math. Stat. **18**, 434–438 (1947)
63. Shapley, L.S.: Stochastic games. Proc. Natl. Acad. Sci. USA **39**, 1095–1100 (1953)
64. Shreve, S.E., Soner, H.M.: Optimal investment and consumption with transaction costs. Ann. Appl. Probab. **4**, 609–692 (1994)
65. Stachurski, J.: Economic Dynamics: Theory and Computation. MIT Press, Cambridge (2009)

66. Stokey, N.L., Lucas, R.E. Jr, Prescott, E.: Recursive Methods in Economic Dynamics. Harvard University Press, Cambridge (1989)
67. Strotz, R.H.: Myopia and inconsistency in dynamic utility maximization. Rev. Econ. Stud. **23**, 165–180 (1956)
68. Topkis, M.D.: Submodularity and Complementarity. Princeton University Press, Princeton (1998)
69. Vieille, V., Weibull, J.W.: Multiple solutions under quasi-exponential discounting. Econ. Theory **39**, 513–526 (2009)
70. Warga, J.: Optimal Control of Differential and Functional Equations. Academic Press, New York (1972)
71. Wessels, J.: Markov programming by successive approximations with respect to weighted supremum norms. J. Math. Anal. Appl. **58**, 326–335 (1977)
72. Whitt, W.: Representation and approximation of noncooperative sequential games. SIAM J. Control Optim. **18**, 33–48 (1980)