Dynamic Programming with State-Action-Dependent Discounting

Chien-Hsiang Yeh

Australian National University

January 21, 2024

Abstract. In this paper, we extend the discrete-time dynamic programming to the case of state-action-dependent discounting. We establish a sufficient condition known as "eventual discounting" to guarantee the standard optimality results. The condition becomes necessary for the existence of policy value given a compact state space and concave functions. Our research encompasses dynamic programming with both bounded and unbounded rewards. Furthermore, we extend the scope of eventual discounting to applications involving risk-sensitive preferences.

Keywords: Dynamic programming, Markov decision process, State-Action-Dependent Discounting, Optimality

1. Introduction

Dynamic programming or Markov decision processes (MDPs) is a robust framework for modeling and solving sequential decision-making problems. The conventional approach in dynamic programming assumes constant discount factors to capture the trade-off between present and future rewards. However, in economics and finance, discount rates vary over time (Cochrane, 2011, Hills and Nakata, 2018). The constant discounting falls short in explaining the intricacies of real-world decision-making where agents have subjective time preferences or experience exogenous uncertainty in discount rates. For example, regarding shock-dependent discounts, Justiniano and Primiceri (2008) illustrate that the variance in discount factors accounts for a significant portion of consumption volatility. Albuquerque et al. (2016) shows that risk

in time preference accounts for key asset pricing moments, such as equity premium, bond term premium, and weak correlation between stock returns and fundamentals.

Besides the uncertainty in exogenous state-dependent discounts, endogenous time preferences cannot be overlooked in economics. For instance, Choi et al. (2008) demonstrate that endogenously generated short-run international differences in subjective discounting, indicating increasing relative U.S. impatience, result in saving and current account imbalances that align with observed data patterns. Hashimzade et al. (2023) shows that a self-reinforcing redistribution mechanism, through which the endogenous discounting can lead to a higher equilibrium interest rate and a more unequal wealth distribution, in comparison to the benchmark model with a constant discount rate. Maeda and Nagaya (2023) show that the accelerated short-sighted consumption habit leads to earlier depletion of exhaustible resources.

This paper extends the dynamic programming theory to account for state-action-dependent discounting. We establish a comprehensive dynamic programming theory by introducing the concept of eventual discounting, where the expected multiplicative discount factors are eventually less than one for any policies. We demonstrate that eventual discounting implies the eventually contraction of Bellman operator, which is then contracting in some weighted supremum norm. Under these conditions of eventual discounting, we demonstrate the existence of an optimal policy, validate Bellman's principle of optimality, and confirm that value function iteration, Howard policy iteration, and optimistic iteration converge to the desired value function or optimal policy.

In the existing literature, Stachurski and Zhang (2021) provides a complete theory of dynamic programming with state-dependent discounting, while Sargent and Stachurski (2023) presents a theory of dynamic programming of state-action-dependent discounting in finite state and action spaces. Moreover, Toda (2021) and Toda (2023) use the Perov Contraction Theorem to prove that the Bellman operator is contracting in some weighted supremum norm and has a unique fixed point when the exogenous state space is finite and discount factors are state-dependent.[1]

Addressing the gap in the literature concerning action-dependent time preferences, our exploration considers general state and action spaces with state-action-dependent

[1]The weighted function is an upper bound for the reward.

discounting. In particular, we provides the standard optimality results for the dynamic programming with action-dependent discounting as outlined in Uzawa (1968) and Becker and Mulligan (1997), combining it with state-dependent stochastic discounting.

The study begins by examining the dynamic programming framework concerning bounded rewards. We show that the eventual discounting is sufficient for the existence and uniqueness of fixed points for both the Bellman operator and the policy value operators. We demonstrate that the eventually discounting condition renders these operators both eventually contracting in the supremum norm and contracting in some weighted supremum norm. This finding provides a robust foundation for the analysis of the convergence rates of these operators over time.

Furthermore, we explore the necessity of eventual discounting for the Markov decision process when state space is compact and functions are concave. We show that the condition of eventual discounting is necessary for ensuring the existence and uniqueness of policy values or the fixed points to the policy operators. In this context, we observe that the spectral radii of operators for discounted conditional expectation primarily influence the convergence rate of the system. Notably, the eigenfunctions corresponding to these spectral radii are employed as weighting vectors in weighted supremum norms.

In the extension, we generalize the result to the dynamic programming with unbounded rewards via the Q-transform, following Ma et al. (2022). We show that the standard optimality of dynamic programming holds when the discount factors are state-action-dependent and eventually discounting. The computation method is provided. We prove that value function iteration and action-value function iteration of the expected action-value operator converge to the optimal value or action value, which extends Ma et al. (2022). We also study risk-sensitive preference with eventual discounting, which ensures the uniqueness of optimal value function and the existence of optimal policy.

Related literature. The related literature in the Markov decision process with state- or action-dependent discount factors includes Wei and Guo (2011), Minjárez-Sosa (2015), Wu et al. (2015), Wu and Zhang (2016) and Jasso-Fuentes et al. (2022).

About endogenous time preferences, Uzawa (1968) and Epstein and Hynes (1983) propose a theory in which impatience rises with consumption, suggesting that the

rich are more likely to heavily discount future consumption. Although Uzawa time preference may be counterintuitive and not be supported by empirical evidence, it succeeds in theories of small open economics such as solving non-stationary issues (Schmitt-Grohé and Uribe, 2003, Guest and McDonald, 2001, Dutta and Yang, 2013). Helpman and Razin (1982) point out that when the rate of time preference remains lower than the interest rate, individuals rationally choose to accumulate foreign assets to finance their increasing consumption. The constant-discount representation of preferences cannot effectively capture this dynamic behavior. Consequently, studies on small open economies commonly employ Uzawa-type endogenous discount factors (Obstfeld, 1990, Mendoza, 1991, Schmitt-Grohé and Uribe, 2003, Durdu et al., 2009, Bodenstein, 2011, 2013, Vasilev, 2022b). For instance, Vasilev (2022a) considers Uzawa time preference, where the higher level of real income today leads to a lower discount rate, to explain the propagation of cyclical fluctuations in Bulgaria. Durdu et al. (2009) uses Uzawa endogenous time preference to model the financial globalization and the risk of Sudden Stop problem.

In contrast to Uzawa time preferences, Becker and Mulligan (1997) propose a framework that time patience is marginally increasing in future-oriented capital. Dutta and Yang (2013) establish endogenous discount factors such that marginal impatience is increasing in consumption (Uzawa type) and decreasing in future-oriented capital (Becker-Mulligan type). Their model is consistent with the empirical evidence from Australia that current consumption and turnover in future-oriented capital are positively correlated.

In empirical studies on time preferences and wealth, Lawrance (1991) reveals that the affluent exhibit greater patience than the less affluent from estimating the consumption Euler equations and the data in Panel Study of Income Dynamics. Huffman et al. (2019) examine heterogeneity in time preferences among elderly Americans and suggest that impatience correlates with lower wealth. Samwick (1998) indicates that time patience tends to increase with both income and age, by estimating the distribution of discount rates from the wealth data in Survey of Consumer Finances 1992. Cohen et al. (2020) provides a survey of the related literature.

A survey of dynamic programming with state-dependent discount rates can be found in Stachurski and Zhang (2021). In asset pricing and state-dependent discounts, since variation in asset returns is significantly due to variation in discount factors, asset pricing models consider stochastic discount factors depending on the state of

consumption growth to include adjustments for risk (Lucas Jr, 1978, Rosenberg and Engle, 2002, Cochrane, 2009, Hansen and Renault, 2010). (Campbell and Ammer, 1993, Cochrane, 2011) point out that variation in asset returns is predominantly due to variation in discount factors.

In literature of risk-sensitive preference, Hansen and Sargent (1995) develop a recursive dynamics of discounted costs for a linear quadratic exponential Gaussian linear control model, introducing risk adjustment into the framework. Moreover, Bäuerle and Jaśkiewicz (2018) studies a one-sector optimal growth model with unbounded shocks and rewards, in the framework of Hansen and Sargent (1995). They demonstrates the optimality equation for the non-expected utility and establish the Euler equation. Their analysis is based on an inequality involving associated random variables, a concept that is also utilized in the present paper. Weil (1993) develop a stochastic optimal consumption model with constant absolute risk aversion to study precautionary savings and the permanent income hypothesis. Backus et al. (2015) investigate a business cycle model incorporating aggregate risk and ambiguity. They observe that heightened uncertainty typically leads to a reduction in consumption.

If state and action spaces are finite, optimality and comparative statics can be established using fixed-point theory in complete lattices. Relevant techniques and insights can be found in works by Zhou (1994), Olszewski (2021), Balbus et al. (2022), Stachurski et al. (2022b), and Sargent and Stachurski (2023). Additionally, literature on the uniqueness of fixed-point problems for recursive preferences and dynamics includes Marinacci and Montrucchio (2010), Borovička and Stachurski (2020), Bloise et al. (2021), Ren and Stachurski (2021), and Christensen (2022).

The paper is structured as follows. Section 2 sets up the recursive decision process and presents its optimality results under eventual discounting. Section 3 presents the optimality of an eventually discounting Markov decision process. Section 4 studies the necessity of eventual discounting. Section 5 gives applications. Section 6 extends to unbounded rewards. Section 7 treats extensions in risk-sensitive preference.

2. Recursive Decision Process with Eventual Discounting

In this section, we introduce the framework of the recursive decision process and dynamic programming with state-action-dependent discounting. We also show the optimality results and the convergences of the conventional computation methods

under eventual discounting. The application of recursive decision processes in Markov decision processes is presented in the next section.

2.1. **Preliminary.** Let X be a metric space. Denote the family of Borel measurable (and bounded) real-valued functions on X as $m\mathsf{X}$ ($mb\mathsf{X}$). Denote the family of real-valued continuous (resp. upper semicontinuous) and bounded functions on X as $cb\mathsf{X}$ (resp. $ub\mathsf{X}$). Define $\mathscr{B}(\mathsf{X})$ as the Borel σ-algebra on X. Given an everywhere positive function w on X, let $\|\cdot\|_w$ be the weighted supremum norm $\|v\|_w \coloneqq \sup_{x\in\mathsf{X}} |v(x)|/w(x)$ for all $v \in m\mathsf{X}$ and let $\|\cdot\|$ be the supremum norm. Let U be a metric space. A self-map F on U is called *globally stable* if T has a unique fixed point $u^* \in U$ and $F^k u \to u^*$ for all $u \in U$. Also, F is called *eventually contracting* if there is $k \in \mathbb{N}$ such that F^k is contracting on U. Denote $\mathbb{1} \in m\mathsf{X}$ as $\mathbb{1} \equiv 1$. Throughout, $\leqslant$ is the pointwise order on $\mathbb{R}^{\mathsf{X}}$.

2.2. **Recursive Decision Process.** The fundamental idea of dynamic programming is to solve sequential decision problems recursively through a *generic Bellman equation*

$$v(x) = \sup_{a\in\Gamma(x)} B(x,a,v) \qquad (x \in \mathsf{X} \text{ and } v \in mb\mathsf{X}) \tag{1}$$

where

(i) X is a Borel space, referred to as the *state space*,

(ii) A is a Borel space, referred to as the *action space*,

(iii) $\Gamma\colon \mathsf{X} \to \mathsf{A}$ is a nonempty correspondence, referred to as the *feasible correspondence*, such that $\Gamma(x)$ is a measurable subset of A for all $x \in \mathsf{X}$, which defines

- the set of the *feasible state-action pairs*

$$\mathsf{G} \coloneqq \{(x,a) \in \mathsf{X}\times\mathsf{A}\colon a \in \Gamma(x)\}, \text{ and}$$

- the set of *feasible stationary policies*

$$\Sigma \coloneqq \{\sigma \in \mathsf{A}^{\mathsf{X}}\colon \sigma \text{ is Borel-measurable and } \sigma(x) \in \Gamma(x) \text{ for all } x \in \mathsf{X}\},$$

(iv) a *value aggregator* $B\colon \mathsf{G}\times mb\mathsf{X} \to \mathbb{R}$ satisfies the monotonicity condition

$$v, w \in mb\mathsf{X} \text{ and } v \leqslant w \Longrightarrow B(x,a,v) \leqslant B(x,a,w) \quad \text{for all } (x,a) \in \mathsf{G}. \tag{2}$$

A *recursive decision process* (RDP) is a tuple $\mathscr{R} = (\mathsf{X}, \mathsf{A}, \Gamma, B)$. We interpret $B(x, a, v)$ as the lifetime rewards, contingent on current state x and action a, using v to evaluate future states. We assume the following regular conditions for the primitives of an RDP throughout this section.

Condition 2.1.

(a) Γ is nonempty, continuous, and compact-valued, and

(b) $(x, a) \mapsto B(x, a, v)$ is bounded and measurable on G for all $v \in mb\mathsf{X}$, and also continuous whenever $v \in cb\mathsf{X}$.

An RDP is called *regular* if it satisfies Condition 2.1. The regular conditions make the dynamic programming well-defined, indicating the existence of an optimal policy or maximizer for Bellman equation (1) when the function v within the value aggregator is continuous.[2]

Given $v \in mb\mathsf{X}$, a policy $\sigma \in \Sigma$ is called *v-greedy* if

$$\sigma(x) \in \operatorname*{arg\,max}_{a \in \Gamma(x)} B(x, a, v) \qquad \text{for all } x \in \mathsf{X}.$$

The regular condition 2.1 guarantees the existence of a greedy policy for any $v \in cb\mathsf{X}$, which is a fundamental requirement in dynamic programming. The next lemma shows that whenever $\mathscr{R}$ is regular, T_σ and T are well-defined on $mb\mathsf{X}$ and $cb\mathsf{X}$, respectively, and v-greedy policies exist for all $v \in cb\mathsf{X}$.

Lemma 2.1. *If $\mathscr{R}$ is regular, then the following statements are true.*

(a) T_σ is a self-map on $mb\mathsf{X}$ for all $\sigma \in \Sigma$.

(b) T is a self-map on $cb\mathsf{X}$.

(c) For all $v \in cb\mathsf{X}$, there exists a v-greedy policy.

Lemma 2.1 implies that T_σ is a self-map for any $\sigma \in \Sigma$ whenever $\mathscr{R}$ is regular. It also implies that for any $v \in cb\mathsf{X}$ there exists a v-greedy policy. This ensures that the policy iteration is well-behaved. Then, Howard operator H and optimistic policy operator W are well-defined.

[2]In this paper, we call such assumptions for the existence of optimal solutions as regular conditions.

2.3. **Eventual Discounting.** Analogous to the standard contracting Markov decision process that a constant discount factor is strictly less than one, we adopt the assumptions for the stochastic discount factors outlined in Stachurski and Zhang (2021), Toda (2021) and Sargent and Stachurski (2023). Denote $\rho(A)$ as the spectral radius for a bounded linear operator $A\colon mb\mathsf{X} \to mb\mathsf{X}$:

$$\rho(A) \coloneqq \lim_{n\to\infty} \|A^n\|^{1/n}$$

where $\|A\|$ denotes the operator norm of A.[3] Assume that there is a Borel measurable $k\colon \mathsf{G} \times \mathsf{X} \to \mathbb{R}_+$ such that

$$|B(x,a,v) - B(x,a,w)| \leqslant \int_{\mathsf{X}} |v(x') - w(x')| k(x,a,x') \mathrm{d}x' \tag{3}$$

for all $(x,a) \in \mathsf{G}$ and $v, w \in mb\mathsf{X}$. Given $\sigma \in \Sigma$, let L_σ be a positive linear operator on $mb\mathsf{X}$ defined by[4]

$$L_\sigma h(x) \coloneqq \int_{\mathsf{X}} h(x') k(x, \sigma(x), x') \mathrm{d}x' \qquad (x \in \mathsf{X}, h \in mb\mathsf{X}). \tag{4}$$

We then have

$$|T_\sigma v(x) - T_\sigma w(x)| \leqslant L_\sigma |v - w|(x) \qquad (x \in \mathsf{X}, v, w \in mb\mathsf{X}).$$

We say that T_σ is *eventually discounting* if $\rho(L_\sigma) < 1$. Define the *eventual-discount factor* d_n^σ under policy $\sigma \in \Sigma$ by

$$d_n^\sigma \coloneqq \sup_{x\in\mathsf{X}} L_\sigma^n \mathbb{1}(x) \qquad (n \in \mathbb{N}),$$

and $\ell(x,x') \geqslant 0$ for all $(x,x') \in \mathsf{X}^2$. We can show that $\rho(L_\sigma) < 1$ if and only if $d_{n_\sigma}^\sigma < 1$ for some $n_\sigma \in \mathbb{N}$. It generalizes the constant discount factor $\beta < 1$, since it can be shown that $\rho(L_\sigma) < 1$ implies the global stability of T_σ. To ensure that T is globally stable, we introduce the following conditions.

Assumption 2.1.

(i) There is $k\colon \mathsf{G}\times\mathsf{X} \to \mathbb{R}_+$ such that $(x,a) \mapsto \int_{\mathsf{X}} k(x,a,x')h(x')\mathrm{d}x'$ is continuous and bounded on G for any $h \in cb\mathsf{X}$ and

$$|B(x,a,v) - B(x,a,w)| \leqslant \int_{\mathsf{X}} k(x,a,x')|v(x') - w(x')|\mathrm{d}x'$$

for all $(x,a) \in \mathsf{G}$ and $v, w \in mb\mathsf{X}$.

[3] $\|A\| \coloneqq \inf\{c \geqslant 0\colon \|Av\| \leqslant c\|v\| \text{ for all } v \in mb\mathsf{X}\}$.

[4] A linear operator operator L on $mb\mathsf{X}$ is positive if $v \in mb\mathsf{X}$ and $v \geqslant 0$ imply $Lv \geqslant 0$.

(ii) For any $\sigma \in \Sigma$, $\rho(L_\sigma) < 1$, where L_σ is defined by (4).

(iii) For any $\sigma \in \Sigma$, there exists a $w_\sigma \in mb\mathsf{X}$ and $\lambda_\sigma \geqslant 0$ such that $w_\sigma \geqslant \mathbb{1}$ and $L_\sigma w_\sigma \leqslant \lambda_\sigma w_\sigma$. Moreover, $\sup_{\sigma\in\Sigma} \lambda_\sigma < 1$.

Note that if X and A are finite, Assumption 2.1 (i) and (ii) are sufficient for the optimality of dynamic programming; that is, we do not require Assumption 2.1 (iii) when there are finite policies. Since it may be difficult to check the existence of n_σ for that $d^\sigma_{n_\sigma} < 1$ for all policies when there are infinitely many policies, we introduce the following stricter assumptions of eventual discounting.

Assumption 2.2. There is a Borel measurable function $\ell\colon \mathsf{X} \times \mathsf{X} \to \mathbb{R}_+$ such that

(i) $L\colon mb\mathsf{X} \to mb\mathsf{X}$ is a positive linear operator,

(ii) $|B(x,a,v) - B(x,a,w)| \leqslant L|v-w|(x)$ for all $x \in \mathsf{X}$, and

(iii) $\rho(L) < 1$, where

$$Lh(x) \coloneqq \int_{\mathsf{X}} h(x')\ell(x,x')\mathrm{d}x' \qquad (x \in \mathsf{X}, h \in mb\mathsf{X}). \tag{5}$$

We say that $\mathscr{R}$ is *eventually discounting* if either Assumption 2.1 or 2.2 holds. Assumption 2.1 implies that T_σ is contracting in the weighted supremum norm $\|\cdot\|_{w_\sigma}$, where w_σ is used as the weighting function for any $\sigma \in \Sigma_C$. It implies that T_σ eventually discounting that d^n_σ is eventually less than one for some large enough $n \geqslant 0$, whence v_σ exists and is well-defined. The assumption $\sup_{\sigma\in\Sigma} \lambda_\sigma < 1$ further ensures that the Bellman operator is eventually contracting so that we can compute v^* by VFI. On the other hand, Assumption 2.2 supposes that all operators L_σ are bounded above by some L such that $\rho(L) < 1$. Hence, we can show that Assumption 2.2 directly guarantees that the Bellman operator is eventually contracting.

The next example with state-dependent discounting, generalized from Chapter 10 of Stokey (1989), is a regular RDP.

Example 2.1 (One-Sector Optimal Growth)**.** An economy contains many identical and infinitely lived households. There is a single good $y_t = F(k_t, \ell_t)$ produced by capital k_t and labor ℓ_t inputs. Labor is supplied inelastically $\ell_t = 1$ for all t and capital depreciates at a rate $\delta \in (0,1)$. Denote c_t as consumption and i_t as investment. A

social planner solves the problem

$$\begin{aligned} & \sup \mathbb{E}_{k_0,z_0} \sum_{t=0}^{\infty} \left(\prod_{i=0}^{t-1} \beta(z_i, c_i) \right) U(c_t) \\ \text{s.t.} \quad & c_t + i_t \leqslant z_t y_t, \\ & 0 \leqslant \ell_t \leqslant 1, \\ & k_{t+1} = (1-\delta)k_t + i_t \qquad (t \in \mathbb{N}_0), \\ & k_0 \geqslant 0 \quad \text{and} \quad z_0 \geqslant 0 \text{ given}, \end{aligned}$$

where $\{z_t\}_{t\geqslant 0}$ is a sequence of exogenous shocks generated by a transition kernel Q on $(\mathsf{Z}, \mathscr{B}(\mathsf{Z}))$ with $\mathsf{Z} = [1, \bar{z}]$ for $1 < \bar{z} < \infty$. The state is $x = (k, z)$, and the action is $a = c$. Assume that F is continuously differentiable, strictly increasing, and strictly concave with

$$\begin{aligned} & F(0, \ell) = 0, \quad F_k(k, \ell) > 0, \quad F_\ell(k, \ell) > 0, \quad (k, \ell > 0) \\ & \lim_{k \to 0} F_k(k, 1) = \infty, \quad \lim_{k \to \infty} F_k(k, 1) = 0. \end{aligned}$$

Moreover, assume that U is continuous, β is continuous and strictly positive, and Q satisfies the Feller property such that $z \mapsto \int h(z')Q(z, \mathrm{d}z')$ is bounded and continuous for all $h \in cb\mathsf{Z}$. Let $\bar{k} > 0$ be such that $\bar{k} = \bar{z}F(\bar{k}, 1) + (1-\delta)\bar{k}$. Then, the set of maintainable capital stock is $[0, \bar{k}]$. Let the state space be $\mathsf{X} = [0, \bar{k}] \times [1, \bar{z}]$, the action space be $\mathsf{A} = [0, \bar{z}F(\bar{k}, 1)]$, the feasible correspondence be $\Gamma(k, z) = \{c \in \mathsf{A} \colon 0 \leqslant c \leqslant zF(k, 1)\}$. Hence, $(x, a) = ((k, z), c)$. Define the stochastic transition kernel by

$$P((k, z), c, (k', z')) = Q(z, z')\mathbb{1}\{k' = (1-\delta)k + zF(k, 1) - c\}.$$

for all $((k, z), c, (k', z')) \in \mathsf{G} \times \mathsf{X}$. Let the value aggregator B be

$$B((k, z), c, v) = U(c) + \mathbb{E}_{(k,z)}[\beta(z, c)v(k', z')] \qquad (((k, z), c, v) \in \mathsf{G} \times mb\mathsf{X}),$$

where (k', z') is generated by $P((k, z), c, \cdot)$. We can check that $\mathscr{R} = (\mathsf{X}, \mathsf{A}, \Gamma, B)$ is a regular RDP.[5] Moreover, we have

$$\begin{aligned} |B((k, z), c, v) - B((k, z), c, w)| &= |\mathbb{E}_{(k,z)}[\beta(z, c)v(k', z')] - \mathbb{E}_{(k,z)}[\beta(z, c)w(k', z')]| \\ &\leqslant \mathbb{E}_{(k,z)}\beta(z, c)|v(k', z') - w(k', z')|. \end{aligned}$$

for all $((k, z), c) \in \mathsf{G}$ and $v, w \in mb\mathsf{X}$. Let $L_\sigma \colon mb\mathsf{X} \to mb\mathsf{X}$ be

$$L_\sigma h(k, z) = \mathbb{E}^{\sigma}_{(k,z)}\beta(z, \sigma(k, z))h(k', z') \qquad ((k, z) \in \mathsf{X}, h \in mb\mathsf{X}),$$

[5] It is also a regular Markov decision process defined in Section 3.

where $\mathbb{E}^{\sigma}_{k,z}$ is the expectation conditioning on $(k_0, z_0) = (k, z)$ under transition kernel P defined above. Then, we have

$$d^{\sigma}_{n} = \sup_{(k,z)\in \mathsf{X}} L^{n}_{\sigma}\mathbb{1}(k,z) = \sup_{(k,z)\in \mathsf{X}} \mathbb{E}^{\sigma}_{(k,z)} \prod_{t=0}^{n} \beta(k_t, z_t).$$

Assume $\beta(z, c) \leqslant \bar{\beta}(z)$ for all k and c such that there exists an $n \in \mathbb{N}$:

$$\sup_{(k,z)\in \mathsf{X}} \mathbb{E}^{\sigma}_{(k,z)} \prod_{t=0}^{n} \beta(k_t, z_t) \leqslant \sup_{(k,z)\in \mathsf{X}} \mathbb{E}_{(k,z)} \prod_{t=0}^{n} \bar{\beta}(z_t) < 1. \tag{6}$$

Define $L\colon mb\mathsf{X} \to mb\mathsf{X}$ by

$$Lh(k,z) = \mathbb{E}_{(k,z)}\bar{\beta}(z)h(k', z') \qquad ((k,z) \in \mathsf{X}, h \in mb\mathsf{X}).$$

Hence, (6) implies $\sup_x L^n\mathbb{1}(x) < 1$ and then $\rho(L) < 1$. To this end, $\mathscr{R}$ is eventually discounting since Assumption 2.2 holds. Moreover, we have $\sup_\sigma d^{\sigma}_{n_\sigma} < 1$ by (6), which further implies Assumption 2.1.[6] □

2.4. **Dynamic Programming.** Let $\mathscr{R} = (\mathsf{X}, \mathsf{A}, \Gamma, B)$ be an RDP. Given $\sigma \in \Sigma$, a *policy operator* $T_\sigma\colon mb\mathsf{X} \to mb\mathsf{X}$ is defined by

$$T_\sigma v(x) \coloneqq B(x, \sigma(x), v) \qquad (x \in \mathsf{X} \text{ and } v \in mb\mathsf{X}).$$

If T_σ has a unique fixed point v_σ, then we call v_σ as *σ-value* function. We show in Section 2.5 that all policy operators have unique fixed points under appropriate assumptions on discount factors, which implies that σ-value exists and is unique for any $\sigma \in \Sigma$. The *(optimal) value function* of $\mathscr{R}$ is

$$v(x) \coloneqq \sup_{\sigma\in\Sigma} v_\sigma(x) \qquad (x \in \mathsf{X}).$$

A policy $\sigma^* \in \Sigma$ is called *optimal* if $v_\sigma = v^*$; that is, if

$$v_\sigma(x) \geqslant v_s \qquad \text{for all } s \in \Sigma \text{ and } x \in \mathsf{X}.$$

The *Bellman operator* $T\colon mb\mathsf{X} \to mb\mathsf{X}$ is defined by

$$Tv(x) \coloneqq \sup_{a\in\Gamma(x)} B(x, a, v) = \sup_{\sigma\in\Sigma} T_\sigma(x) \qquad (x \in \mathsf{X} \text{ and } v \in mb\mathsf{X}).$$

[6]The complete argument is shown in Section 3.

By the definition of greedy policies, σ is v-greedy if and only if $Tv = T_\sigma v$. We say v satisfies the *Bellman equations* if $Tv = v$; that is v satisfies (1). We say that *Bellman's principle of optimality* holds if

$$\sigma \in \Sigma \text{ is optimal for } \mathscr{R} \iff \sigma \text{ is } v^*\text{-greedy.}$$

That is, Bellman's principle of optimality holds whenever $v_\sigma = v^* \iff Tv^* = T_\sigma v^*$. Bellman's principle of optimality ensures that we can find the optimal policy if we can compute v^*, and v^*-greedy policy exists.

The conventional dynamic programming algorithms are defined as follows (see, e.g., Bertsekas (2022)). Let $\mathscr{R}$ be regular. A sequence $\{v_k\}_{k\geqslant 0} \subset cb\mathsf{X}$ is called a *value function iteration* (VFI) if $v_{k+1} = Tv_k$ for $k \geqslant 0$ with any $v_0 \in cb\mathsf{X}$. For policy iteration algorithms, we introduce the following conditions to ensure the existence of greedy policies and continuity of policy values during iterations. Note that if we know the greedy policies exist and are continuous for any value function, we do not require the following conditions.[7]

Condition 2.2.

(a) Γ is nonempty, continuous, compact-valued, and convex-valued,
(b) $(x, a) \mapsto B(x, a, v)$ is bounded and measurable on G for all $v \in mb\mathsf{X}$, and
(c) $a \mapsto B(x, a, v)$ is strictly quasi-concave for $x \in \mathsf{X}$ all $v \in cb\mathsf{X}$.

Define $\Sigma_C \subset \Sigma$ as the subset of all continuous policies. Let Condition 2.2 hold. Then, the optimal policy σ^* is continuous. It can be shown that $v^* = \sup_{\sigma\in\Sigma_C} v_\sigma$ and $Tv = \sup_{\sigma\in\sigma_C} T_\sigma v$ for all $v \in cb\mathsf{X}$. It then suffices to restrict the policy iterations to continuous policies. A sequence $\{\sigma_\kappa\}_{k\geqslant 0} \subset \Sigma_C$ is called a *Howard policy iteration* (HPI) if σ_{k+1} is v_{σ_k}-greedy for $k \geqslant 0$ and any $\sigma_0 \in \Sigma_C$. Define the *Howard operator* $H\colon cb\mathsf{X} \to cb\mathsf{X}$ by

$$Hv = v_\sigma \text{ where } \sigma \text{ is } v\text{-greedy.}$$

The iteration, $\{H^k v_0\}_{k\geqslant 0}$ with $v_0 = v_\sigma$ for $\sigma \in \Sigma_C$, of the Howard operator H is an abstract version of HPI. A sequence $\{v_k\}_{k\geqslant 0} \subset mb\mathsf{X}$ is called an *optimistic policy iteration* (OPI) if, fixing $m \in \mathbb{N}$, $v_{k+1} = T^m_{\sigma_k} v_k$ where σ_k is v_k-greedy for $k \geqslant 0$ with

[7] For example, state and action spaces are finite.

$v_0 = v_\sigma$ for any $\sigma \in \Sigma_C$. Observe that if $m \to \infty$, then OPI is the same as HPI; if $m = 1$, OPI is the same as VFI. Define optimistic policy operator $W : cb\mathsf{X} \to cb\mathsf{X}$ by

$$Wv = T_\sigma^m v \quad \text{where } \sigma \text{ is the first } v\text{-greedy policy} \qquad (v \in cb\mathsf{X}). \tag{7}$$

2.5. **Optimality.** As discussed in the introduction and the following section, we extend the existing dynamic programming theory to incorporate the cases of action-dependent discounting, including action-dependent time preferences introduced in Uzawa (1968) and Becker and Mulligan (1997). In this section, we show the optimality results, including the existence of optimal policies and Bellman's Principle of Optimality, when an RDP is eventually discounting. Moreover, the computation algorithms, including VFI, HPI, and OPI, converge to the optimal value function or optimal policies. For the following theorem, if Assumption 3.1 holds, define $w(x) = \inf_\sigma w_\sigma(x) \geqslant 1$ for all $x \in \mathsf{X}$.

Theorem 2.1. *Suppose that $\mathscr{R}$ is regular and either Assumption 2.1 or 2.2 holds. Then, the following statements are true.*

(a) T_σ is eventually contracting on $mb\mathsf{X}$ for all $\sigma \in \Sigma$,
(b) v^* is the unique solution to the Bellman equation in $cb\mathsf{X}$,
(c) VFI converges to v^*,
(d) Bellman's principle of optimality holds, and
(e) at least one optimal (continuous) policy exists,

Moreover,

(i) if Assumption 2.1 holds, then T is contracting on $(mb\mathsf{X}, \|\cdot\|_w)$ with modulus $\sup_{\sigma\in\Sigma} \lambda_\sigma$, and T_σ is contracting on $(mb\mathsf{X}, \|\cdot\|_{w_\sigma})$ with modulus λ_σ for all $\sigma \in \Sigma$, and
(ii) if Assumption 2.2 holds, then T is eventually contracting on $cb\mathsf{X}$.

If, in addition, Condition 2.2 holds, then

(α) HPI converges to σ^*, and
(β) OPI converges to v^*,

where all the iterated greedy policies to HPI and OPI are continuous.

Theorem 2.1 generalizes the conventional dynamic programming theory to the case of state-action-dependent discounting. In detail, Theorem 2.1 shows that the optimal policy exists, holds, VFI and OPI converge to v^*, and HPI converges to σ^*. The Bellman's principle of optimality guarantees that we can first compute v^* by VFI and find σ^* by

$$\sigma^*(x) = \arg\max_a B(x, a, v^*) \qquad (x \in \mathsf{X}).$$

It is observed that the condition of eventual discounting is sufficient for eventual contracting or global stability of T_σ or T.

Moreover, the value function v^* is continuous, and it suffices to search continuous policies for optimal policies. Note that if there is a non-continuous function satisfying the Bellman equation, then it is equal to or greater than the optimal value v^*, and there is no policy σ that its σ-value attains that function.[8] Since we are interested in v^*, which is continuous when the RDP is regular, we can restrict the iteration of VFI on the continuous functions. In addition, if Condition 2.2 holds and $v \in cb\mathsf{X}$, then any policy $\sigma \in \Sigma$ is dominated by some continuous policy $\sigma_c \in \Sigma_C$ that $T_\sigma v \leqslant T_{\sigma_c} v$. To this end, it also suffices to focus on the continuous policies for HPI and OPI. Note that Condition 2.2 is not necessary if we can ensure that the continuous greedy policy exists for any $v \in cb\mathsf{X}$. In particular, if action and state spaces are finite, the greedy policy exists so that we do not need Condition 2.2.

2.6. **Blackwell's Condition.** The generalized Blackwell's condition for the global stability of the Bellman operator T or policy operator T_σ is provided as follows.

Proposition 2.1. *Let T be an order-preserving self-map on $U \subset mb\mathsf{X}$. If there exists a positive linear operator G on $mb\mathsf{X}$ such that $\rho(G) < 1$ and*

$$T(v + c) \leqslant Tv + Gc \quad \text{for all } c, v \in U \text{ with } c \geqslant 0,$$

then T is eventually contracting on U.

Therefore, if we can check that, for any $\sigma \in \Sigma$, there exists a positive linear operator L_σ on $mb\mathsf{X}$ such that

$$T_\sigma(v + c)(x) = B(x, \sigma(x), v + c) \leqslant B(x, \sigma(x), v) + L_\sigma c(x) = T_\sigma v(x) + L_\sigma c(x)$$

for all $x \in \mathsf{X}$ and $c, v \in mb\mathsf{X}$ with $c \geqslant 0$, then T_σ is eventually contracting and has a unique fixed point v_σ.

[8] That is, if $Tw = w$ for some non-continuous $w \in mb\mathsf{X}$, then $w \geqslant v^*$ and $w > v_\sigma$ for any $\sigma \in \Sigma$.

3. Markov Decision Process with Eventual Discounting

In this section, we focus on a Markov decision process with state-action-dependent discounting and bounded rewards.

3.1. **Markov Decision Process.** A Markov decision process is an RDP such that the value aggregator is separated into reward and expected future continuing value. In detail, a *Markov decision process* $\mathscr{M}$ is an RDP $(\mathsf{X}, \mathsf{A}, \Gamma, B)$ such that

$$\begin{aligned} B(x,a,v) &\coloneqq \int_{\mathsf{X}} [r(x,a,x') + \beta(x,a,x')v(x')]P(x,a,x')\mathrm{d}x' \\ &= \mathbb{E}_{x,a}[r(x,a,x') + \beta(x,a,x')v(x')] \end{aligned} \tag{8}$$

for all $(x,a) \in \mathsf{G}$ and $v \in mb\mathsf{X}$, where

- $r\colon \mathsf{G} \times \mathsf{X} \to \mathbb{R}$ is a Borel measurable function, referred to as the *reward*,
- $\beta\colon \mathsf{G} \times \mathsf{X} \to \mathbb{R}_+$ be a Borel measurable and everywhere positive function, referred to as the *discount factors*, and
- $P\colon \mathsf{G} \times \mathsf{X} \to \mathbb{R}_+$ is a *stochastic kernel* on X contingent on current state and action; that is, $B \mapsto P(x,a,B)$ is a probability measure on X for all $(x,a) \in \mathsf{G}$ and $(x,a) \mapsto P(x,a,B)$ is a measurable function on G for all $B \in \mathscr{B}(\mathsf{X})$.

To simplify notation, denote β_σ, P_σ and r_σ as $\beta_\sigma(x,x') \coloneqq \beta(x,\sigma(x),x')$, $P_\sigma(x,x') \coloneqq P(x,\sigma(x),x')$, and $r_\sigma(x) \coloneqq \mathbb{E}_x^\sigma r(x,\sigma(x),x')$, respectively, for all x, x' and any $\sigma \in \Sigma$, where $\mathbb{E}_x^\sigma$ denotes the expectation under P_σ transition kernel conditioning on x. Also, denote $r(x,a) \coloneqq \mathbb{E}_{x,a} r(x,a,x')$ for any $(x,a) \in \mathsf{G}$. Given an MDP $\mathscr{M}$ and $\sigma \in \Sigma$, the policy operator T_σ following (8) becomes

$$T_\sigma v(x) \coloneqq r_\sigma(x) + \int_{\mathsf{X}} \beta_\sigma(x,x')v(x')P_\sigma(x,x')\mathrm{d}x' \qquad (x \in \mathsf{X}, v \in mb\mathsf{X}).$$

If T_σ has a unique fixed point $v_\sigma \in mb\mathsf{X}$, then iteration implies $v_\sigma = T_\sigma v_\sigma = T_\sigma^n v_\sigma$. Letting $n \to \infty$, if it converges, we have

$$v_\sigma(x) \coloneqq \mathbb{E}_x^\sigma \sum_{t=0}^{\infty} \left(\prod_{i=0}^{t-1} \beta_\sigma(X_i, X_{i+1}) \right) r_\sigma(X_t) \qquad (x \in \mathsf{X}). \tag{9}$$

where $\prod_{i=0}^{-1} \beta_i^\sigma = 1$ by convention and $\{X_t\}_{t \in \mathbb{N}_0}$ is a stochastic process such that $X_0 = x$ and X_{t+1} is generated by $P_\sigma(X_t, \cdot)$ for all $t \in \mathbb{N}_0$. We introduce the following regular conditions on an MDP.

Condition 3.1.

(i) Γ is nonempty, compact-valued, and continuous.

(ii) $(x, a) \mapsto r(x, a)$ is continuous and bounded on G.

(iii) β is bounded and strictly positive,

(iv) $(x, a) \mapsto \int_{\mathsf{X}} f(y)\beta(x, a, y)P(x, a, y)\mathrm{d}y$ is bounded and continuous on G for every $f \in cb\mathsf{X}$.

We say that an MDP $\mathscr{M}$ is *regular* if condition 3.1 is satisfied. Condition 3.1 ensures that $\mathscr{M}$ is a regular RDP and then the greedy policies exist by Lemma 2.1.

Lemma 3.1. *If an MDP $\mathscr{M}$ satisfies Condition 3.1 holds, then it is a regular RDP, and the following statements are true.*

(a) T_σ is a self-map on $mb\mathsf{X}$ for all $\sigma \in \Sigma$.

(b) T and T_σ are self-maps on $cb\mathsf{X}$ for all $\sigma \in \Sigma_C$.

(c) For all $v \in cb\mathsf{X}$, there exists a v-greedy policy.

To ensure the continuity of optimal policies or greedy policies, we introduce the conditions with concavity.

Condition 3.2.

(i) Γ is nonempty, continuous, compact-valued, and convex-valued.

(ii) $(x, a) \mapsto r(x, a)$ is continuous and bounded on G, and $a \mapsto r(x, a)$ is strictly concave for all $x \in \mathsf{X}$.

(iii) β is bounded and strictly positive,

(iv) $(x, a) \mapsto \int_{\mathsf{X}} f(y)\beta(x, a, y)P(x, a, y)\mathrm{d}y$ is bounded and continuous on G for every $f \in cb\mathsf{X}$,

(v) $a \mapsto \int_{\mathsf{X}} f(y)\beta(x, a, y)P(x, a, y)\mathrm{d}y$ is concave for all $x \in \mathsf{X}$ and $f \in cb\mathsf{X}$.

Condition 3.2 guarantees that HPI and OPI are well-behaved.

3.2. **Eventual Discounting.** Let $\mathscr{M}$ be an MDP. To this end, (3) becomes

$$|B(x, a, v) - B(x, a, w)| \leqslant \int_{\mathsf{X}} |v(x') - w(x')|\beta(x, a, x')P(x, a, x')\mathrm{d}x'$$

for all $(x, a) \in \mathsf{G}$ and $v, w \in mb\mathsf{X}$. The corresponding L_σ is defined by

$$L_\sigma h(x) = \int_{\mathsf{X}} |h(x')b(x, \sigma(x), x')P(x, \sigma(x), x')\mathrm{d}x' = \mathbb{E}_x^\sigma \beta_\sigma(x, x')h(x') \tag{10}$$

for all $x \in \mathsf{X}$ and $h \in mb\mathsf{X}$. We say that T_σ is *eventually discounting* if $\rho(L_\sigma) < 1$, or equivalently, there is $n_\sigma \in \mathbb{N}$ such that

$$d_{n_\sigma}^\sigma = \sup_x L_\sigma^{n_\sigma} \mathbb{1}(x) = \sup_x \mathbb{E}_x^\sigma \prod_{t=0}^{n_\sigma - 1} \beta_\sigma(X_t, X_{t+1}) < 1.$$

Assumption 3.1. For any $\sigma \in \Sigma$, either $\rho(L_\sigma) < 1$ or there exists $\{n_\sigma\} \in \mathbb{N}$ such that $d_{n_\sigma}^\sigma < 1$. Moreover, $\sup_{\sigma\in\Sigma} \rho(L_\sigma) < 1$ or $\sup_{\sigma\in\Sigma} d_{n_\sigma}^\sigma < 1$ for some $\{n_\sigma\}_{\sigma\in\Sigma} \subset \mathbb{N}$.

Assumption 3.2. There is $\ell \colon \mathsf{X}^2 \to \mathbb{R}$ such that $\beta(x, a, x')P(x, a, x') \leqslant \ell(x, x')$ for any $(x, a, x') \in \mathsf{G} \times \mathsf{X}$ and $\rho(L) < 1$, where L is defined by

$$Lh(x) = \int_{\mathsf{X}} h(x')\ell(x, x')\mathrm{d}x'$$

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

We say that $\mathscr{M}$ is *eventually discounting* if either Assumption 3.1 or 3.2 is satisfied. Since Assumption 3.2 implies $L_\sigma \mathbb{1} \leqslant L\mathbb{1}$ for any $\sigma \in \Sigma$, Assumption 3.2 is a sufficient condition to Assumption 3.1.

Lemma 3.2. *If $\mathscr{M}$ is regular and Assumption 3.2 holds, then there exists an $n \in \mathbb{N}$ such that $d_n^\sigma < 1$ for all $\sigma \in \Sigma$ and $\sup_\sigma \rho(L_\sigma) < 1$ for all $\sigma \in \Sigma$.*

Analogously, we can show that Assumption 3.1 implies Assumption 2.1. The following lemma shows the relationship between spectral radius $\rho(L_\sigma)$ and eventual discount factor $d_{n_\sigma}^\sigma$.

Lemma 3.3. *Suppose that $\mathscr{R}$ is regular. Then, the following statements are true.*

(a) For any $\sigma \in \Sigma$, $\rho(L_\sigma) < 1$ if and only if there is an $n_\sigma \in \mathbb{N}$ such that $d_{n_\sigma}^\sigma < 1$.
(b) For any $\sigma \in \Sigma$, $\rho(L_\sigma) = \lim_{n\to\infty}(d_n^\sigma)^{1/n}$.
(c) $\sup_{\sigma\in\Sigma} d_{n_\sigma}^\sigma < 1$ implies $\sup_{\sigma\in\Sigma} n_\sigma < \infty$
(d) $\sup_{\sigma\in\Sigma} \rho(L_\sigma) < 1$ if and only if $\sup_{\sigma\in\Sigma} d_{n_\sigma}^\sigma < 1$ for some $\{n_\sigma\}_{\sigma\in\Sigma} \subset \mathbb{N}$.

Lemma 3.3 provides a method to compute $\rho(L_\sigma)$ by taking the limit of $(d_n^\sigma)^{1/n}$. Moreover, it implies that $\rho(L_\sigma) < 1$ and $d_{n_\sigma}^\sigma < 1$ (for some $n_\sigma \in \mathbb{N}$) are equivalent.

Furthermore, 3.3 (d) shows that the statements in Assumption 3.1 are equivalent: $\sup_{\sigma\in\Sigma}\rho(L_\sigma)<1 \iff \sup_{\sigma\in\Sigma_C} d^\sigma_{n_\sigma}<1$ for $\{n_\sigma\}_{\sigma\in\Sigma}\subset\mathbb{N}$.

Example 3.1 (Firm valuation with stochastic interest rates)**.** Assume that the state space X is finite and follows a stochastic kernel $P\colon \mathsf{X}\times\mathsf{X}\to\mathbb{R}_+$. Suppose that the discount factors are

$$\beta_t := \frac{1}{1+r_t} \qquad (t\in\mathbb{N})$$

where $r_t = r(X_t)$ denotes the (real) interest rates following a stochastic process. If $\pi_t=\pi(X_t)$ is the profit at time t, then the expected present value of the firm is

$$v(x)=\mathbb{E}_x\sum_{t=0}^{\infty}\left(\prod_{i=0}^{t}\beta_i\right)\pi_t,$$

given current state $X_0=x$. Let $A(x,x')=P(x,x')/(1+r(x))$ for all $(x,x')\in\mathsf{X}\times\mathsf{X}$. If $\rho(A)<1$, then Assumption 3.2 is satisfied and $v=\pi+Av$ (Sargent and Stachurski, 2023). □

Example 3.2 (Uzawa Time Preferences)**.** Mendoza (1991), Schmitt-Grohé and Uribe (2003), Vasilev (2022a,b), and Izadi and Lamsoo (2022) study a small open economy where a representative household has Uzawa preference that the richer are more impatient than the poor. Uzawa preference has the merits that it stabilizes the small open economy and generates a non-degenerate distribution of wealth (Guest and McDonald, 2001). The household chooses consumption c_t and working hours h_t to maximize the utility

$$\begin{aligned}
&\mathbb{E}_0\sum_{t=0}^{\infty}\theta_t U(c_t,h_t)\\
&U(c,h)=\frac{(c-\nu^{-1}h^\nu)^{1-\gamma}}{1-\gamma}\\
&\theta_{t+1}=b(c_t,h_t)\theta_t,\quad t\geqslant 0 \qquad \text{with } \theta_0=1\\
&b(c,h)=(1+c-\nu^{-1}h^\nu)^{-\psi},
\end{aligned}$$

where $\nu>1$ is the labor supply elasticity, $\psi>0$ is the elasticity of discounting factor with respect to component $1+c-\nu^{-1}h^\nu$, and $\gamma>1$ measures the degree of relative risk aversion. The composite commodity $c_t-\nu^{-1}h^\nu$ is assumed to be positive, and then the utility is bounded above.[9] We can see that $b(c,h)<1$ when $c>\nu^{-1}h^\nu$.

[9]Otherwise, since $\gamma>1$, if $c\leqslant\nu^{-1}h^\nu$, then $U(c,h)\to\infty$ as $c\uparrow\nu^{-1}h^\nu$. Then, households can arbitrarily increase utility. Since the domain, $c\neq\nu^{-1}h^\nu$, is open, there is no maximizer in this case.

Hence, the feasible correspondence is the subset of $\{(c,h)\colon c > \nu^{-1}h^{\nu}\}$ such that the discount factors are strictly less than one. To solve the problem by discretization in programming, we can assume there is $\varepsilon > 0$ such that $c - \nu^{-1}h^{\nu} \geqslant \varepsilon > 0$ and then $b(c,h) \leqslant (1+\varepsilon)^{-\psi} < 1$ for any (c,h), so Assumption 2.1 holds.[10] □

Example 3.3. Durdu et al. (2009) consider a small open economy such that a representative household chooses the optimal consumption c_t and maximizes the preference

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \exp\left\{-\sum_{\tau=0}^{t-1} \psi \ln(1+c_\tau)\right\} \frac{c_t^{1-\gamma}}{1-\gamma},$$

where $\psi > 0$ is the elasticity of the rate of time preference with respect to $1+c$. Since $\beta_t = \exp\{-\psi \ln(1+c_t)\} < 1$ for $c_t > 0$, if we discretize the consumption space and consider inner solutions, then we have $c \geqslant \varepsilon$ and $\beta(c) \leqslant 1/(1+\varepsilon)^{\psi}$ for some $\varepsilon > 0$. □

Example 3.4 (Uzawa Time Preferences with Stochastic Discounting)**.** This example considers both the Uzawa time preference of Durdu et al. (2009) and state-dependent discounting of Hubmer et al. (2021):[11]

$$\beta_t = \beta(c_t, Z_t) = Z_t \exp\{-\psi \ln(1+c_t)\},$$

where c_t is consumption and Z_t is an AR(1) process follows

$$Z_{t+1} = \rho_Z Z_t + (1-\rho_Z)\mu_Z + \sigma_\varepsilon \varepsilon_{t+1} \qquad \{\varepsilon_t\} \overset{IID}{\sim} N(0,1) \tag{11}$$

with $\rho_Z = 0.992, \mu_Z = 0.944, \sigma_\varepsilon = 0.0006$. They discretize the process onto a grid of $N = 15$ states by Tauchen's method which allows us to write the operator L, defined in Assumption 2.2, as a matrix

$$L_{ij} = \beta(x_i)P(x_i, x_j), \qquad 1 \leqslant i, j \leqslant N.$$

The spectral radius of matrix L is 0.9469, computed by Stachurski and Zhang (2021). Then, since $\beta_t \leqslant Z_t$, there exists $n \in \mathbb{N}$ such that

$$\mathbb{E}_0 \beta_0 \beta_1 \cdots \beta_{n-1} \leqslant \mathbb{E}_0 Z_0 Z_1 \cdots Z_{n-1} < 1$$

[10] On the other hand, the steady state satisfies $\beta(c,h)(1+r) = 1$ so that $\beta(c,h) < 1$, where $r > 0$ is the real interest rate. For example, Vasilev (2022a) calibrate the parameters such that $b(c,h) = 0.982$. If we solve the model around the steady state, the discount factor is bounded away from one.

[11] The literature of state-dependent discounting with AR(1) process includes Hills and Nakata (2018), Hills et al. (2019) and Nakata (2016).

for any consumption path. □

Example 3.5 (Becker-Mulligan Time Preferences)**.** Becker and Mulligan (1997) propose the endogenous time preferences that are increasing in future-oriented capital. Stern (2006) considers Becker-Mulligan time preferences in an optimal growth model:

$$\sup_{\{c_t,s_t,k_t\}_{t\geqslant 0}} \sum_{t=1}^{\infty}\prod_{i=1}^{t-1}\beta(s_i)U(c_t)$$
$$\text{s.t. } c_t+\pi s_t+k_t\leqslant f(k_{t-1}), c_t\geqslant 0, s_t\geqslant 0, k_t\geqslant 0, \text{ for all } t\in\mathbb{N}$$

where k_0 is given, $\beta\colon [0,\infty)\to(0,\infty)$ is continuous, concave and strictly increasing, π is the price of s_t, and k_t denotes capital.[12] Stern (2006) assumes that f is continuous and strictly increasing, and there exists a $k_m\geqslant 0$ such that $\beta(k_m/\pi)<1$ and $f(k)<k$ whenever $k>k_m$. Hence, we have $\beta(s)\leqslant\beta(k_m/\pi)<1$ for all $s\in[0,k_m/\pi]$, so Assumption 2.1 is satisfied.

Erol et al. (2011) also consider a similar model:

$$\sup_{\{c_t,k_{t+1}\}_{t\geqslant 1}} \sum_{t=0}^{\infty}\prod_{i=1}^{t}\beta(k_i)U(c_t)$$
$$\text{s.t. } c_t+k_{t+1}\leqslant f(k_t), c_t\geqslant 0, k_t\geqslant 0, \qquad \text{for all } t\in\mathbb{N}$$

where k_0 is given. Erol et al. (2011) assumes that $\beta\colon[0,\infty)\to(0,\infty)$ is continuous, differentiable, strictly increasing and $\sup_{k>0}\beta(k)<1$, so Assumption 2.1 is satisfied.[13] □

Example 3.6 (Becker-Mulligan Preference with Stochastic Discounting)**.** This example modifies the Becker-Mulligan Preference of Erol et al. (2011) with stochastic uncertainty. Consider a model from Example 3.5:

$$\sup_{\{c_t,k_{t+1}\}_{t\geqslant 1}} \mathbb{E}_0\sum_{t=0}^{\infty}\left(\prod_{i=1}^{t}Z_{i-1}\beta(k_i)\right)U(c_t)$$
$$\text{s.t. } c_t+k_{t+1}\leqslant f(k_t), c_t\geqslant 0, k_t\geqslant 0, \qquad \text{for all } t\in\mathbb{N}$$

where k_0 is given, and $\{Z_t\}_{t\geqslant 0}$ is an AR(1) process satisfying that there exists $n\in\mathbb{N}$ such that $\mathbb{E}_0 Z_0Z_1\cdots Z_{n-1}<1$. Similar to Erol et al. (2011), assumes that

[12] The time preference $\beta(s_t)$ is interpreted as the degree to which generation t cares for generation $t+1$, while variable s_t represents actions or resources that the parent could take to strengthen the relationship with her child.

[13] Erol et al. (2011) also assumes that k has a compact support.

$\beta\colon [0,\infty) \to (0,\infty)$ is continuous, differentiable, strictly increasing and $\sup_{k>0} \beta(k) \leqslant 1$. To this end, Assumption 2.1 is satisfied that

$$\mathbb{E}_0 Z_0 \beta(k_1) Z_1 \beta(k_2) \cdots Z_{n-1} \beta(k_n) \leqslant \mathbb{E}_0 Z_0 Z_1 \cdots Z_{n-1} < 1.$$

for any $\{k_t\}_{t\geqslant 0}$. □

3.3. **Optimality.** In this section, we present the optimality of an MDP with eventual discounting. Our results show that the standard methods to compute the optimal value and policies hold for the state-action-dependent dynamic programming such as Uzawa or Becker-Mulligan time preferences with stochastic discounting in Example 3.4 and Example 3.6.

Since we can show that Assumption 3.1 implies Assumption 2.1, the following theorem follows directly from Theorem 2.1. That is, an eventually discounting MDP is an eventually discounting RDP when $\sup_{\sigma\in\Sigma} \rho(L_\sigma) < 1$.

In the following theorem, define

$$w_\sigma = \mathbb{E}_x^\sigma \sum_{t=0}^{\infty} \prod_{i=0}^{t-1} \beta_i^\sigma \quad \text{and} \quad \lambda_\sigma = \sup_{x\in \mathsf{X}} \frac{w_\sigma(x) - 1}{w_\sigma(x)}.$$

Let $w(x) = \inf_\sigma w_\sigma(x)$ for all $x \in \mathsf{X}$.

Theorem 3.1. *If $\mathscr{M}$ is regular and Assumption 3.1 holds, then the following statements are true.*

(a) T_σ is eventually contracting on $mb\mathsf{X}$ and contracting on $(mb\mathsf{X}, \|\cdot\|_{w_\sigma})$ with modulus λ_σ for all $\sigma \in \Sigma$,
(b) T is contracting on $(cb\mathsf{X}, \|\cdot\|_w)$ with modulus $\sup_{\sigma\in\Sigma} \lambda_\sigma$,
(c) v^* is the unique solution to the Bellman equation in V,
(d) VFI converges to v^*,
(e) Bellman's principle of optimality holds, and
(f) at least one optimal (continuous) policy exists,

If, in addition, Condition 3.2 holds, then

(α) HPI converges to σ^*, and
(β) OPI converges to v^*,

where all the iterated greedy policies to HPI and OPI are continuous.

Theorem 3.1 generalizes the traditional constant-discounting dynamic programming theory to the case of state-action-dependent discounting. In particular, Theorem 3.1 shows that if $\mathbb{R}$ is regular and eventually discounting, then the optimal policy exists, Bellman's principle of optimality holds, and VFI converge to v^*. Moreover, Theorem 3.1 proves that T or T_σ are both eventually contracting in supremum norm and contracting in some weighted supremum norm. The latter immediately allows us to analyze the convergence time.

To this end, Theorem 3.1 demonstrates that the optimality of Example 3.4 of Uzawa preferences or Example 3.6 of Becker-Mulligan time preferences holds true. In addition, the optimal policy can be obtained by the v^*-greedy policy, where v^* is computed by VFI.

4. Necessity of Eventual Discounting

In this section, we investigate the necessity of eventual discounting or spectral radius conditions for both the existence and uniqueness of the policy values or the fixed points of policy operators in the environment of compact state space. Since the convergence of Howard policy iteration and optimistic policy iteration depend on the global stability of policy operators, eventual discounting for policy operators is essential to an MDP. Furthermore, we demonstrate that the convergence rate of T_σ is the corresponding spectral radius of the expected discounting operator.

4.1. An MDP with a Compact State Space.

We first establish an MDP with a compact state space and positive rewards. Let $\mathscr{M} = (\mathsf{X}, \mathsf{A}, \Gamma, B)$ be an MDP, where value aggregator B follows (8). We consider the following regular conditions. Let $mb\mathsf{X}_+$ (resp. $cb\mathsf{X}_+$) denote the set of everywhere positive functions in $mb\mathsf{X}$ (resp. $cb\mathsf{X}$.) We consider the following regular conditions throughout this section.

Condition 4.1.

(i) X is compact,

(ii) β and P are continuous, and

(iii) $(x, a) \mapsto r(x, a)$ is strictly positive and continuous on G, and $a \mapsto r(x, a)$ is strictly concave for all $x \in \mathsf{X}$,

(iv) Γ is nonempty, continuous, compact-valued and convex-valued,

(v) $(x,a) \mapsto \int_{\mathsf{X}} f(y)\beta(x,a,y)P(x,a,y)\mathrm{d}y$ is continuous for all $f \in cb\mathsf{X}$, and

(vi) $a \mapsto \int_{\mathsf{X}} f(y)\beta(x,a,y)P(x,a,y)\mathrm{d}y$ is concave for all $x \in \mathsf{X}$ and $f \in cb\mathsf{X}$,

Condition 4.1 is strict since it is difficult that Condition 4.1 (iii) and (vi) hold at the same time. For simplicity, Condition 4.1 (vi) holds if β and P are only state-dependent, or if β does not depend on future states, P is not action-dependent, and $a \mapsto \beta(x,a)$ is concave.

Condition 4.1 is similar to Condition 3.2. The main differences are that X is compact and r is strictly positive, which can be attained by scaling up a bounded reward. Then, a continuous v-greedy policy exists for any $v \in cb\mathsf{X}$. Moreover, T and T_σ are self-maps on $cb\mathsf{X}_+$ for $\sigma \in \Sigma_C$.

Lemma 4.1. *If Condition 4.1 holds, then the following statements are true.*

(a) T_σ is a self-map on $mb\mathsf{X}_+$ for all $\sigma \in \Sigma$.

(b) T and T_σ are self-maps on $cb\mathsf{X}_+$ for all $\sigma \in \Sigma_C$.

(c) For all $v \in cb\mathsf{X}_+$, there exists a continuous v-greedy policy.

An MDP $\mathscr{M}$ is *ergodic* if for each policy $\sigma \in \Sigma$ the induced P_σ-Markov chain is ergodic/irreducible. The next assumption guarantees that $\mathscr{M}$ is ergodic. Given $\sigma \in \Sigma$, let P_σ^n be defined by $P_\sigma^n = P_\sigma$ and $P_\sigma^n = \int P_\sigma(x,y)P_\sigma^{n-1}(y,z)\mathrm{d}y$ for all $(x,z) \in \mathsf{X} \times \mathsf{X}$ and $i \in \mathbb{N}$.

Assumption 4.1. For all $\sigma \in \Sigma$, there exists an $n \in \mathbb{N}$ such that the transition density P_σ^n is positive everywhere.

Fix $\sigma \in \Sigma_C$. Define $k_\sigma(x,x') \coloneqq \beta_\sigma(x,x')P_\sigma(x,x')/\bar{\beta}_\sigma(x)$ for all $(x,x') \in \mathsf{X} \times \mathsf{X}$, where

$$\bar{\beta}_\sigma(x) \coloneqq \int \beta_\sigma(x,x')P_\sigma(x,x')\mathrm{d}x'.$$

Note that since β is strictly positive, $\bar{\beta}$ is strictly positive. We can write L_σ as

$$L_\sigma v(x) = \bar{\beta}_\sigma(x) \int v(x')k_\sigma(x,x')\mathrm{d}x' \qquad (x \in \mathsf{X}, v \in cb\mathsf{X}_+).$$

Given $\sigma \in \Sigma_C$, denote e_σ as the eigenvector of L_σ corresponding to its spectral radius satisfying $L_\sigma e_\sigma = \rho(L_\sigma)e_\sigma$. We can show that L_σ^2 is a compact operator such that $e_\sigma \in cb\mathsf{X}$ exists and is everywhere positive. The following proposition shows that

the spectral radius and the corresponding eigenfunction dominate the contraction of the policy operator, in the sense that $\rho(L_\sigma) < 1$ if and only if T_σ is contracting on $(cb\mathsf{X}_+, \|\cdot\|_{e_\sigma})$ with modulus $\rho(L_\sigma)$.

Proposition 4.1. *If Assumption 4.1 and Condition 4.1 hold, then for all $\sigma \in \Sigma_C$ the following statements are equivalent.*

(a) $\rho(L_\sigma) < 1$.
(b) There exists an $n_\sigma \in \mathbb{N}$ *such that* $d^\sigma_{n_\sigma} < 1$.
(c) T_σ *is globally stable on* $cb\mathsf{X}_+$.
(d) T_σ *is contracting on* $(cb\mathsf{X}_+, \|\cdot\|_{e_\sigma})$ *with modulus* $\rho(L_\sigma)$.
(e) T_σ *has a fixed point in* $cb\mathsf{X}_+$.
(f) T_σ *has a unique fixed point in* $cb\mathsf{X}_+$.

Moreover, if $\rho(L_\sigma) \geqslant 1$, then T_σ has no fixed point in $cb\mathsf{X}_+$.

Since $L_\sigma e_\sigma = \rho(L_\sigma) e_\sigma$ for any $\sigma \in \Sigma_C$, we can easily verify that Assumption 2.1 holds if $\sup_{\sigma\in\Sigma_C} \rho(L_\sigma) < 1$. Thus, if the MDP is eventually discounting, then the optimal properties in Theorem 3.1 are true. We summarize the results in the following theorem, which is a corollary of Theorem 2.1.

To apply Theorem 2.1, we use the eigenvectors as the weighting vector for the weighted supremum. Without loss of generality, we normalize e_σ such that $e_\sigma(x) \geqslant 1$ for all $x \in \mathsf{X}$ and $\sigma \in \Sigma_C$. In this case, we have $e \coloneqq \inf_\sigma e_\sigma \geqslant \mathbb{1}$. The normalization is well-defined since X is compact and e_σ is continuous and everywhere positive, so there exists an $\underline{e_\sigma} \in \mathbb{R}$ such that $\inf_x e_\sigma(x) \geqslant \underline{e_\sigma} > 0$.

Theorem 4.1. *Suppose that Condition 4.1 and Assumption 4.1 hold. If Assumption 3.1 holds, then the following statements are true.*

(a) T_σ *is eventually contracting on* $mb\mathsf{X}_+$ *for all* $\sigma \in \Sigma$, *and* T_σ *is contracting with modulus* $\rho(L_\sigma)$ *on* $(cb\mathsf{X}_+, \|\cdot\|_{e_\sigma})$ *for all* $\sigma \in \Sigma_C$,
(b) T *is contracting with modulus* $\sup_{\sigma\in\Sigma_C} \rho(L_\sigma)$ *on* $(cb\mathsf{X}_+, \|\cdot\|_e)$,
(c) v^* *is the unique solution to the Bellman equation in* $cb\mathsf{X}_+$,
(d) HPI converges to σ^*,
(e) VFI converges to v^*,
(f) OPI converges to v^*,

(g) Bellman's principle of optimality holds, and

(h) at at least one optimal continuous policy exists.

where all the iterated greedy policies to HPI and OPI are continuous.

5. Applications

In this section, we apply the main results to optimal growth and optimal default problems.

5.1. Optimal Growth.

We can apply the main results to the following examples of optimal growth models.

Example 5.1 (One-sector Optimal Growth, Continued)**.** Let $\mathscr{M}$ be the MDP defined in Example 2.1. The Bellman equation is

$$v(k,z) = \sup_{c\in[0,zF(k,1)]} \Big\{ U(c) + \beta(z,c) \int_{\mathsf{X}} v(k',z') Q(z,\mathrm{d}z') \mathbb{1}\{k' = (1-\delta)k + zF(k,1) - c\} \mathrm{d}k' \Big\}.$$

Assume that there is an $n \in \mathbb{N}$ such that

$$d_n^\sigma \leqslant d_n := \sup_x \sup_\sigma \mathbb{E}_x^\sigma \prod_{t=0}^{n-1} \beta(Z_t, \sigma(Z_t)) < 1$$

for all $\sigma \in \Sigma$. Then, we have $\sup_{\sigma\in\Sigma_C} d_{n_\sigma}^\sigma < 1$, so Theorem 3.1 shows that the optimality of dynamic programming holds and the related dynamic programming algorithms converge to the unique optimal value. □

Example 5.2 (Uzawa Time Preference with Stochastic Discounting, Continued)**.** This example continues Example 3.4 by modifying the optimal saving problem in Hubmer et al. (2021) to the case of Uzawa endogenous time preference. Following Hubmer et al. (2021), suppose that the policy operator is

$$T_\sigma v(x,z) = u(R(x,z)x + y(x,z) - \sigma(x)) + \beta(R(x,z)x + y(x,z) - \sigma(x), z) \int v(\sigma(x), z') Q(z, \mathrm{d}z')$$

where $x \in \mathsf{X} := \mathbb{R}_+$ is the present asset, z is the exogenous shocks in $\mathbb{R}$ as Example 3.4, $R(x,z)$ is the gross return rates on asset holdings, $y(x,z)$ is the labor net income,

$R(x,z)x + y(x,z) - \sigma(x) = c$ is the consumption, and $\sigma(x)$ is the asset or saving leaving to the next period. Assume that β is defined as Example 3.4 and $\{Z_t\}$ follows Example 3.4. Let the utility function be

$$u(c) \coloneqq \frac{c^{1-\gamma}}{1-\gamma} \qquad (\gamma > 1).$$

The feasible correspondence is

$$\Gamma(x,z) \coloneqq \{x' \in \mathbb{R} : \bar{x} \leqslant x' \leqslant R(x,z)x + y(x,z)\}$$

To solve the problem numerically, we discretize both X and Z into finite grid points. To this end, the reward function is bounded, and the continuity assumptions in Condition 3.1 are satisfied. As discussed in Example 3.4, the MDP is eventually discounting. Therefore, all of the conclusions in Theorem 3.1 hold. □

5.2. **Optimal Default.** This example considers an optimal saving problem with default following Arellano (2008), Hatchondo et al. (2009), Yue (2010), Hatchondo et al. (2016) and Ma et al. (2022). We assume state-action-dependent discounting. A country with current assets w_t chooses between continuing to participate in international financial markets and defaulting. Let $y_t = y(z_t, \xi_t)$ be output, where $\{z_t\}$ is a Markov process and $\{\xi_t\}$ is an IID shock. Assume that default results in permanent exclusion from financial markets which yields the lifetime value

$$v^d(y,z) = \mathbb{E}_z \sum_{t=0}^{\infty} \prod_{i=0}^{t-1} \beta^d(z_i) u(y_t),$$

where $\beta^d(z)$ represents the stochastic discount dependent on state z given defaulting. The value of continued participation in the financial market is

$$v^c(w,y,z) = \sup_{-b \leqslant w' \leqslant R(w+y)} u(w+y-w'/R) + \beta(z,w')\mathbb{E}_z v(w',y',z')$$

where $b > 0$ is a constant borrowing constraint, $\beta(z,w')$ is the discount factor depends on state z and wealth w', in the spirit of Becker-Mulligan time preferences in Example 3.5, and v is the value function satisfying

$$v(w,y,z) = \max\{v^d(y,z), v^c(w,y,z)\}.$$

The Bellman equation is

$$v(w,y,z) = \sup_{\substack{\delta\in\{0,1\}\\ -b\leqslant w'\leqslant R(w+y)}} \left\{ \delta \left[\mathbb{E}_z \sum_{t=0}^{\infty} \prod_{i=0}^{t-1} \beta^d(z_i) u(y_t) \right] \right.$$
$$\left. + (1-\delta)\Big[u(w+y-w'/R) + \beta(z,w')\mathbb{E}_z v(w',y',z') \Big] \right\}.$$

Let $\mathsf{X} = W \times Y \times Z$ where W, Y and Z are domains of w_t, y_t and z_t, respectively. Assume that u is continuous and either u is bounded or z_t and ξ_t have compact supports. Suppose that β^d and β are continuous, bounded, strictly positive, that there are $n, m \in \mathbb{N}$ such that

$$\sup_{(w,z,y)\in\mathsf{X}} \mathbb{E}_z \prod_{t=0}^{n-1} \beta^d(z_t) < 1, \text{ and}$$
$$\sup_{\sigma\in\Sigma} \sup_{(w,z,y)\in\mathsf{X}} \mathbb{E}_z \prod_{t=0}^{m-1} \beta(z_t, \sigma_{w'}(w_t,y_t,z_t)) < 1$$

where $\sigma(w,y,z) = (\sigma_\delta(w,y,z), \sigma_{w'}(w,y,z))$ is the policy of action (δ, w') given state $(w,y,z) \in \mathsf{X}$. Therefore, v^d is bounded and continuous and the MDP is eventually discounting, so all conclusions in Theorem 3.1 hold.

5.3. **Asset Pricing.** There is an ex-dividend contract that trades at prices Π_t and pays dividend D_t. To this end, purchasing this contract at t and selling at $t+1$ pays $\Pi_{t+1} + D_{t+1}$. Let the Lucas stochastic discount factor be

$$M_{t+1} = \bar{\beta} \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma},$$

where $\bar{\beta} > 0$ is a constant discount factor measuring the impatience of the agent. Given the absence of arbitrage, the price at time t must satisfy

$$\Pi_t = \mathbb{E}_t M_{t+1}(\Pi_{t+1} + D_{t+1}). \tag{12}$$

Let $\{X_t\}_{t\geqslant 0}$ be a P-Markov process on a compact state space X. Suppose that the dividend growth obeys

$$\ln \frac{D_{t+1}}{D_t} = \mu_d + X_t + \sigma_d \eta_{d,t+1}$$

where $\{\eta_{d,t}\}_{t\geqslant 0}$ is IID and standard normal. Moreover, consumption growth obeys

$$\ln \frac{C_{t+1}}{C_t} = \mu_c + X_t + \sigma_c \eta_{c,t+1}$$

where $\{\eta_{c,t}\}_{t\geqslant 0}$ is IID and standard normal. Let $V_t \coloneqq \Pi_t/D_t$ be the price-dividend ratio. Then, we obtain

$$\begin{aligned} V_t = \frac{\Pi_t}{D_t} &= \mathbb{E}_t\left[M_{t+1}\frac{D_{t+1}}{D_t}\left(\frac{P_{t+1}}{D_{t+1}}+1\right)\right] \\ &= \mathbb{E}_t\left[\bar{\beta}\exp(-\gamma\mu_c+\mu_d+(1-\gamma)X_t-\gamma\sigma_c\eta_{c,t+1}+\sigma_d\eta_{d,t+1})\,(V_{t+1}+1)\right]. \end{aligned} \tag{13}$$

Conditioning on $X_t = x$, (13) yields the value function

$$v(x) = \int_{\mathsf{X}} \bar{\beta}\exp\left(-\gamma\mu_c+\mu_d+(1-\gamma)x+\frac{\gamma^2\sigma_c^2+\sigma_d^2}{2}\right)(1+v(x'))P(x,x')\mathrm{d}x'$$

for all $x \in \mathsf{X}$ and $v \in cb\mathsf{X}$. Define $A\colon cb\mathsf{X} \to cb\mathsf{X}$ by

$$Af(x,x') \coloneqq \int_{\mathsf{X}} \bar{\beta}\exp\left(-\gamma\mu_c+\mu_d+(1-\gamma)x+\frac{\gamma^2\sigma_c^2+\sigma_d^2}{2}\right)f(x')P(x,x')\mathrm{d}x'$$

for all $x \in \mathsf{X}$ and $f \in cb\mathsf{X}$. We define the corresponding discount function

$$\beta(x) \coloneqq \bar{\beta}\exp\left(-\gamma\mu_c+\mu_d+(1-\gamma)x+\frac{\gamma^2\sigma_c^2+\sigma_d^2}{2}\right)$$

for all $x \in \mathsf{X}$. Now, if P satisfies the Feller property, then a version of Theorem 3.1 without policy (or Lemma A.14) shows that A has a unique fixed point if $\rho(A) < 1$. If we further assume that P is continuous and there is $n \in \mathbb{N}$ such that P^n is everywhere positive, then a version of Proposition 4.1 without policy shows that A has a unique fixed point if and only if $\rho(A) < 1$.

5.3.1. *Incomplete Market with Subjective Discounting.* This section considers the asset pricing with heterogeneous expectations in Harrison and Kreps (1978). Investors are risk-neutral. Suppose that there are finitely many investor classes, denoted by set A. Agents in each class $a \in \mathsf{A}$ have a subjective probability distribution $P_a\colon \mathsf{X}\times\mathsf{X} \to \mathbb{R}_+$. Fix the random states $\{X_t\} \subset \mathsf{X}$. We further assume that agents in class $a \in \mathsf{A}$ have subjective time preferences such that $\beta_t = \beta_a(X_t, X_{t+1})$ for $a \in \mathsf{A}$. Let $\Pi_t = \pi(X_t)$ and $D_t = d(X_t)$ for all t. Harrison and Kreps (1978) shows that the price scheme is consistent if and only if[14]

$$\Pi_t = \max_{a\in\mathsf{A}} \mathbb{E}^a[\beta_t(\Pi_{t+1}+D_{t+1})].$$

[14]A price is called consistent when it prevents any investor from achieving an excessive expected return through adroit and legitimate speculation.

The Bellman equation is

$$\begin{aligned}\Pi_t &= \max_{a\in\mathsf{A}} \mathbb{E}^a[\beta_t(\Pi_{t+1} + D_{t+1})].\\ \pi(x) &= \max_{a\in\mathsf{A}} \mathbb{E}^a[\beta_a(x,x')(\pi(x') + d(x'))]\\ &= \max_{a\in\mathsf{A}} \int_{\mathsf{X}} \beta_a(x,x')[\pi(x') + d(x')]P_a(x,x')\mathrm{d}x'\end{aligned}$$

for $x \in \mathsf{X}$. Define $r(x,a) \coloneqq \int_{\mathsf{X}} \beta_a(x,x')\pi(x')P_a(x,x')\mathrm{d}x'$ and

$$T_a\pi(x) \coloneqq r(x,a) + \int_{\mathsf{X}} \beta_a(x,x')\pi(x')P_a(x,x')\mathrm{d}x'$$

for all $a \in \mathsf{A}$ and $x \in \mathsf{X}$. If r is bounded and continuous, and the MDP is eventually discounting that either Assumption 3.1 or 3.2 holds, then Theorem 3.1 shows that the optimal price π^* is unique, the optimality of dynamic programming holds, and the VFI will converge to π^*.

6. Unbounded Rewards

In this section, we study an eventually discounting MDP with an unbounded reward function, which generalizes the main results in Section 3.

6.1. MDP with Unbounded Rewards.

Throughout this section, we assume the following regular condition with unbounded reward $r\colon \mathsf{G} \to \mathbb{R} \cup \{-\infty, \infty\}$.[15]

Condition 6.1.

(i) Γ is nonempty, compact-valued.

(ii) r is u.s.c..

(iii) β is bounded, continuous, and strictly positive.

(iv) $a \mapsto \int_{\mathsf{X}} f(y)\beta(x,a,y)P(x,a,y)\mathrm{d}y$ is continuous on $\Gamma(x)$ for all $x \in \mathsf{X}$ and for $f \in cb\mathsf{X}$.

[15]The conditions for the case of measurable functions are presented in the appendix.

If Condition 6.1 is satisfied, define spaces $\mathcal{V}$ and $\mathcal{G}$ by

$$\begin{aligned}\mathcal{V} &\coloneqq \{v\colon \mathsf{X} \to \mathbb{R} \cup \{-\infty\}\colon v \text{ is u.s.c., } v/\kappa \text{ is bounded above,} \\ &\qquad\qquad \text{and } (x,a) \mapsto \mathbb{E}_{x,a} v(x') \text{ is bounded below}\}, \\ \mathcal{G} &\coloneqq \{g\colon \mathsf{G} \to \mathbb{R}\colon g \text{ is u.s.c., } \|g\|_\kappa < \infty, \\ &\qquad\qquad \text{and } a \mapsto g(x,a) \text{ is u.s.c. on } \Gamma(x) \text{ for all } x \in \mathsf{X}\}.\end{aligned}$$

The regular conditions, with the assumptions below, ensure the existence of a greedy policy. In this section, we say that an MDP is *regular* if Condition 6.1 holds.

The pairs $(\mathcal{V}, \|\cdot\|_\kappa)$ and $(\mathcal{G}, \|\cdot\|_\kappa)$ are Banach spaces. Define the maximal reward function $\bar{r}$ and (conditionally) expected maximal reward function $\hat{r}$ as

$$\bar{r} \coloneqq \sup_{a \in \Gamma(x)} r(x,a) \quad (x \in \mathsf{X}) \quad \text{and} \quad \hat{r}(x,a) \coloneqq \mathbb{E}_{x,a} \bar{r}(x) \quad ((x,a) \in \mathsf{G}). \tag{14}$$

We introduce a mild assumption that $\hat{r}$ is bounded below and $\bar{r}$ is bounded above by some function following Ma et al. (2022).

Assumption 6.1.

(i) There exist a $\kappa \in mb\mathsf{X}$ such that $\kappa\colon \mathsf{X} \to [1,\infty)$ and a constant $d \geqslant 0$ such that $\bar{r}(x) \leqslant d\kappa(x)$ and $a \mapsto \mathbb{E}_{x,a}\kappa(x')$ is continuous for all $x \in \mathsf{X}$. If Condition 6.1 holds, then κ is also continuous.

(ii) There exist constant $n \in \mathbb{N}$ and $\alpha \in \mathbb{R}$ and a linear operator $L\colon \mathcal{V} \to \mathcal{V}$ such that $\rho(L) < 1$, $\|L\| < \infty$, $\|L^n\| < 1$, $\alpha \in (0, 1/\|L^n\|^{1/n})$ and

$$\mathbb{E}_{x,a}\beta(x,a,x')\kappa(x')L^t\mathbb{1}(x') \leqslant \alpha\kappa(x)L^{t+1}\mathbb{1}(x)$$

for all $(x,a) \in \mathsf{G}$ and $t \in \{0,1,\ldots,n\}$.

(iii) $\hat{r}$ is bounded below.

If $\kappa \equiv \mathbb{1}$ and $\alpha = 1$, then Assumption 6.1 (ii) is the same as Assumption 2.2. Hence, Assumption 2.2 extends the eventual contracting of Assumption 2.2.

Recall that the Bellman equation is

$$v(x) = \sup_{a \in \Gamma(x)} \{r(x,a) + \mathbb{E}_{x,a}[\beta(x,a,x')v(x')]\}.$$

Define the *action-value function* $g(x,a) \coloneqq \mathbb{E}_{x,a}\beta(x,a,x')v(x')$ for any $(x,a) \in \mathsf{G}$, which is the expected (discounted) future value conditioning on (x,a). The Bellman

equation can be written as $v(x) = \sup_{a\in\Gamma(x)}\{r(x,a) + g(x,a)\}$. Changing (x,a) to (x',a'), multiplying $\beta(x,a,x')$, and taking expectation yield

$$g(x,a) = \mathbb{E}_{x,a}\left[\beta(x,a,x') \sup_{a'\in\Gamma(x')}\{r(x',a') + g(x',a')\}\right] \tag{15}$$

for all $(x,a) \in \mathsf{G}$. Define the *expected value operator* $E\colon \mathbb{R}^{\mathsf{X}} \to \mathbb{R}^{\mathsf{G}}$ by

$$Ev(x,a) \coloneqq \mathbb{E}_{x,a}\beta(x,a,x')v(x') \qquad ((x,a) \in \mathsf{G}, v \in \mathbb{R}^{\mathsf{X}}).$$

Define the *maximum value operator* $M\colon \mathcal{G} \to \mathcal{V}$ by

$$Mg(x) = \sup_{a\in\Gamma(x)}\{r(x,a) + g(x,a)\} \qquad (x \in \mathsf{X}, g \in \mathcal{G}).$$

We say that a policy $\sigma \in \Sigma$ is *g-greedy* if $r(x,\sigma(x)) + g(x,\sigma(x)) = Mg(x)$ for all $x \in \mathsf{X}$. Let g^* be the solution to (15). We will show that the value function is $v^*(x) = \sup_{a\in\Gamma(x)}\{r(x,a) + g^*(x,a)\}$ for all $x \in \mathsf{X}$, and a policy $\sigma \in \Sigma$ is optimal if and only if σ is g^*-greedy.

Define *expected value Bellman operator* R on $\mathcal{G}$ by

$$Rg(x,a) \coloneqq \mathbb{E}_{x,a}\left[\beta(x,a,x') \sup_{a'\in\Gamma(x')}\{r(x',a') + g(x',a')\}\right] \qquad ((x,a) \in \mathsf{G}, g \in \mathcal{G}).$$

The *action-value function iteration* (AFI) is the iteration $\{g_t\}_{t\geqslant 0} \subset \mathcal{G}$ such that $g_{t+1} = Rg_t$ for all $t \in \mathbb{N}_0$ with $g_0 \in \mathcal{G}$.

Theorem 6.1. *If $\mathcal{M}$ is regular and Assumption 6.1 holds, then the following statements are true.*

(a) v^* and v_σ are well-defined for all $\sigma \in \Sigma$,

(b) R is eventually contracting on $(\mathcal{G}, \|\cdot\|_\kappa)$ and T is eventually contracting on $(\mathcal{V}, \|\cdot\|_\kappa)$,

(c) R admits a unique fixed point g^* in $\mathcal{G}$ and T admits a unique fixed point v^* in $\mathcal{V}$,

(d) $v^* = Mg^* \in \mathcal{V}$ and $g^* = Ev^* \in \mathcal{G}$,

(e) VFI converges to v^*, AFI converges to g^*,

(f) at least one optimal policy exists, and

(g) a feasible policy is optimal if and only if it is g^*-greedy if and only it is v^*-greedy.

6.2. **Application in Optimal Savings.** An agent solves an optimal savings problem with borrowing constraint:

$$\begin{aligned} &\sup \mathbb{E} \sum_{t=0}^{\infty} \prod_{i=0}^{t-1} \beta_i u(c_t) \\ \text{s.t. } & 0 \leqslant c_t \leqslant w_t \\ & w_{t+1} = R_{t+1}(w_t - c_t) + y_{t+1} \\ & (w_0, y_0) \text{ given.} \end{aligned} \tag{16}$$

Here $\beta_t \in (0, \infty)$ is the discount factor, $u \colon \mathbb{R}_+ \to \mathbb{R} \cup \{-\infty, \infty\}$ is a utility function, c_t denotes consumption, $w_t \geqslant 0$ denotes wealth, $y_t \geqslant 0$ denotes non-financial income, and R_t denotes the gross rate of return on financial income. Assume that asset return, income, and discount factors satisfy

$$R_t = R(\varepsilon_t), \quad y_t = y(z_t, \varepsilon_t), \quad \text{and} \quad \beta_t = \beta(z_t, z_{t+1})$$

where β is strictly positive, bounded and continuous, z_t is a Markov process with state space Z and ε_t is an IID shock that could be vector-valued. We first note that Condition 6.1 is satisfied. The Bellman equation is

$$v(w, z) = \sup_{0 \leqslant c \leqslant w} \{u(c) + \mathbb{E}_z[\beta(z, z')v(R(\varepsilon')(w - c) + y(z', \varepsilon'), z')]\}$$

To this end, the state is $x = (w, z) \in \mathbb{R}_+ \times \mathsf{Z}$ and the action is $a = c \in \Gamma(x) = [0, w]$. Suppose that u is u.s.c., increasing, $\inf_z \mathbb{E}u(y(z', \varepsilon')) > -\infty$, and there exist $p > 0$ and $q > 1$ such that

$$u(c) \leqslant pc + q \text{ for all } c > 0.$$

Define the weighting function by $\kappa(x) = pw + q$ for all $x = (w, z)$. To ensure Assumption 6.1, assume

$$\sup_z \mathbb{E}_z \beta(z, z') y(z', \varepsilon') < \infty, \quad \text{and} \quad \mathbb{E}R(\varepsilon') \geqslant 1, \tag{17}$$

In addition, assume that there is $m \in \mathbb{N}$ such that

$$\sup_{z \in \mathsf{Z}} \mathbb{E}_z \beta_0 \beta_1 \beta_2 \cdots \beta_{m-1} (\mathbb{E}R(\varepsilon'))^m < 1.$$

This assumption is the eventual discounting for Assumption 6.1 (b). Define operator L by $Lh(x) \coloneqq \mathbb{E}_z \beta(z, z') h(x') \mathbb{E}R(\varepsilon')$ for all $x \in \mathsf{X}$.

Lemma 6.1. *If $\mathscr{M}$ follows (16) and all the above corresponding assumptions hold, then there is large enough $q(n) > 1$ such that*

$$\mathbb{E}_{x,a}\beta(x,a,x')\kappa(x')L^n\mathbb{1}(x') \leqslant \kappa(x)L^{n+1}\mathbb{1}(x)$$

for all $x \in \mathsf{X}$ and all $n \in \mathbb{N}$.

By Lemma 6.1, it remains to show $\rho(L) < 1$. Since $\sup_{z\in\mathsf{Z}} \mathbb{E}_z\beta_0\beta_1\beta_2\cdots\beta_{m-1}(\mathbb{E}R(\varepsilon'))^m < 1$ for some $m \in \mathbb{N}$, we have $\|L^m\mathbb{1}\| < 1$ and then $\rho(L) < 1$. Therefore, Assumption 6.1 is satisfied, and then all conclusions of Theorem 6.1 hold.

7. Extension: Risk-Sensitive Preferences

We study the risk-sensitive preference models with state-action-dependent discounting in this section, where the agent is risk-averse in future utility and future consumption. Let an MDP satisfy the following regular conditions.

Condition 7.1.

(i) $(x,a) \mapsto r(x,a)$ is u.s.c.,
(ii) β is bounded and strictly positive,
(iii) Γ is nonempty, compact-valued, and continuous,
(iv) $(x,a) \mapsto (-\beta(x,a)/\theta)\ln\int \exp(-\theta h(y))P(x,a,y)\mathrm{d}y$ is u.s.c. for any $h \in cb\mathsf{X}$,

Let $\theta > 0$ be the agents' risk-sensitive coefficient. For any feasible policy $\sigma \in \Sigma$ and Borel measurable function $v\colon \mathsf{X} \to \mathbb{R}\cup\{-\infty\}$, let

$$T_\sigma v(x) \coloneqq r(x,\sigma(x)) - \frac{\beta(x,\sigma(x))}{\theta}\ln\mathbb{E}_x^\sigma e^{-\theta v(x')} \qquad (x \in \mathsf{X}). \tag{18}$$

The *σ-value function* is defined by

$$v_\sigma(x) = \limsup_{x\to\infty} T_\sigma^n\bar{r}(x) \qquad (x \in \mathsf{X}).$$

The associated Bellman equation is

$$v(x) = \sup_{a\in\Gamma(x)}\left\{r(x,a) - \frac{\beta(x,a)}{\theta}\ln\mathbb{E}_{x,a}e^{-\theta v(x')}\right\} \ (x \in \mathsf{X}). \tag{19}$$

Suppose that the state process evolves following

$$x_{t+1} = f(x_t, a_t, \varepsilon_{t+1}), \tag{20}$$

where f is Borel measurable, and $\{\varepsilon_t\}$ is an IID process taking values in $\mathbb{R}^m$. For each t, let $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{mt})$.

Assumption 7.1.

(i) $r\colon \mathsf{G} \to \mathbb{R}$ and $\beta\colon \mathsf{G} \to \mathbb{R}$ are increasing in x,

(ii) r_σ and β_σ are increasing for any $\sigma \in \Sigma$

(iii) $f(x, a, \varepsilon')$ is increasing in (x, ε'), and $f(x, \sigma(x), \varepsilon')$ is increasing in (x, ε') for any $\sigma \in \Sigma$,

(iv) $\Gamma(x_1) \subset \Gamma(x_2)$ if $x_1 \leqslant x_2$,

(v) $\varepsilon_{1t}, \dots, \varepsilon_{mt}$ are independent for each t.

Let $g(x, a) \coloneqq -\beta(x, a)/\theta \log \mathbb{E}_{x,a} \exp(-\theta v(x'))$ for all $(x, a) \in \mathsf{G}$. Let $\mathcal{G}$ be the continuous functions g on G such that $\|g\| < \infty$ and g is increasing in X. Similar to (15), the transformed Bellman equation is

$$g(x, a) = -\frac{\beta(x, a)}{\theta} \ln \mathbb{E}_{x,a} \exp\left(-\theta \sup_{a' \in \Gamma(x')} \{r(x', a') + g(x', a')\} \right). \tag{21}$$

for $(x, a) \in \mathsf{G}$ and $g \in \mathcal{G}$. Define the *risk-sensitive expected value Bellman operator* R by letting $Rg(x, a)$ be the right-hand side of (21) for any $g \in \mathcal{G}$ and $(x, a) \in \mathsf{G}$.

Assumption 7.2. $\bar{r}$ is bounded above and $\hat{r}$ is bounded below, where $\bar{r}$ is defined by (14) and $\hat{r}$ is defined by

$$\hat{r}(x, a) \coloneqq -\frac{1}{\theta} \ln \mathbb{E}_{x,a} \exp(-\theta \bar{r}(x')) \qquad ((x, a) \in \mathsf{G}).$$

Define $\bar{\beta}(x) \coloneqq \sup_{a \in \Gamma(x)} \beta(x, a)$ for all $x \in \mathsf{X}$.

Assumption 7.3. There exists an $\ell\colon \mathsf{X}^2 \to \mathbb{R}_+$ be such that $\beta(x, a)P(x, a, x') \leqslant \ell(x, x')$ for all $(x, a, x') \in \mathsf{G} \times \mathsf{X}$. Moreover, $\rho(L) < 1$ or there exists an $n \in \mathbb{N}$ such that $\sup_{x \in \mathsf{X}} L^n \bar{\beta}(x) < 1$, where $L\colon mb\mathsf{X} \to mb\mathsf{X}$ is defined by $Lh(x) \coloneqq \int \ell(x, x')h(x')\mathrm{d}x'$ for $x \in \mathsf{X}$ and $h \in mb\mathsf{X}$.

If Assumption 7.1, 7.2, and 7.3 are satisfied, the optimality results follow from the following theorem.

Theorem 7.1. *If Condition 7.1, Assumption 7.1, 7.2, and 7.3 hold, then the following statements are true.*

(a) $R\mathcal{G} \subset \mathcal{G}$ *and* R *is eventually contracting on* $(\mathcal{G}, \|\cdot\|)$,

(b) R *admits a unique fixed point in* $\mathcal{G}$,

(c) v^* *is well defined and*

$$v^*(x) = \sup_{a\in\Gamma(x)} \{r(x,a) + g^*(x,a)\} \quad g^*(x,a) = -\frac{\beta(x,a)}{\theta} \log \mathbb{E}_{x,a} e^{-\theta v^*(x)},$$

(d) at least one optimal policy exists, and

(e) a feasible policy is optimal if and only if it is g^**-greedy.*

Appendix A. Appendix

A.1. **Proofs in Section 2.**

Proof of Lemma 2.1. Let Condition 2.1 hold. Clearly, since $\sigma \in \Sigma$ is measurable, Condition 2.1 implies that T_σ is a self-map on $mb\mathsf{X}$. We next show that T is a self-map on $cb\mathsf{X}$. Fix $v \in cb\mathsf{X}$. Since $(x,a) \mapsto B(x,a,v)$ is bounded and continuous on G, and Γ is continuous and compact-valued, it follows from the (measurable) Maximum theorem that $x \mapsto \sup_{a\in\Gamma(x)} B(x,a,v) = Tv(x)$ is continuous and the correspondence $\tau\colon \mathsf{X} \to \mathsf{A}$ defined by

$$\tau(x) = \arg\max_{a\in\Gamma(x)} B(x,a,v)$$

is nonempty, u.s.c, and compact-valued and admits a Borel measurable selector σ satisfying $\sigma(x) \in \tau(x)$ for all $x \in \mathsf{X}$. Then, σ is a Borel measurable v-greedy policy. Since $Tv(x) = B(x,a,v)$ for $a \in \tau(x)$, Tv is bounded. Therefore, Tv is a self-map on $cb\mathsf{X}$. □

Lemma A.1. *If* $\mathscr{R}$ *is regular and* $\rho(L_\sigma) < 1$ *for all* $\sigma \in \Sigma$*, then* T_σ *is eventually contracting, globally stable on* $mb\mathsf{X}$*, and has a unique fixed point* $v_\sigma \in mb\mathsf{X}$ *for all* $\sigma \in \Sigma$*. Moreover, if* $\sigma \in \Sigma$ *is continuous, then* $v_\sigma \in cb\mathsf{X}$*.*

Proof of Lemma A.1. Suppose that $\mathscr{R}$ is regular and $\rho(L_\sigma) < 1$ for all $\sigma \in \Sigma$. Fix $\sigma \in \Sigma$. Let L_σ be the operator defined by (4). Then, we have

$$|T_\sigma v(x) - T_\sigma w(x)| \leqslant L_\sigma |v - w|(x) \qquad (x \in \mathsf{X} \text{ and } v, w \in mb\mathsf{X}).$$

Since L_σ is order-preserving, iteration implies that $|T_\sigma^n v - T_\sigma^n w| \leqslant L_\sigma^n |v - w|$ for all $n \in \mathbb{N}$. Since $\rho(L_\sigma) < 1$, Lemma 3.3 implies that there is an $n_\sigma \in \mathbb{N}$ such that $d_{n_\sigma}^\sigma < 1$ and iteration implies

$$L_\sigma^{n_\sigma} h(x) \leqslant d_{n_\sigma}^\sigma \|h\|. \qquad (h \in mb\mathsf{X})$$

Therefore, the eventual discounting $d_{n_\sigma}^\sigma < 1$ implies that

$$|T_\sigma^{n_\sigma} v - T_\sigma^{n_\sigma} w| \leqslant L_\sigma^{n_\sigma} |v - w| \leqslant d_{n_\sigma}^\sigma \|v - w\| < \|v - w\|.$$

Taking supremum over X on the left, we conclude that $T_\sigma^{n_\sigma}$ is a contracting map with modulus $d_{n_\sigma}^\sigma$. Finally, the generalized Contraction Mapping theorem shows that T_σ is globally stable and has a unique fixed point $v_\sigma \in mb\mathsf{X}$.[16] If $\sigma \in \Sigma$ is continuous, then T_σ is a self-map on $cb\mathsf{X}$. Then, T_σ is globally stable on both $cb\mathsf{X}$ and $mb\mathsf{X}$. Since $cb\mathsf{X} \subset mb\mathsf{X}$ and v_σ is unique in both $mb\mathsf{X}$ and $cb\mathsf{X}$, we have $v_\sigma \in cb\mathsf{X}$. □

Lemma A.2. *If $\mathscr{R}$ is regular and Assumption 2.1 holds, then for all $u, v \in cb\mathsf{X}$ there exists a $\sigma \in \Sigma_C$ such that $\|Tu - Tv\|_{w_\sigma} \leqslant \lambda_\sigma \|u - v\|_{w_\sigma}$.*

Proof of Lemma A.2. Let $\mathscr{R}$ be regular and Assumption 2.1 hold. Fix $u, v \in cb\mathsf{X}$. Since $(x, a) \mapsto \int_{\mathsf{X}} k(x, a, x')|v(x') - w(x')|\mathrm{d}x'$ is continuous and Γ is continuous and compact-valued, it follows from the Maximum theorem that there exists a $\sigma' \in \Sigma$ such that

$$\sigma'(x) \in \arg\max_{a \in \Gamma(x)} \int_{\mathsf{X}} k(x, a, x')|v(x') - w(x')|\mathrm{d}x' \qquad (x \in \mathsf{X}).$$

[16] See, e.g., Cheney et al. (2001) for generalized Contraction Mapping theorem.

Hence, since there is $w_\sigma \in mb\mathsf{X}$ and $\lambda_\sigma < 1$ such that $L_\sigma w_\sigma \leqslant \lambda_\sigma w_\sigma$ for any $\sigma \in \Sigma_C$ by Assumption 2.1, we have

$$\begin{aligned}
|Tu(x) - Tv(x)| &= |\sup_a B(x,a,u) - \sup_a B(x,a,v)| \\
&\leqslant \sup_a |B(x,a,u) - B(x,a,v)| \\
&\leqslant \sup_\sigma \int_{\mathsf{X}} k(x,a,x')|u(x') - v(x')|\mathrm{d}x' \\
&= \int_{\mathsf{X}} k(x,\sigma'(x),x')|u(x') - v(x')|\mathrm{d}x' \\
&\leqslant \int_{\mathsf{X}} k(x,\sigma'(x),x') w_{\sigma'}(x') \sup_y \frac{|u(y) - v(y)|}{w_{\sigma'}(y)} \mathrm{d}x' \\
&= \|u - v\|_{w_{\sigma'}} L_{\sigma'} w_{\sigma'}(x) \\
&\leqslant \|u - v\|_{w_{\sigma'}} \lambda_{\sigma'} w_{\sigma'}(x)
\end{aligned} \tag{22}$$

for all $x \in \mathsf{X}$. Dividing both sides by $w_{\sigma'}(x)$ and taking supremum, we have $\|Tu - Tv\|_{w_{\sigma'}} \leqslant \lambda_{\sigma'} \|u - v\|_{w_{\sigma'}}$. □

Lemma A.3. *If $\mathscr{R}$ is regular and Assumption 2.1 holds, then T is contracting on $(cb\mathsf{X}, \|\cdot\|_w)$ with modulus $\sup_{\sigma\in\Sigma} \lambda_\sigma$ and has a unique fixed point in $cb\mathsf{X}$, where $w(x) \coloneqq \inf_{\sigma\in\Sigma} w_\sigma(x) \geqslant 1$ for all $x \in \mathsf{X}$.*

Proof of Lemma A.3. Let $\mathscr{R}$ be regular and Assumption 2.1 hold. Lemma 2.1 implies that T is a self-map on $cb\mathsf{X}$. Fix $u, v \in cb\mathsf{X}$. It follows from Lemma A.2 that there exists a $\sigma' \in \Sigma$ such that for all $x \in \mathsf{X}$

$$\begin{aligned}
\sup_x \frac{|Tu(x) - Tv(x)|}{w_{\sigma'}(x)} &\leqslant \lambda_{\sigma'} \sup_y \frac{|u(y) - v(y)|}{w_{\sigma'}(y)} \\
&\leqslant \left(\sup_{\sigma\in\Sigma} \lambda_\sigma\right) \sup_y \frac{|u(y) - v(y)|}{\inf_\sigma w_\sigma(y)} = \left(\sup_{\sigma\in\Sigma} \lambda_\sigma\right) \|u - v\|_w.
\end{aligned}$$

Taking the supremum over Σ on the left yields

$$\|Tu - Tv\|_w \leqslant \left(\sup_{\sigma\in\Sigma} \lambda_\sigma\right) \|u - v\|_w.$$

Therefore, T is contracting in $\|\cdot\|_w$ norm with modulus $\lambda = \sup_{\sigma\in\Sigma}$, and then T has a unique fixed point $\bar{v} \in cb\mathsf{X}$. □

Lemma A.4. *If $\mathscr{R}$ is regular and Assumption 2.1 holds, then T_σ is contracting on $(mb\mathsf{X}, \|\cdot\|_{w_\sigma})$ with modulus λ_σ. for all $\sigma \in \Sigma$.*

Proof of Lemma A.4. The proof is similar to Lemma A.3. □

Lemma A.5. *If $\mathscr{R}$ is regular and Assumption 2.2 holds, then T is eventually contracting and globally stable on $cb\mathsf{X}$ and has a unique fixed point $\bar{v} \in cb\mathsf{X}$, and T_σ is eventually contracting and globally stable on $mb\mathsf{X}$, and have unique fixed points $v_\sigma \in mb\mathsf{X}$ for all $\sigma \in \Sigma$.*

Proof of Lemma A.5. Let $\mathscr{R}$ be regular and Assumption 2.2 hold. Lemma 2.1 implies that T is a self-map on $cb\mathsf{X}$. Fix $v, w \in cb\mathsf{X}$. Then, we have

$$\begin{aligned} |Tv(x) - Tw(x)| &= \left| \sup_\sigma T_\sigma v(x) - \sup_\sigma T_\sigma w(x) \right| \\ &= \sup_\sigma |T_\sigma v(x) - T_\sigma w(x)| \\ &\leqslant \sup_\sigma |B(x, \sigma(x), v) - B(x, \sigma(x), w)| \\ &\leqslant L|v - w|(x) \qquad (x \in \mathsf{X}). \end{aligned}$$

Hence, we obtain $|Tv - Tw| \leqslant L|v - w|$. Since L preserve orders, iteration gives

$$|T^n v - T^n w| \leqslant L^n |v - w| \leqslant L^n \mathbb{1} \|v - w\| \leqslant \|L^n \mathbb{1}\| \|v - w\|$$

for $n \in \mathbb{N}$. Taking the supremum on the left, we have $\|T^n v - T^n w\| \leqslant \|L^n \mathbb{1}\| \|v - w\|$ for all $n \in \mathbb{N}$. Since L is a positive linear operator, we have $\rho(L) = \lim_{n\to\infty} \|L^n \mathbb{1}\|^{1/n}$ by Theorem 9.1 of Krasnosel'skii et al. (2012) as the proof of Lemma 3.3. Since $\rho(L) < 1$, there exists an $m \in \mathbb{N}$ such that $\|L^m \mathbb{1}\| < 1$. We conclude that $\|T^m v - T^m w\| < \|v - w\|$ for all $v, w \in cb\mathsf{X}$. Therefore, the generalized Contracting Mapping theorem (see, e.g., Cheney et al. (2001)) shows that T is globally stable on X, which implies that T has a unique fixed point $\bar{v} \in cb\mathsf{X}$. The proof for T_σ is similar. □

Lemma A.6. *If $\mathscr{R}$ is regular, T_σ is globally stable on $mb\mathsf{X}$ for all $\sigma \in \Sigma$, and T is globally stable on $cb\mathsf{X}$, then the following statements are true.*

(a) If $\bar{v}$ is the unique fixed point of T on $cb\mathsf{X}$, then $v^* = \bar{v}$ and there exists a $\sigma^* \in \Sigma$ such that $v^* = v_{\sigma^*}$.

(b) If T has a non-continuous fixed point $\bar{v} \in mb\mathsf{X}$, then $v^* < \bar{v}$ and $v_\sigma < \bar{v}$ for all $\sigma \in \Sigma$.

Proof of Lemma A.6. Let $\mathscr{R}$ be regular and suppose that T_σ is globally stable on $mb\mathsf{X}$ for all $\sigma \in \Sigma$, and T is globally stable on $cb\mathsf{X}$. We first show (a). Fix $\sigma \in \Sigma$ and

let $\bar{v}$ be the unique fixed point of T on $cb\mathsf{X}$. Since $\bar{v}$ is the fixed point of T and $T\bar{v} = \sup_{\sigma\in\Sigma} T_\sigma \bar{v}$, we obtain $\bar{v} = T\bar{v} \geqslant T_\sigma \bar{v}$. Since T_σ is order-preserving, iteration gives $\bar{v} \geqslant T_\sigma \bar{v} \geqslant \cdots \geqslant T_\sigma^n \bar{v}$ for $n \in \mathbb{N}$. In addition, since T_σ is globally stable, we have $\bar{v} \geqslant v_\sigma$ as $n \to \infty$, which implies $\bar{v} \geqslant v^*$ by taking supremum over Σ.

Conversely, Lemma 2.1 implies that there exists a $\sigma^* \in \Sigma$ such that $T_{\sigma^*}\bar{v} = T\bar{v}$. Then, we have $T_{\sigma^*}\bar{v} = \bar{v}$, whence $\bar{v} = v_{\sigma^*}$ by the uniqueness of fixed point of T_{σ^*}. By the definition of v^*, we obtain $v^* \geqslant v_{\sigma^*} = \bar{v}$. We conclude that $v^* = \bar{v}$ and an optimal policy σ^* exists.

To show (b), assume that there is a non-continuous $\bar{v} \in mb\mathsf{X}$ such that $T\bar{v} = \bar{v}$. Suppose that there is $\sigma \in \Sigma$ such that $v_\sigma \geqslant \bar{v}$. Then, we have $v^* \geqslant v_\sigma \geqslant \bar{v}$. Moreover, since $\bar{v}$ is a fixed point of T, we have $\bar{v} = T\bar{v} \geqslant T_\sigma \bar{v}$ for any $\sigma \in \Sigma$. Iteration implies $\bar{v} \geqslant T_\sigma^n \bar{v}$, so $\bar{v} \geqslant v_\sigma$ by the global stability of T_σ. It implies that $\bar{v} \geqslant v^*$. Then, we obtain $\bar{v} = v^*$, which is continuous by (a). Therefore, we must have $v_\sigma < \bar{v}$ for any $\sigma \in \Sigma$. The same argument also implies $v^* < \bar{v}$. □

Lemma A.7. *If Condition 2.2 holds, then the following statements are true.*

(a) T and T_σ are self-maps on $cb\mathsf{X}$, and

(b) for any $v \in cb\mathsf{X}$ there exists a continuous v-greedy policy.

Proof of Lemma A.7. Let $\mathscr{R}$ be regular and Assumption 2.2 holds. Then, if $\sigma \in \Sigma$ is continuous and $v \in cb\mathsf{X}$, then $x \mapsto T_\sigma(x) = B(x, \sigma(x), v)$ is continuous. Hence, T_σ is a self-map on $cb\mathsf{X}$. Since Γ is convex-valued and compact-valued, and $a \mapsto B(x, a, v)$ is strictly quasi-concave for any $x \in \mathsf{X}$ and $v \in cb\mathsf{X}$, it follows from the Maximum theorem that the Tv is continuous and then T is a self-map on $cb\mathsf{X}$. Moreover, for any $v \in cb\mathsf{X}$, the maximizer correspondence $x \mapsto \max_{a\in\Gamma(x)} B(x, a, v)$ is single-valued and continuous, whence a continuous v-greedy policy exists. □

Lemma A.8 (Bellman's Principle of Optimality)**.** *Suppose that $\mathscr{R}$ is regular and T_σ is globally stable for all $\sigma \in \Sigma$. If v^* is a fixed point to T, then Bellman's principle of optimality holds.*

Proof of Lemma A.8. Let all the stated assumptions hold. We want to show: $\sigma \in \Sigma$ is optimal (i.e., $v_\sigma = v^*$) if and only if σ is v^*-greedy. Suppose that $\sigma \in \Sigma$ is v^*-greedy: $T_\sigma v^* = Tv^*$. Since $Tv^* = v^*$, we obtain $T_\sigma v^* = v^*$ so that $v_\sigma = v^*$. Conversely,

suppose that $\sigma \in \Sigma$ is optimal: $v_\sigma = v^*$. Since T_σ has a unique fixed point v_σ, we obtain $v^* = T_\sigma v^*$. Since $Tv^* = v^*$, we have $T_\sigma v^* = v^* = Tv^*$, so that σ is v^*-greedy. □

Lemma A.9. *If Condition 2.2 holds, T_σ is globally stable on $mb\mathsf{X}$ for all $\sigma \in \Sigma$, and T is globally stable on $cb\mathsf{X}$, then $v^* = \sup_{\sigma\in\Sigma_C} v_\sigma$ and $Tv = \sup_{\sigma\in\Sigma_C} T_\sigma v$ for all $v \in cb\mathsf{X}$.*

Proof of Lemma A.9. Let all the assumptions hold. By Lemma A.7, there is a continuous $\sigma' \in \Sigma$ such that $Tv = T_{\sigma'}v$, which implies that $Tv = \sup_{\sigma\in\Sigma} T_\sigma v = T_{\sigma'}v \leqslant \sup_{\sigma\in\Sigma_C} T_\sigma v$. Moreover, since $v^* = \bar{v} \in cb\mathsf{X}$ by Lemma A.6, we have $\sup_{\sigma\in\Sigma} v_\sigma = v^* = v_{\sigma^*} \leqslant \sup_{\sigma\in\Sigma_C} v_\sigma$, where σ^* is a continuous v^*-greedy policy. □

Lemma A.10 (HPI)**.** *Suppose that Condition 2.2 holds, T_σ is globally stable on $cb\mathsf{X}$ with fixed point $v_\sigma \in cb\mathsf{X}$ for all $\sigma \in \Sigma_C$, and T is globally stable on $cb\mathsf{X}$. If $v^* \in cb\mathsf{X}$ is the unique fixed point of T on $cb\mathsf{X}$, then $\{H^n v_{\sigma_0}\}_{n\geqslant 0}$ converges to v^* for any $\sigma_0 \in \Sigma_C$, and v^* is the unique fixed point of H.*

Proof of Lemma A.10. Let all the stated assumptions hold. Let $\{\sigma_k\}_{k\geqslant 0} \subset \Sigma_C$ be such that $\sigma_0 \in \Sigma_C$ and $v_{\sigma_k} = Hv_{\sigma_{k-1}}$ for all $k \in \mathbb{N}$; that is, $\sigma_k \in \Sigma_C$ satisfies $T_{\sigma_k} v_{\sigma_{k-1}} = Tv_{\sigma_{k-1}}$. Note that for any $k \geqslant 0$, the continuity of σ_k implies that T_{σ_k} is a self-map on $cb\mathsf{X}$, v_{σ_k} is continuous, and then a continuous v_{σ_k}-greedy policy σ_{k+1} exists by Lemma A.7. By definition, $T_{\sigma_k} v_{\sigma_{k-1}} = Tv_{\sigma_{k-1}} \geqslant T_{\sigma_{k-1}} v_{\sigma_{k-1}} = v_{\sigma_{k-1}}$. Applying T_{σ_k} on both sides repeatedly, since T_{σ_k} is order-preserving, we have $T^n_{\sigma_k} v_{\sigma_{k-1}} \geqslant T_{\sigma_k} v_{\sigma_{k-1}} \geqslant Tv_{\sigma_{k-1}} \geqslant v_{\sigma_{k-1}}$. Taking n to infinity, the global stability of the policy operator implies that $v_{\sigma_k} \geqslant Tv_{\sigma_{k-1}} \geqslant v_{\sigma_{k-1}}$ for all $k \geqslant 0$. Since T is order-preserving, observe that $Tv_{\sigma_k} \geqslant T^2 v_{\sigma_{k-1}}$ for all $k \in \mathbb{N}$ and then $v_{\sigma_k} \geqslant Tv_{\sigma_{k-1}} \geqslant T^2 v_{\sigma_{k-2}}$. Induction yields $v_{\sigma_k} \geqslant T^k v_{\sigma_0}$ for $k \in \mathbb{N}$. Since $v^* \geqslant v_{\sigma_k}$ by definition, we obtain $v^* \geqslant v_{\sigma_k} \geqslant T^k v_{\sigma_0}$. Taking $k \to \infty$, since T is globally stable with unique fixed point v^*, we have $v_{\sigma_k} \to v^*$.

Next, let σ^* be v^* greedy policy. Since $T_{\sigma^*} v^* = Tv^* = v^*$ and T_{σ^*} has a unique fixed point v_{σ^*}, we have $v_{\sigma^*} = v^*$ and $Hv^* = v_{\sigma^*} = v^*$, whence v^* is a fixed point of H. To see that v^* is the unique fixed point of H, let $\bar{v}$ be the fixed point of H. Then, we have $\bar{v} = H\bar{v} = v_\sigma$ where σ is $\bar{v}$-greedy. It implies that $T\bar{v} = Tv_\sigma = T_\sigma v_\sigma = v_\sigma = \bar{v}$. Since v^* is the unique fixed point of T by global stability, we have $\bar{v} = v^*$. □

Lemma A.11. *If Condition 2.2 holds, then W, defined by (7), is an order-preserving self-map on V_u, and we have*

$$v \in V_u \coloneqq \{v \in cb\mathsf{X} \colon Tv \geqslant v\} \Longrightarrow Tv \leqslant Wv \leqslant T^m v.$$

Proof of Lemma A.11. Let Condition 2.2 hold. Then, T and T_σ are order-preserving self-maps on $cb\mathsf{X}$ for all $\sigma \in \Sigma_C$, and v-greedy continuous policy exists for all $v \in cb\mathsf{X}$ by Lemma A.7. Fix any $v \in V_u$ and let $\sigma \in \Sigma_C$ be such that $T_\sigma v = Tv$. Since $T_\sigma u \leqslant Tu$ for any $u \in cb\mathsf{X}$, T and T_σ are order-preserving and $v \leqslant Tv$, we see that $Wv \in V_u$:

$$Wv = T_\sigma T_\sigma^{m-1} v \leqslant T T_\sigma^{m-1} v \leqslant T T_\sigma^{m-1} Tv = T T_\sigma^{m-1} T_\sigma v = TWv.$$

Then, W is a self-map on V_u. Also since T_σ^m is order-preserving, W is order-preserving. Next, regarding the first inequality, since T_σ is order-preserving, $v \leqslant Tv$, and $T_\sigma v = Tv$, we have

$$T_\sigma^{m-1} v \leqslant T_\sigma^{m-1} Tv = T_\sigma^{m-1} T_\sigma v = Wv.$$

Repeating the same iteration, we have $T_\sigma^{m-j} v \leqslant Wv$ for $j < m$. In particular, for $j = m-1$, we have $T_\sigma v \leqslant Wv$. Since $T_\sigma v = Tv$, we obtain $Tv \leqslant Wv$. For the second inequality, since $T_\sigma v \leqslant Tv$ by definition, and T and T_σ are order-preserving, we have $Wv = T_\sigma^m v \leqslant T^m v$. □

Lemma A.12 (OPI)**.** *If Condition 2.2 holds and T is globally stable on $cb\mathsf{X}$ with unique fixed point $v^* \in cb\mathsf{X}$, then the OPI iteration $\{v_k\}$ converges to v^* with $v_0 = v_\sigma$ for some $\sigma \in \Sigma_C$.*

Proof of Lemma A.12. Let all the stated assumptions hold. Pick $\sigma \in \Sigma_C$ and let $v_0 = v_\sigma$. Let $\{v_k\}_{k \geqslant 0}$ be the OPI iteration. First, claim that

$$T^k v_0 \leqslant W^k v_0 \leqslant T^{km} v_0 \qquad (k \in \mathbb{N}).$$

Since $v_0 = T_\sigma v_0 \leqslant Tv_0$, $v_0 \in V_u$. Hence, it follows from Lemma A.11 that the claim holds for $k = 1$. Suppose that the claim holds for some $k \in \mathbb{N}$. Since all operators are order-preserving and self-map on V_u, Lemma A.11 with $W^k v_0, T^{km} v_0 \in V_u$ implies the iteration:

$$T^{k+1} v_0 \leqslant T W^k v_0 \leqslant W W^k v_0 \leqslant W T^{km} v_0 \leqslant T^{(k+1)m} v_0.$$

Therefore, the claim holds for all $k \in \mathbb{N}$ by induction. Now, since $T^k v \to v^*$ as $k \to \infty$, the above claim implies that $W^k v_0 \to v^*$. Since OPI iteration follows $v_k = W^k v_0$, we conclude that $\{v_k\}$ converges to v^*. □

Proof of Theorem 2.1. Let $\mathscr{R}$ be regular. Suppose either Assumption 2.1 or 2.2 holds. We first prove (a), (i), and (ii). If Assumption 2.1 holds, then Lemma A.3 implies that T is contracting on $(cb\mathsf{X}, \|\cdot\|_w)$ with modulus $\sup_{\sigma\in\Sigma_C} \lambda_\sigma$, Lemma A.1 implies that T_σ is eventually contracting on $(mb\mathsf{X}, \|\cdot\|)$ for any $\sigma \in \Sigma$, and Lemma A.4 implies that T_σ is contracting on $(mb\mathsf{X}, \|\cdot\|_{w_\sigma})$ with modulus λ_σ for all $\sigma \in \Sigma$. If Assumption 2.2 holds, then Lemma A.5 shows that T and T_σ are eventually contracting on $cb\mathsf{X}$ and $mb\mathsf{X}$, respectively, for all $\sigma \in \Sigma$. Part (b), (c), and (e) follow from Lemma A.6 with Part (i) and (ii). Part (d) follows from Lemma A.8 with Part (i), (ii) and (b). Part (α) follows from Lemma A.10 with Part (i), (ii) and (b). Part (β) follows from Lemma A.12 with Part (i), (ii) and (b). □

A.2. **Proofs in Section 3.**

Proof of Lemma 3.3. Let $\mathscr{R}$ be regular. Fix $\sigma \in \Sigma$ and let the linear operator L_σ be defined by (4). It follows from Theorem 1.5.5 of Bühler and Salamon (2018) that $\rho(L_\sigma) := \lim_{n\to\infty} \|L_\sigma^n\|^{1/n}$ always exists and is bounded above by $\|L_\sigma\|$. Applying Theorem 9.1 of Krasnosel'skii et al. (2012), since (i) L_σ is a positive linear operator on V, (ii) the positive cone in V is solid and normal under the pointwise partial order,[17] and (iii) $\mathbb{1}$ lies interior to the positive cone in V, we obtain

$$\rho(L_\sigma) = \lim_{n\to\infty} \|L_\sigma^n \mathbb{1}\|^{1/n} = \lim_{n\to\infty} \left\{ \sup_x |L_\sigma^n \mathbb{1}(x)| \right\}^{1/n} = \lim_{n\to\infty} (d_n^\sigma)^{1/n}. \tag{23}$$

Hence, if $\rho(L_\sigma) < 1$, there is an $n_\sigma \in \mathbb{N}$ such that $d_{n_\sigma}^\sigma < 1$. Conversely, suppose that $d_{n_\sigma}^\sigma < 1$ for some $n_\sigma \in \mathbb{N}$. Since any $n \in \mathbb{N}$ can be written uniquely as $n = kn_\sigma + i$

[17] A cone is solid if it has an interior point; it is normal if $0 \leqslant x \leqslant y$ implies $\|x\| \leqslant M\|y\|$. The cone of non-negative functions in $mb\mathsf{X}$ or $cb\mathsf{X}$ is both solid and normal.

for some $k, i \in \mathbb{N}_0$ with $i < n_\sigma$, we have that, for sufficiently large n,

$$\begin{aligned}(d_n^\sigma)^{1/n} &= \sup_x L_\sigma^n \mathbb{1}(x) = \sup_x L_\sigma^{kn_\sigma} L_\sigma^{n-kn_\sigma} \mathbb{1}(x) \\ &= \sup_x \int \ell_\sigma^{kn_\sigma}(x, x') \int \ell_\sigma^{n-kn_\sigma}(x', x'') \mathbb{1}(x'') \mathrm{d}x'' \mathrm{d}x' \\ &\leqslant \sup_x \int \ell_\sigma^{kn_\sigma}(x, x') \left(\sup_{x'} \int \ell_\sigma^{n-kn_\sigma}(x', x'') \mathrm{d}x'' \right) \mathrm{d}x' \\ &= (d_{kn_\sigma}^\sigma d_{n-kn_\sigma}^\sigma)^{1/n} \\ &\leqslant (d_{n_\sigma}^\sigma)^{k/n} (d_i^\sigma)^{1/n} \leqslant (d_{n_\sigma}^\sigma)^{k/n} (M)^{1/n}.\end{aligned} \tag{24}$$

where the second inequality follows from the same argument for the first inequality. Since $k/n \to 1/n_\sigma$ as $n \to \infty$, the right-hand side converges to $(d_{n_\sigma}^\sigma)^{1/n_\sigma} < 1$ as $n \to \infty$. Hence, $\rho(L_\sigma) < 1$. Since σ is arbitrarily chosen, the statements hold for all $\sigma \in \Sigma$.

To show (c), suppose $\sup_{\sigma\in\Sigma} d_{n_\sigma}^\sigma < 1$ and $\sup_{\sigma\in\Sigma} n_\sigma = \infty$. Then, $\sup_{\sigma\in\Sigma} n_\sigma = \infty$ implies $\sup_{\sigma\in\Sigma} d_n^\sigma \geqslant 1$ for all $n \in \mathbb{N}$; otherwise, $\sup_\sigma n_\sigma \leqslant n$ for some n that $\sup_{\sigma\in\Sigma} d_n^\sigma < 1$. However, it contradicts with $\sup_{\sigma\in\Sigma} d_{n_\sigma}^\sigma < 1$.

To show (d), we first suppose $\sup_{\sigma\in\Sigma} d_{n_\sigma}^\sigma < 1$. Then, by (24) and letting k and M be defined as above, we have for $n \in \mathbb{N}$

$$\begin{aligned}\sup_\sigma \rho(L_\sigma) &= \sup_\sigma \lim_{n\to\infty} (d_n^\sigma)^{1/n} \leqslant \sup_\sigma \lim_{n\to\infty} (d_{n_\sigma}^\sigma)^{k/n} M^{1/n} \\ &= \sup_\sigma (d_{n_\sigma}^\sigma)^{1/n_\sigma} \leqslant (\sup_\sigma d_{n_\sigma}^\sigma)^{1/\sup_\sigma n_\sigma} < 1,\end{aligned}$$

where the last inequality follows from $\sup_\sigma n_\sigma < \infty$ by (c). Conversely, suppose that $\sup_{\sigma\in\Sigma} \rho(L_\sigma) < 1$. Then, we have $d_{n_\sigma}^\sigma < 1$ for all σ by (a). Assume that $\sup_{\sigma\in\Sigma} d_{n_\sigma}^\sigma = 1$. Then, we have $\sup_{\sigma\in\Sigma} d_n^\sigma = 1$ for all n; otherwise, we have a contradiction by letting $n_\sigma = n$. Fix $\varepsilon > 0$. Since $\rho(L_\sigma) = \lim_{n\to\infty} (d_n^\sigma)^{1/n}$, there exist an $N \in \mathbb{N}$ such that for all $n \geqslant N$ we have

$$(d_n^\sigma)^{1/n} - \varepsilon \leqslant \rho(L_\sigma) \leqslant (d_n^\sigma)^{1/n} + \varepsilon.$$

Taking supremum on the right, we have

$$(d_n^\sigma)^{1/n} - \varepsilon \leqslant \rho(L_\sigma) \leqslant (\sup_\sigma d_n^\sigma)^{1/n} + \varepsilon = 1 + \varepsilon.$$

Hence, we also have

$$1 - \varepsilon = (\sup_\sigma d_n^\sigma)^{1/n} - \varepsilon \leqslant \sup_\sigma \rho(L_\sigma) \leqslant 1 + \varepsilon.$$

Since this inequalities hold for any $\varepsilon > 0$, we obtain $\sup_\sigma \rho(L_\sigma) = 1$, a contradiction. □

Proof of Lemma 3.1. Let $\mathscr{M}$ be an MDP satisfying Condition 3.1 holds. Since $(x, a) \mapsto B(x, a, v) = r(x, a) + \int_{\mathsf{X}} v(x')\beta(x, a, y)P(x, a, y)\mathrm{d}x'$ is continuous and bounded on G for $v \in cb\mathsf{X}$ by assumption, Condition 2.1 holds. Clearly, the value aggregator B satisfies monotonicity condition (2). Therefore, $\mathscr{M}$ is regular. Lemma 2.1 concludes the remaining statements. □

Lemma A.13. *If Condition 3.2 holds, then the following statements are true.*

(a) T and T_σ are self-maps on $cb\mathsf{X}$, and

(b) for any $v \in cb\mathsf{X}$ there exists a continuous v-greedy policy.

Proof of Lemma A.13. The statements follow from that if Condition 3.2 holds, then Condition 2.2 holds. □

Proof of Lemma 3.2. Let $\mathscr{M}$ be regular and Assumption 3.2 hold. Then, we have $L_\sigma \mathbb{1} \leqslant L\mathbb{1}$ for all $\sigma \in \Sigma$. Pick $\sigma \in \Sigma$. Since L_σ and L are order-preserving, iteration implies $L_\sigma^n \mathbb{1} \leqslant L^n \mathbb{1}$ for all $n \in \mathbb{N}$. Let $\{X_t\}$ be a P_σ-Markov process with $X_0 = x$ and $\beta_t = \beta(X_t, \sigma(X_t), X_{t+1})$ for all $t \in \mathbb{N}_0$. Then, iteration implies

$$L_\sigma^n \mathbb{1}(x) = \mathbb{E}_x^\sigma\{\beta_0\beta_1 \cdots \beta_{n-1}\} \leqslant L^n\mathbb{1}(x) \leqslant \|L^n\mathbb{1}\|.$$

Finally, taking supremum over X, we have $d_n^\sigma = \|L_\sigma^n\mathbb{1}\| \leqslant \|L^n\mathbb{1}\|$ for all $n \in \mathbb{N}$. Since $\lim_{n\to\infty} \|L^n\mathbb{1}\|^{1/n} = \rho(L) < 1$, there exists a $n \in \mathbb{N}$ satisfying $\|L^n\mathbb{1}\| < 1$ and then $d_n^\sigma < 1$ for all $\sigma \in \Sigma$. Taking $n \to \infty$ and supremum over Σ yields

$$\sup_\sigma \rho(L_\sigma) = \sup_\sigma \lim_{n\to\infty} \|L_\sigma^n\mathbb{1}\|^{1/n} \leqslant \lim_{n\to\infty} \|L^n\mathbb{1}\|^{1/n} = \rho(L) < 1.$$

□

Lemma A.14. *If $\mathscr{M}$ is regular and Assumption 3.1 holds, then T_σ is eventually contracting and globally stable on $mb\mathsf{X}$, and v_σ defined by (9) is the unique fixed point of T_σ in $mb\mathsf{X}$ for any $\sigma \in \Sigma$.*

Proof of Lemma A.14. Suppose that $\mathscr{M}$ is regular and Assumption 3.1 holds. Fix $\sigma \in \Sigma$. It follows from Lemma 3.3 that the assumption implies $d^n_{n_\alpha} < 1$ for some $n_\sigma \in \mathbb{N}$. Let L_σ be the operator defined by (10). Then, we have

$$\begin{aligned}
|T_\sigma v(x) - T_\sigma w(x)| &= \left| \int_{\mathsf{X}} (v(x') - w(x'))\beta_\sigma(x, x')P_\sigma(x, x')\mathrm{d}x' \right| \\
&\leqslant \int_{\mathsf{X}} |v(x') - w(x')|\beta_\sigma(x, x')P_\sigma(x, x')\mathrm{d}x' \\
&= L_\sigma|v - w|(x) \qquad (x \in \mathsf{X}, v, w \in mb\mathsf{X}).
\end{aligned}$$

Since L_σ is order-preserving, iteration implies $|T^n_\sigma v - T^n_\sigma w| \leqslant L^n_\sigma|v - w|$ for all $v, w \in mb\mathsf{X}$ and $n \in \mathbb{N}$. Let $\{X_t\}$ be a stochastic process generated by P_σ with $X_0 = x$. Let $\beta_t = \beta(X_t, \sigma(X_t), X_{t+1})$ for all $t \in \mathbb{N}_0$. The definition of L_σ yields $L_\sigma h(x) = \mathbb{E}_x \beta_0 h(X_1)$ and, by iteration,

$$L^{n_\sigma}_\sigma h(x) = \mathbb{E}^\sigma_x\{\beta_0\beta_1 \cdots \beta_{n_\sigma - 1}h(X_{n_\sigma})\} \leqslant d^\sigma_{n_\sigma}\|h\|. \qquad (h \in mb\mathsf{X})$$

Therefore, the condition of eventual discounting $d^\sigma_{n_\sigma} < 1$ implies that

$$|T^{n_\sigma}_\sigma v - T^{n_\sigma}_\sigma w| \leqslant L^{n_\sigma}_\sigma|v - w| \leqslant d^\sigma_{n_\sigma}\|v - w\| < \|v - w\|.$$

Taking supremum on the left, we conclude that $T^{n_\sigma}_\sigma$ is a contracting map with modulus $d^\sigma_{n_\sigma}$. Finally, the generalized Contraction Mapping theorem shows that T_σ is globally stable and has a unique fixed point v_s.

Next, we check that v_σ, defined by (9), is a fixed point of T_σ by definition. Since $v_s = T_\sigma v_s = T^n_\sigma v_s$, we have

$$\begin{aligned}
v_s(x) &= r_\sigma(x) + \mathbb{E}^\sigma_x \beta_0 v_s(X_1) \\
&= r_\sigma(x) + \mathbb{E}^\sigma_x \beta_0 [r_\sigma(X_1) + \mathbb{E}^\sigma_{X_1} \beta_1 v_s(X_2)] \\
&= \cdots \\
&= \mathbb{E}^\sigma_x \left[\sum_{t=0}^{n} \prod_{i=0}^{t-1} \beta_i r_\sigma(X_t) \right] + \mathbb{E}^\sigma_x \beta_0\beta_1 \cdots \beta_n v_s(X_{n+1}) \\
&= \lim_{n\to\infty} \mathbb{E}^\sigma_x \left[\sum_{t=0}^{n} \prod_{i=0}^{t-1} \beta_i r_\sigma(X_t) \right] = v_\sigma(x) \qquad (x \in \mathsf{X}).
\end{aligned}$$

where we use

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}_x^\sigma\beta_0\beta_1\cdots\beta_n v_s(X_{n+1}) &\leqslant \lim_{n\to\infty}\left(\sup_x \mathbb{E}_x^\sigma\beta_0\beta_1\cdots\beta_n\|v_s\|\right)\\ &\leqslant \lim_{\substack{n\to\infty,\\ n=kn_\sigma+i,\\ i<n_\sigma, i,k\in\mathbb{N}}}\left(\sup_x\mathbb{E}_x^\sigma\prod_{t=0}^{kn_\sigma-1}\beta_t\mathbb{E}^\sigma_{X_{kn_\sigma}}\prod_{j=kn_\sigma}^{kn_\sigma+i-1}\beta_j\|v_s\|\right)\\ &\leqslant \left(\lim_{k\to\infty}(d^\sigma_{n_\sigma})^k\right)\sup_{i<n_\sigma} d_i^\sigma\|v_s\| = 0.\end{aligned}$$

Therefore, $v_s = v_\sigma$. □

Lemma A.15. *Suppose that $\mathscr{M}$ is regular and $\bar{v}\in cb\mathsf{X}$ is a fixed point of T. If Assumption 3.1 holds, then the following statements are true.*

(a) $\|\bar{v}\|\leqslant M_\beta\|r\|/(1-\gamma)$,
(b) $\|v_\sigma\|\leqslant M_\beta\|r\|/(1-\gamma)$ *for all* $\sigma\in\Sigma$, *and*
(c) $\|v\|\leqslant M_\beta\|r\|/(1-\gamma)$ *for* $v\in cb\mathsf{X}$ *implies* $\|Tv\|, \|T_\sigma v\|\leqslant M_\beta\|r\|/(1-\gamma)$ *for all* $\sigma\in\Sigma$,

where $M_\beta := \sup_{\sigma\in\Sigma}\left[\sum_{t=0}^{n_\sigma-1}\|\beta\|^t\right]$ *and* $\gamma := \sup_{\sigma\in\Sigma} d^\sigma_{n_\sigma}$.

Proof of Lemma A.15. Let all the stated assumptions hold. Since $\bar{v}$ is the fixed point of T and Lemma 3.1 implies that there exists a $\sigma\in\Sigma$ such that $T_\sigma\bar{v} = T\bar{v}$, we obtain $\bar{v} = T_\sigma\bar{v}$. Let $\{X_t\}$ be a stochastic process generated by P_σ conditioning $X_0 = x$. Let $\beta_t = \beta(X_t,\sigma(X_t),X_{t+1})$ for all $t\in\mathbb{N}_0$. The iteration of $\bar{v} = T_\sigma\bar{v} = T_\sigma^{n_\sigma}\bar{v}$ gives

$$\begin{aligned}|\bar{v}(x)| = |T_\sigma\bar{v}| &= \left|r_\sigma(x) + \mathbb{E}_x^\sigma\beta_0\bar{v}(X_1)\right|\\ &= \left|\mathbb{E}_x^\sigma\left[\sum_{t=0}^{n_\sigma-1}\prod_{i=0}^{t-1}\beta_i r_\sigma(X_t)\right] + \mathbb{E}_x^\sigma\beta_0\beta_1\cdots\beta_{n_\sigma-1}\bar{v}(X_{n_\sigma})\right|\\ &\leqslant \sup_{\sigma\in\Sigma}\left\{\|r\|\mathbb{E}_x^\sigma\left[\sum_{t=0}^{n_\sigma-1}\prod_{i=0}^{t-1}\beta_i\right] + \|\bar{v}\|\mathbb{E}_x^\sigma\beta_0\beta_1\cdots\beta_{n_\sigma-1}\right\}\\ &\leqslant \|r\|\sup_{\sigma\in\Sigma}\sup_x\mathbb{E}_x^\sigma\left[\sum_{t=0}^{n_\sigma-1}\prod_{i=0}^{t-1}\beta_i\right] + \|\bar{v}\|\sup_{\sigma\in\Sigma}d^\sigma_{n_\sigma}\\ &\leqslant \|r\|\sup_{\sigma\in\Sigma}\left[\sum_{t=0}^{n_\sigma-1}\|\beta\|^t\right] + \|\bar{v}\|\sup_{\sigma\in\Sigma}d^\sigma_{n_\sigma}\qquad(x\in\mathsf{X})\end{aligned}$$

Hence, taking sup over X yields $\|\bar{v}\| \leqslant M_\beta \|r\| + \gamma \|\bar{v}\|$. Since $\gamma < 1$, we have the bound $\|\bar{v}\| \leqslant M_\beta \|r\|/(1-\gamma)$. Next, since v_σ is the fixed point of T_σ by Lemma A.14, a similar iteration generates the same bound.

To prove part (c), suppose that $v \in V$ with $\|v\| \leqslant M_\beta \|r\|/(1-\gamma)$. Then, it follows from Lemma 3.1 that $Tv = T_\sigma v$ for some $\sigma \in \Sigma$ and the same iteration shows $\|Tv\| \leqslant M_\beta \|r\| + \|v\|\gamma \leqslant M_\beta \|r\|/(1-\gamma)$. Clearly, for any T_σ, the same iteration gives $\|T_\sigma v\| \leqslant M_\beta \|r\|/(1-\gamma)$. □

Lemma A.16. *If $\mathscr{M}$ is regular and Assumption 3.1 holds, then Assumption 2.1 is satisfied when*

$$w_\sigma = \mathbb{E}_x^\sigma \sum_{t=0}^{\infty} \prod_{i=0}^{t-1} \beta_i^\sigma \quad \text{and} \quad \lambda_\sigma = \sup_{x \in \mathsf{X}} \frac{w_\sigma(x) - 1}{w_\sigma(x)}.$$

Proof of Lemma A.16. Let $\mathscr{M}$ be regular and Assumption 3.1 hold. Then, Assumption 2.1 (a) holds by regular condition, and Assumption 2.1 (b) follows from Assumption 3.1 and Lemma 3.3. It follows from Lemma A.14 that T_σ is globally stable and has a unique fixed point v_σ, for all $\sigma \in \Sigma$. Fix $\sigma \in \Sigma_C$. Let the reward be $r \equiv \mathbb{1}$ and $w_\sigma(x) = \mathbb{E}_x^\sigma \sum_{t=0}^{\infty} \prod_{i=0}^{t-1} \beta_i^\sigma$, where $\beta_t^\sigma = \beta_\sigma(X_t, X_{t+1})$ given a stochastic process $\{X_t\}$ generated by P_σ with $X_0 = x$. Since β is bounded and strictly positive, and $\prod_{i=1}^{-1} \beta_i^\sigma = 1$, w_σ is bounded and $w_\sigma \gg \mathbb{1}$. Clearly, w_σ is the fixed point of T_σ when $r = \mathbb{1}$, whence Lemma A.15 implies that w_σ is bounded. Therefore, we have

$$w_\sigma(x) = 1 + \int w_\sigma(x') \beta_\sigma(x, x') P_\sigma(x, x') \mathrm{d}x' \qquad (x \in \mathsf{X}).$$

Rewriting this equation, we have for all $x \in \mathsf{X}$

$$\begin{aligned} L_\sigma w_\sigma(x) &= \int w_\sigma(x') \beta_\sigma(x, x') P_\sigma(x, x') \mathrm{d}x' \\ &= w_\sigma(x) - 1 \leqslant \left(\sup_y \frac{w_\sigma(y) - 1}{w_\sigma(y)} \right) w_\sigma(x) =: \lambda_\sigma w_\sigma(x). \end{aligned}$$

Since w_σ is bounded and $w_\sigma \gg \mathbb{1}$, we obtain $\lambda_\sigma < 1$. We next show that $\sup_{\sigma \in \Sigma} \lambda_\sigma < 1$. Now, Lemma 3.3 implies $\sup_\sigma n_\sigma < \infty$. Since in addition $\sup_\sigma d_{n_\sigma}^{b,\sigma} < 1$, Lemma A.15 implies that $\sup_\sigma \sup_x w_\sigma(x)$ is bounded:

$$\sup_\sigma \sup_x w_\sigma(x) = \sup_\sigma \sup_x \mathbb{E}_x^\sigma \sum_{t=0}^{\infty} \prod_{i=0}^{t-1} \beta_i \leqslant \frac{\sup_\sigma \sum_{t=0}^{n_\sigma - 1} \|\beta\|^t}{1 - \sup_\sigma d_{n_\sigma}^\sigma} < \infty$$

Hence, we observe that

$$\lambda = \sup_\sigma \lambda_\sigma = \sup_\sigma \sup_x \left\{ 1 - \frac{1}{w_\sigma(x)} \right\} = 1 - \frac{1}{\sup_\sigma \sup_x w_\sigma(x)} < 1.$$

Therefore, Assumption 2.1 (c) is satisfied. □

Proof of Theorem 3.1. Let $\mathscr{M}$ be regular. Lemma 3.1 implies that $\mathscr{M}$ is a regular RDP. Moreover, if Condition 3.2 holds, then Condition 2.2 holds. Since Lemma A.16 implies that Assumption 2.1 holds, the conclusions follow from Theorem 2.1. □

A.3. **Proofs in Section 4.** Let $L_1(\pi_\sigma)$ be the set of Borel measurable functions $g\colon \mathsf{X} \to \mathbb{R}$ such that

$$\|g\| \coloneqq \int |g(x)| \pi_\sigma(\mathrm{d}x) < \infty.$$

For $f, g \in L_1(\pi_\sigma)$, we write $f \geqslant g$ if $f(x) \geqslant g(x)$ for π_σ-almost all $x \in \mathsf{X}$. We write $f \gg g$ if $f(x) > g(x)$ for π_σ-almost all $x \in \mathsf{X}$. Define $\mathscr{G}_\sigma$ to be all $f \in L_1(\pi_\sigma)$ such that $f \gg 0$.

Proof of Lemma 4.1. The proof is identical to Lemma A.13 with the fact that r is positive everywhere. □

Lemma A.17. *If Assumption 4.1 and Condition 4.1 hold, then for all $\sigma \in \Sigma_C$ the following statements are true.*

(a) k_σ is continuous and bounded.

(b) There exists an $m \in \mathbb{N}$ such that k_σ^m is everywhere positive.

(c) There exists a unique stationary density π_σ for k_σ on X.

(d) π_σ is everywhere positive and continuous on X

Proof of Lemma A.17. Let the stated assumptions hold. Fix $\sigma \in \Sigma_C$. Since β, P and σ are continuous, β_σ and P_σ are continuous on a compact set X so that they are bounded. Then, k_σ is continuous and bounded, which shows (a). Let $\{X_t^\sigma\}_{t \geqslant 0}$ be the state process such that

$$\mathbb{P}\{X_{t+1}^\sigma \in B | X_t^\sigma = x\} = \int_B k_\sigma(x, y) \mathrm{d}y$$

for every $x \in \mathsf{X}$ and Borel set $B \subset \mathsf{X}$. Since Assumption 4.1 holds and β is strictly positive, the Markov chain induced by P_σ is irreducible, which implies that the Markov

chain induced by k_σ is also irreducible. Hence, there exists an $m \in \mathbb{N}$ such that k_σ^m is everywhere positive, which shows (b), and then $\{X_t^\sigma\}$ is irreducible. It then implies that (c): there exists a unique stationary distribution π_σ: $X_t^\sigma \overset{d}{=} \pi_\sigma$. Finally it follows from the proof of Lemma C1 of Borovička and Stachurski (2020) that (d): π_σ is everywhere positive and continuous. □

Lemma A.18. *If Assumption 4.1 and Condition 4.1 hold, then the following statements are true for all $\sigma \in \Sigma_C$:*

(a) L_σ is a bounded linear operator on $L_1(\pi_\sigma)$.
(b) $L_\sigma g$ is continuous for $g \in \mathscr{G}_\sigma$.
(c) $L_\sigma g \geqslant 0$ when $g \geqslant 0$ and $L_\sigma g \in \mathscr{G}_\sigma$ when $g \in \mathscr{G}_\sigma$.
(d) L_σ is irreducible and L_σ^2 is compact.
(e) $\rho(L_\sigma) > 0$ and there exists a continuous function $e_\sigma \in \mathscr{G}_\sigma$ such that $L_\sigma e_\sigma = \rho(L_\sigma) e_\sigma$.
(f) L_σ is order preserving on $L_1(\pi_\sigma)$.

Proof of Lemma A.18. Let Assumption 4.1 and Condition 4.1 hold. Fix $\sigma \in \Sigma_C$. For (a) and (b), since k_σ is continuous and bounded by Lemma A.17, and $\bar{\beta}_\sigma$ is bounded and everywhere positive, the result follows from the proof in Lemma C2 of Borovička and Stachurski (2020). For (c), the first claim is obvious, and the second follows from that β is everywhere positive. For (d), the proof is identical to Lemma C3 of Borovička and Stachurski (2020) with continuity of π_σ and k_σ from Lemma A.17. Part (e) and (f) are identical to Lemma 8.4 of Stachurski et al. (2022a). □

Lemma A.19. *If Assumption 4.1 and Condition 4.1 hold, then $\|T_\sigma v - T_\sigma w\|_{e_\sigma} \leqslant \rho(L_\sigma) \|v - w\|_{e_\sigma}$ for all $v, w \in cb\mathsf{X}_+$ and all $\sigma \in \Sigma_C$.*

Proof of Lemma A.19. Let Assumption 4.1 and Condition 4.1 hold. Fix $\sigma \in \Sigma_C$. Lemma A.18 implies that there is $e_\sigma \in cb\mathsf{X}_+$ such that $L_\sigma e_\sigma = \rho(L_\sigma) e_\sigma$. Then, for

$v, w \in cb\mathsf{X}_+$ we have

$$\begin{aligned}
|T_\sigma v(x) - T_\sigma w(x)| &= \left|\bar{\beta}_\sigma(x) \int k_\sigma(x,x')(v(x') - w(x'))\mathrm{d}x'\right| \\
&\leqslant \bar{\beta}_\sigma(x) \int k_\sigma(x,x')|v(x') - w(x')|\mathrm{d}x' \\
&\leqslant \bar{\beta}_\sigma(x) \int k_\sigma(x,x')e_\sigma(x') \sup_{x'} \frac{|v(x') - w(x')|}{e_\sigma(x')}\mathrm{d}x' \qquad (25) \\
&= \bar{\beta}_\sigma(x) \int k_\sigma(x,x')e_\sigma(x')\|v - w\|_{e_\sigma}\mathrm{d}x' \\
&= \|v - w\|_{e_\sigma} L_\sigma e_\sigma(x) = \|v - w\|_{e_\sigma} \rho(L_\sigma)e_\sigma(x).
\end{aligned}$$

Dividing both sides with $e_\sigma(x)$ and taking the supremum, we have $\|T_\sigma v - T_\sigma w\|_{e_\sigma} \leqslant \rho(L_\sigma)\|v - w\|_{e_\sigma}$ for all $v, w \in cb\mathsf{X}_+$. □

Proof of Proposition 4.1. Let Assumption 4.1 and Condition 4.1 hold. Part (a) and (b) are equivalent by Lemma 3.3. Fix $\sigma \in \Sigma_C$. Then, we have $r_\sigma \in cb\mathsf{X}_+$ and $T_\sigma v = r_\sigma + L_\sigma v$ for $v \in cb\mathsf{X}_+$. Since L_σ is irreducible and L_σ^2 is compact by Lemma A.18, it follows from Theorem 3.1 of Stachurski et al. (2022b) that part (a) and (c) are equivalent, and T_σ has no fixed point in $cb\mathsf{X}_+$ if $\rho(L_\sigma) \geqslant 1$. Then, part (e) and (f) implies part (a), and clearly part (c) implies (e) and (f). Suppose that $\rho(L_\sigma) < 1$. Lemma A.19 shows that $\|T_\sigma v - T_\sigma w\|_{e_\sigma} \leqslant \rho(L_\sigma)\|v - w\|_{e_\sigma}$ for all $v, w \in cb\mathsf{X}_+$, whence a contraction map so that (a) implies (d). Finally, if T_σ is a contraction map in $\|\cdot\|_{e_\sigma}$, then a similar proof in Lemma A.4 yields that T_σ is globally stable. □

Lemma A.20. *Suppose that Condition 4.1 and Assumption 4.1 hold. If Assumption 3.1 holds, then Assumption 2.1 holds by restricting Assumption 3.1 (iii) to continuous policies: $L_\sigma e_\sigma = \rho(L_\sigma)e_\sigma$ for any $\sigma \in \Sigma_C$ and $\sup_{\sigma\in\Sigma} \rho(L_\sigma) < 1$.*

Proof of Lemma A.20. Suppose that all the stated assumptions hold. It suffices to check Assumption 2.1 (c). By assumption 3.1 and Lemma 3.3, we have $\sup_{\sigma\in\Sigma} \rho(L_\sigma) < 1$. Let e_σ be the eigenvector corresponding to $\rho(L_\sigma)$ for all $\sigma \in \Sigma_C$. Then, the conclusion follows directly from Lemma A.18 by letting $\lambda_\sigma = \rho(L_\sigma)$ and $w_\sigma = e_\sigma$ for all $\sigma \in \Sigma_C$ for Assumption 2.1. □

Proof of Theorem 4.1. Let all the stated assumptions hold. By Lemma A.9, note that the results in Theorem 2.1 hold even if we restrict the Assumption 2.1 (iii) to continuous policies: for any $\sigma \in \Sigma_C$, there exists a $w_\sigma \in cb\mathsf{X}$ and $\lambda \geqslant 0$ such

that $w_\sigma \geqslant \mathbb{1}$ and $L_\sigma w_\sigma \leqslant \lambda_\sigma w_\sigma$, and $\sup_{\Sigma_C} \rho(L_\sigma) < 1$. Lemma A.20 implies that Assumption 2.1 holds with $\sup_{\Sigma_C} \rho(L_\sigma) < 1$. The statements follow from Theorem 2.1, Lemma 4.1 and letting $w_\sigma = e_\sigma$ and $\lambda_\sigma = \rho(L_\sigma)$ for all $x \in \Sigma_C$. Moreover, since we can restrict to constinuous policies in Lemma A.3, the Bellman operator T has a contraction modulus $\sup_{\sigma \in \Sigma_C} \rho(L_\sigma)$ on $\|\cdot\|_e$. □

A.4. **Proofs in Section 6.**

Condition A.1.

(i) Γ is nonempty, compact-valued.

(ii) $a \mapsto r(x,a)$ is u.s.c. for all $x \in \mathsf{X}$.

(iii) β is bounded and strictly positive.

(iv) $a \mapsto \int_{\mathsf{X}} f(y)\beta(x,a,y)P(x,a,y)\mathrm{d}y$ is continuous on $\Gamma(x)$ for all $x \in \mathsf{X}$ and for $f \in mb\mathsf{X}$.

If Condition A.1 is satisfied, define spaces $\mathcal{V}$ and $\mathcal{G}$ by

$$\mathcal{V} \coloneqq \{v \colon \mathsf{X} \to \mathbb{R} \cup \{-\infty\} \colon v \in m\mathsf{X},\ v/\kappa \text{ is bounded above,} \\ \text{and } (x,a) \mapsto \mathbb{E}_{x,a} v(x') \text{ is bounded below}\},$$

$$\mathcal{G} \coloneqq \{g \colon \mathsf{G} \to \mathbb{R} \colon g \in m\mathsf{G}, \|g\|_\kappa < \infty, \\ \text{and } a \mapsto g(x,a) \text{ is u.s.c. on } \Gamma(x) \text{ for all } x \in \mathsf{X}\}$$

where $\kappa \geqslant \mathbb{1}$ is a real-valued function on X, which is further defined in Assumption 6.1. We say that an MDP is *regular* if either one of Condition A.1 or 6.1 holds.

Lemma A.21. *If $\mathscr{M}$ is regular and Assumption 6.1 hold, then $(\mathcal{G}, \|\cdot\|_\kappa)$ and $(\mathcal{V}, \|\cdot\|)$ are Banach spaces.*

Proof of Lemma A.21. The case of Condition 6.1 follows from Ma et al. (2022). Suppose that Condition A.1 and Assumption 6.1 hold. Let $B_\kappa(G)$ be the space of Borel measurable real-valued functions f on G satisfying $\|f\|_\kappa < \infty$. Since $(B_\kappa(G), \|\cdot\|_\kappa)$ is a Banach space, it suffices to show that $\mathcal{G}$ is closed in $B_\kappa(G)$. Let $\{g_n\} \subset \mathcal{G}$ such that $\|g_n - g\|_\kappa \to 0$. Clearly, $\|g\|_\kappa$ is finite. We next show that $a \mapsto g(x,a)$ is u.s.c. for all $x \in \mathsf{X}$. Fix $x' \in \mathsf{X}$. For all $a_0 \in \Gamma(x')$ and $y > g(x', a_0)$, let $\varepsilon = y - g(x', a_0)$. Since $\|g_n - g\|_\kappa \to 0$, for all $\delta > 0$ there exist $N \in \mathbb{N}$ such that for all x and $a \in \Gamma(x)$ we have

$$|g_N(x,a) - g(x,a)| < \kappa(x)\delta.$$

We choose δ such that $\kappa(x')\delta < \varepsilon/3$. Since $g_N(x,\cdot)$ is u.s.c. on $G(x)$ for all $x \in \mathsf{X}$, there exists a neighborhood U of a_0 such that for all $a \in U$

$$g_N(x',a) < g_N(x',a_0) + \varepsilon/3.$$

Hence, the previous inequalities imply

$$\begin{aligned} g(x',a) &< g_N(x',a) + \kappa(x')\delta < g_N(x',a_0) + \kappa(x')\delta + \varepsilon/3 \\ &< g(x',a_0) + 2\kappa(x')\delta + \varepsilon/3 < g(x',a_0) + \varepsilon < y. \end{aligned}$$

for all $a \in U$. Therefore, $a \to g(x',a)$ is u.s.c.. Since x' is arbitrarily picked, $a \mapsto g(x,a)$ is u.s.c. on $\Gamma(x)$ for all $x \in \mathsf{X}$. We conclude that $\mathcal{G}$ is closed in $B_\kappa(G)$ and then $(\mathcal{G}, \|\cdot\|_\kappa)$ is complete. Similarly, to show that $\mathcal{V}$ is closed in $mb\mathsf{X}$, suppose that $\{v_n\} \subset \mathcal{V}$ such that $\|v_n - v\| \to 0$. Let $\varepsilon > 0$. Since $\|v_n - v\|_\kappa \to 0$, there is an $N \in \mathbb{N}$ such that $|v_N(x) - v(x)| < \varepsilon$ for all $x \in \mathsf{X}$. Since in addition v_N/κ is bounded above by some $a_N \in \mathbb{R}$, we have $v(x) \leqslant v_N(x) + \varepsilon \leqslant a_N\kappa(x) + \varepsilon$ for all $x \in \mathsf{X}$, so v/κ is bounded above following from $\kappa \geqslant \mathbb{1}$. Finally, by the definition of $\mathcal{V}$, since $\mathbb{E}_{x,a}v_N(x')$ is bounded below by some $b_N \in \mathbb{R}$, we have

$$\mathbb{E}_{x,a}v(x') \geqslant \mathbb{E}_{x,a}[v_N(x') - \varepsilon] \geqslant b_N - \varepsilon.$$

Therefore, $\mathbb{E}_{x,a}v(x')$ is also bounded below, which implies $v \in \mathcal{V}$, and then $\mathcal{V}$ is closed in $mb\mathsf{X}$. □

Lemma A.22 (Well-defined Value)**.** *If Assumption 6.1 hold, then $v_\sigma(x)$ and $v^*(x)$ are well-defined in $\mathbb{R} \cup \{-\infty\}$ for all $x \in \mathsf{X}$ and $\sigma \in \Sigma$.*

Proof of Lemma A.22. Suppose that Assumption 6.1 holds. Fixing x and $\sigma \in \Sigma$. Since $\bar{r}(x) \leqslant d\kappa(x)$ and $\mathbb{E}_{x,a}\beta(x,a,x')\kappa(x')L^n\mathbb{1}(x') \leqslant \alpha\kappa(x)L^{n+1}\mathbb{1}(x)$ for all $x \in \mathsf{X}$ and $n \geqslant 0$, iteration implies that for all $t \in \mathbb{N}$ we have

$$\begin{aligned} \mathbb{E}_x \prod_{i=0}^{t-1} \beta(x_i,\sigma(x_i),x_{i+1}) r(x_t,\sigma(x_t)) &\leqslant \mathbb{E}_x \prod_{i=0}^{t-1} \beta(x_i,\sigma(x_i),x_{i+1})\bar{r}(x_t) \\ &\leqslant \mathbb{E}_x \prod_{i=0}^{t-1} \beta(x_i,\sigma(x_i),x_{i+1}) d\kappa(x_t) \\ &\leqslant d\mathbb{E}_x \prod_{i=0}^{t-2} \beta(x_i,\sigma(x_i),x_{i+1})\alpha\kappa(x_{t-1})L\mathbb{1}(x_{t-1}) \\ &\leqslant d\alpha^t\kappa(x)(L^t\mathbb{1})(x). \end{aligned}$$

Since there is an $n \in \mathbb{N}$ such that $\alpha^n \|L^n\| < 1$ by assumption, we have

$$\begin{aligned} v_\sigma(x) &= \mathbb{E}_x \sum_{t=0}^{\infty} \prod_{i=0}^{t-1} \beta(x_i, \sigma(x_i), x_{i+1}) r(x_t, \sigma(x_t)) \\ &\leqslant \sum_{t=0}^{\infty} d\alpha^t \|L^t \mathbb{1}\| \kappa(x) \leqslant d\kappa(x) \frac{\sum_{t=0}^{n-1} \alpha^t \|L^t\|}{1 - \alpha^n \|L^n\|} < \infty. \end{aligned}$$

Therefore, $v_\sigma(x)$ is well-defined in $\mathbb{R} \cup \{-\infty\}$ for all $x \in \mathsf{X}$ and $\sigma \in \Sigma$. Moreover, since the upper bound holds for all $\sigma \in \Sigma$, the definition of v^* implies that $v^*(x)$ is bounded above and then well-defined for all $x \in \mathsf{X}$. □

Lemma A.23. *If $\mathscr{M}$ is regular and Assumption 6.1 hold, then $M\mathcal{G} \subset \mathcal{V}$, $E\mathcal{V} \subset \mathcal{G}$, $T\mathcal{V} \subset \mathcal{V}$ and $R\mathcal{G} \subset \mathcal{G}$.*

Proof of Lemma A.23. Let Condition A.1 and Assumption 6.1 hold (the proof for Condition 6.1 is similar.) We first show that $E\mathcal{V} \subset \mathcal{G}$. Let $v \in \mathcal{V}$. Then, v/κ is bounded above and $\mathbb{E}_{(\cdot,\cdot)} v(x')$ is bounded below. Similar to Lemma 8.3.7 of Hernández-Lerma and Lasserre (2012b), we can show that $a \mapsto \mathbb{E}_{x,a} \beta(x, a, x') v(x')$ is u.s.c. on $\Gamma(x)$ for all $x \in \mathsf{X}$. In detail, fix $x \in \mathsf{X}$ and let $\{a_n\}_{t \geqslant 0} \subset \Gamma(x)$ be such that $a_n \to a \in \Gamma(x)$. Since v/κ is bounded above, there exists an $m \in \mathbb{R}$ such that $v \leqslant m\kappa$. Hence, $v - m\kappa$ is non-positive, so there is a non-increasing sequence of bounded measurable functions $\{v_m^k\}$ such that $v_m^k \downarrow v_m$.[18] Then, Assumption 6.1 implies that, for all k,

$$\begin{aligned} \limsup_{n\to\infty} \int v_m(x') \beta(x, a_n, x') P(x, a_n, \mathrm{d}x') &\leqslant \limsup_{n\to\infty} \int v_m^k(x') \beta(x, a_n, x') P(x, a_n, \mathrm{d}x') \\ &= \int v_m^k(x') \beta(x, a, x') P(x, a, \mathrm{d}x') \end{aligned}$$

Letting $k \to \infty$, the Monotone Convergence theorem yields that

$$\limsup_{n\to\infty} \int v_m(x') \beta(x, a_n, x') P(x, a_n, \mathrm{d}x') \leqslant \int v_m(x') \beta(x, a, x') P(x, a, \mathrm{d}x').$$

To this end, since $x \in \mathsf{X}$ is arbitrary, $a \mapsto \mathbb{E}_{x,a} \beta(x, a, x') v_m(x') = E v_m(x, a)$ is u.s.c. for all $x \in \Gamma(x)$. It implies that $Ev(x, \cdot)$ is u.s.c. on $\Gamma(x)$. Furthermore, since $v/\kappa \leqslant m$, Assumption 6.1 implies that, for all $(x, a) \in \mathsf{G}$,

$$Ev(x, a) = \mathbb{E}_{x,a} \beta(x, a, x') v(x') \leqslant \mathbb{E}_{x,a} \beta(x, a, x') m\kappa(x') \leqslant m\kappa(x) \alpha L \mathbb{1}(x).$$

[18] If Condition 6.1 holds, then κ is continuous and v is u.s.c., implying $v - m\kappa$ is u.s.c.. Then, there is a $\{v_m^k\} \subset cb\mathsf{X}$ such that $v_m^k \downarrow v_m$.

Hence, Ev/κ is bounded above by $m\alpha\|L\|$. Moreover, to see that Ev is bounded below, let $B \coloneqq \{x \in \mathsf{X} \colon v(x) < 0\}$. Since v/κ is bounded above, we have

$$\int_{\mathsf{X}\setminus B} v(x')P(x,a,x')\mathrm{d}x' \leqslant \int_{\mathsf{X}\setminus B} m\kappa(x')P(x,a,x')\mathrm{d}x' \leqslant m\|\kappa\|.$$

for all $(x,a) \in \mathsf{G}$. Since $\mathbb{E}_{(\cdot,\cdot)}v(x')$ is bounded below, there is $e \in \mathbb{R}$ such that

$$\int_{B} v(x')P(x,a,x')\mathrm{d}x' + \int_{\mathsf{X}\setminus B} v(x')P(x,a,x')\mathrm{d}x' \geqslant e$$

for all $(x,a) \in \mathsf{G}$. The above two inequalities imply

$$\int_{B} v(x')P(x,a,x')\mathrm{d}x' \geqslant e - \int_{\mathsf{X}\setminus B} v(x')P(x,a,x')\mathrm{d}x' \geqslant e - m\|\kappa\|,$$

for all $(x,a) \in \mathsf{G}$. Since β is bounded, we see that $Ev(\cdot,\cdot) = \mathbb{E}_{(\cdot,\cdot)}\beta(\cdot,\cdot,x')v(x')$ is also bounded below: for all $(x,a), \in \mathsf{G}$

$$\begin{aligned}\mathbb{E}_{x,a}\beta(x,a,x')v(x') &= \int_{B} \beta(x,a,x')v(x')P(x,a,\mathrm{d}x') + \int_{\mathsf{X}\setminus B} \beta(x,a,x')v(x')P(x,a,\mathrm{d}x') \\ &\geqslant \|\beta\|(e - m\|\kappa\|).\end{aligned}$$

Since Ev is bounded below and Ev/κ is bounded above, we conclude that $\|Ev\|_\kappa < \infty$, which implies $Ev \in \mathcal{G}$ and then $E\mathcal{V} \subset \mathcal{G}$.

Next, we show that $M\mathcal{G} \subset \mathcal{V}$. Let $g \in \mathcal{G}$. Since $r(x,\cdot)$ and $g(x,\cdot)$ are u.s.c. on $\Gamma(x)$ for all $x \in \mathsf{X}$, Proposition D.5 of Hernández-Lerma and Lasserre (2012a) implies that $x \mapsto Mg(x) = \sup_{a\in\Gamma(x)}\{r(x,a) + g(x,a)\}$ is measurable. Moreover, Assumption 6.1 implies that for all $x \in \mathsf{X}$

$$\begin{aligned}Mg(x) = \sup_{a\in\Gamma(x)} \{r(x,a) + g(x,a)\} &\leqslant \sup_{a\in\Gamma(x)} \{r(x,a)\} + \sup_{a\in\Gamma(x)} \{g(x,a)\} \\ &\leqslant d\kappa(x) + \|g\|_\kappa\kappa(x).\end{aligned}$$

Therefore, Mg/κ is bounded above. Also, since $\|g\|_\kappa < \infty$ and κ is bounded, we have $|g(\cdot)| \leqslant \|g\|_\kappa\kappa(\cdot) < \infty$ so that g is bounded below by some $\underline{g} \in \mathbb{R}$. Then, we have, for all $(x,a) \in \mathsf{G}$,

$$\begin{aligned}\mathbb{E}_{x,a}Mg(x') &= \mathbb{E}_{x,a} \sup_{a'\in\Gamma(x')} \{r(x',a') + g(x',a')\} \geqslant \mathbb{E}_{x,a} \sup_{a'\in\Gamma(x')} \{r(x',a') + \underline{g}\} \\ &= \mathbb{E}_{x,a}\bar{r}(x') + \underline{g} = \hat{r}(x,a) + \underline{g}.\end{aligned}$$

Since $\hat{r}$ is bounded below by assumption, $\mathbb{E}_{(\cdot,\cdot)}Mg(x')$ is bounded below. Therefore, we have $Mg \in \mathcal{V}$ and then $M\mathcal{G} \subset \mathcal{V}$. Now, since $T = ME$ and $R = EM$, we obtain that $T\mathcal{V} = ME\mathcal{V} \subset M\mathcal{G} \subset \mathcal{V}$ and $R\mathcal{G} = EM\mathcal{G} \subset E\mathcal{V} \subset \mathcal{G}$. □

Lemma A.24. *Let $\mathscr{M}$ be regular and Assumption 6.1 hold. If $\bar{g}$ is the unique fixed point of R, then $\bar{v} = M\bar{g}$ is the unique fixed point of T and $\bar{g} = E\bar{v}$. In addition, R is globally stable on $(\mathcal{G}, \|\cdot\|_w)$ if and only if T is globally stable on $(\mathcal{V}, \|\cdot\|_w)$, where w is a everywhere positive function on X.*

Proof of Lemma A.24. Let $\mathscr{M}$ be regular and Assumption 6.1 hold. Let $\bar{g}$ be unique the fixed point of R. Since $T = ME$ and $R = EM$, we have $M\bar{g} = MR\bar{g} = MEM\bar{g} = TM\bar{g}$, so that $M\bar{g}$ is a fixed point of T. Suppose that $v \neq M\bar{g}$ is a fixed point of T. Then, since v is a fixed point of T, we obtain $Ev = ETv = EMEv = REv$ implies that Ev is a fixed point of R. Since $\bar{g}$ is the unique fixed point of R, we must have $\bar{g} = Ev$ and then $M\bar{g} = MEv = Tv = v$. Therefore, $M\bar{g}$ must be the unique fixed point of T.

The above statement also shows $\bar{g} = E\bar{v}$. For the second statement, observe the iteration $T^n v = (ME)^n v = M(EM)^{n-1}Ev = MR^{n-1}Ev$ for any $v \in \mathcal{V}$. Since R is globally stable and $Ev \in \mathcal{G}$, we have $R^{n-1}Ev \to \bar{g}$ as $n \to \infty$. Hence, $T^n v \to M\bar{g}$ as $n \to \infty$ for any $v \in \mathcal{V}$. Similarly, we can show that the global stability of T implies the global stability of R. □

Lemma A.25. *If $\mathscr{M}$ is regular and Assumption 6.1 hold, then R is eventually contracting and globally stable on $(\mathcal{G}, \|\cdot\|_\kappa)$, and T is globally stable on $(\mathcal{V}, \|\cdot\|_\kappa)$.*

Proof of Lemma A.25. Suppose that $\mathscr{M}$ is regular and Assumption 6.1 hold. Lemma A.23 shows that $R\mathcal{G} \subset \mathcal{G}$ and $T\mathcal{V} \subset \mathcal{V}$. Fix $g, h \in \mathcal{G}$. Assumption 6.1 (b) implies that

for all $(x, a) \in \mathsf{G}$ we have

$$\begin{aligned}
|Rg(x,a) - Rh(x,a)| &= \left| \mathbb{E}_{x,a}\beta(x,a,x') \left[\sup_{a' \in \Gamma(x')} \{r(x',a') + g(x',a')\} \right.\right. \\
&\qquad \left.\left. - \sup_{a' \in \Gamma(x')} \{r(x',a') + h(x',a')\} \right] \right| \\
&\leqslant \mathbb{E}_{x,a} \left(\beta(x,a,x') \sup_{a' \in \Gamma(x')} |g(x',a') - h(x',a')| \right) \\
&\leqslant \mathbb{E}_{x,a}\beta(x,a,x') \|g - h\|_\kappa \kappa(x') \\
&\leqslant \|g - h\|_\kappa \kappa(x) \alpha L \mathbb{1}(x).
\end{aligned}$$

Then, iteration implies

$$\begin{aligned}
|R^2 g(x,a) - R^2 g(x,a)| &\leqslant \mathbb{E}_{x,a}\beta(x,a,x') \sup_{a' \in \Gamma(x')} |Rg(x',a') - Rg(x',a')| \\
&\leqslant \mathbb{E}_{x,a}\beta(x,a,x') \|g - h\|_\kappa \kappa(x') \alpha L \mathbb{1}(x') \\
&\leqslant \|g - h\|_\kappa \kappa(x) \alpha^2 L^2 \mathbb{1}(x).
\end{aligned}$$

Induction yields

$$|R^n g(x,a) - R^n h(x,a)| \leqslant \|g - h\|_\kappa \kappa(x) \alpha^n L^n \mathbb{1}(x)$$

for all $(x, a) \in \mathsf{G}$ and $n \in \mathbb{N}$. Dividing $\kappa(x)$ on the both sides and taking supremum, we obtain

$$\|R^n g - R^n h\|_\kappa \leqslant \|g - h\|_\kappa \alpha^n \|L^n \mathbb{1}\| \leqslant \|g - h\|_\kappa \alpha^n \|L^n\|.$$

Since there exists an $n \in \mathbb{N}$ such that $\|L^n\| < 1$ and $\alpha < 1/\|L^n\|^{1/n}$, R is eventually contracting on $(\mathcal{G}, \|\cdot\|_\kappa)$, whence it is also globally stable by the generalized Banach contraction mapping theorem. Finally, Lemma A.24 implies that T is globally stable. □

Lemma A.26. *If $\mathscr{M}$ is regular and Assumption 6.1 hold, then a v-greedy policy and a g-greedy policy exist for all $v \in \mathcal{V}$ and $g \in \mathcal{G}$.*

Proof of Lemma A.26. Let Condition A.1 and Assumption 6.1 hold. Let $v \in \mathcal{V}$. Since Condition A.1 holds and $a \mapsto \mathbb{E}_{x,a}$ is continuous for all $x \in \mathsf{X}$, it follows from the proof of Lemma A.23 that $a \mapsto r(x,a) + \mathbb{E}_{x,a}\beta(x,a,x')v(x')$ is u.s.c. for all $x \in \mathsf{X}$. Since $\Gamma(x)$ is compact for all $x \in \mathsf{X}$, the v-greedy policy exists by the Maximum theorem.

Let $g \in \mathcal{G}$. Then, $a \mapsto r(x,a) + g(x,a)$ is u.s.c. for all $x \in \mathsf{X}$, so the g-greedy policy exists by the Maximum theorem. The proof for Condition 6.1 is similar. □

Lemma A.27. *If $\mathscr{M}$ is regular and Assumption 6.1 hold, and $\bar{g}$ is a fixed point of R, then $v^* = M\bar{g}$, $\bar{g} = Ev^*$, and an optimal policy exists. Moreover, the following statements are equivalent.*

(a) a policy $\sigma \in \Sigma$ is optimal,

(b) σ is $\bar{g}$-greedy, and

(c) σ is v^*-greedy.

Proof of Lemma A.27. Let $\mathscr{M}$ be regular and Assumption 6.1 hold. Lemma A.24 implies that $\bar{v} = M\bar{g}$ is a unique fixed point of T, so $\bar{v} = T\bar{v} \geqslant T_\sigma \bar{v}$ for all $\sigma \in \Sigma$. Fix $\sigma \in \Sigma$. Let $\{x_t\}$ be a P_σ-Markov process and $\beta_t^\sigma = \beta(x_t, \sigma(x_t), x_{t+1})$ for all $t \in \mathbb{N}_0$. Then, since $\bar{v} \geqslant T_\sigma \bar{v}$ and $\bar{g} = M\bar{v}$ by Lemma A.24, iteration implies that for all $x \in \mathsf{X}$ and $\sigma \in \Sigma$ we have

$$
\begin{aligned}
&\bar{v}(x_0) \geqslant T_\sigma \bar{v}(x_0) = r(x_0, \sigma(x_0)) + \mathbb{E}_{x_0,\sigma(x_0)} \beta_0 \bar{v}(x_1) \\
&\geqslant r(x_0, \sigma(x_0)) + \mathbb{E}_{x_0,\sigma(x_0)} \beta_0^\sigma [r(x_1, \sigma(x_1)) + \mathbb{E}_{x_1,\sigma(x_1)} \beta_1^\sigma \bar{v}(x_2)] \\
&\geqslant \mathbb{E}_{x_0,\sigma(x_0)} \sum_{t=0}^{N} \prod_{i=0}^{t-1} \beta_i^\sigma r(x_t, \sigma(x_t)) + \mathbb{E}_{x_0,\sigma(x_0)} \beta_0^\sigma \beta_1^\sigma \cdots \beta_{N-1}^\sigma \mathbb{E}_{x_N,\sigma(x_N)} \beta_N^\sigma \bar{v}(x_{N+1}) \quad (26) \\
&= \mathbb{E}_{x_0,\sigma(x_0)} \sum_{t=0}^{N} \prod_{i=0}^{t-1} \beta_i^\sigma r(x_t, \sigma(x_t)) + \mathbb{E}_{x_0,\sigma(x_0)} \beta_0^\sigma \beta_1^\sigma \cdots \beta_{N-1}^\sigma \bar{g}(x_N, \sigma(x_N)).
\end{aligned}
$$

Now, Assumption 6.1 implies

$$
\begin{aligned}
|\mathbb{E}_{x_0,\sigma(x_0)} \beta_0^\sigma \beta_1^\sigma \cdots \beta_{N-1}^\sigma \bar{g}(x_N, \sigma(x_N))| &\leqslant \mathbb{E}_{x_0,\sigma(x_0)} \beta_0^\sigma \beta_1^\sigma \cdots \beta_{N-1}^\sigma |\bar{g}(x_N, \sigma(x_N))| \\
&\leqslant \mathbb{E}_{x_0,\sigma(x_0)} \beta_0^\sigma \beta_1^\sigma \cdots \mathbb{E}_{x_{N-1},\sigma(x_{N-1})} \beta_{N-1}^\sigma \|g\|_\kappa \kappa(X_N) \\
&\leqslant \|g\|_\kappa \mathbb{E}_{x_0,\sigma(x_0)} \beta_0^\sigma \beta_1^\sigma \cdots \beta_{N-2}^\sigma \kappa(x_{N-1}) \alpha L \mathbb{1}(x_{N-1}) \\
&\leqslant \|\bar{g}\|_\kappa \kappa(x_0) \alpha^N L^N \mathbb{1}(x_0) \\
&\leqslant \|\bar{g}\|_\kappa \kappa(x_0) \|\alpha^N L^N \mathbb{1}\|.
\end{aligned}
$$

Since there exists an $n \in \mathbb{N}$ satisfying $\alpha^n \|L\|^n < 1$, letting $N = pn + q$ with $p, q \in \mathbb{N}_0$ and $q < n$, we obtain

$$
\alpha^N \|L^N \mathbb{1}\| \leqslant (\alpha^{np} \|L^{np}\|) \alpha^q \|L^q\| \leqslant (\alpha^n \|L^n\|)^p \max_{q<n} \{\alpha^q \|L^q\|\} \to 0 \text{ as } n(N) \to \infty.
$$

Therefore, letting $N \to \infty$, (26) and Lemma A.22 imply that $\bar{v}(x_0) \geqslant v_\sigma(x_0)$ for all x_0 and $\sigma \in \Sigma$. Hence, we have $\bar{v} \geqslant \sup_\sigma v_\sigma = v^*$. Next, since $\bar{g} = E\bar{v}$ by Lemma A.24 and $\bar{g}$-greedy policy exists by Lemma A.26, there exists σ^* such that

$$\begin{aligned} \bar{v}(x) &= M\bar{g}(x) = r(x, \sigma^*(x)) + g(x, \sigma^*(x)) \\ &= r(x, \sigma^*(x)) + \mathbb{E}_{x,\sigma^*(x)}\beta(x, \sigma^*(x), x')\bar{v}(x') = T_{\sigma^*}\bar{v}(x). \end{aligned} \tag{27}$$

for all $x \in \mathsf{X}$. Hence, the same iteration of (26) with $\bar{v} = T_{\sigma^*}\bar{v}$ implies $\bar{v} = v_{\sigma^*} \leqslant v^*$. We conclude that $\bar{v} = v^*$. The above arguments also show that a g^*-greedy policy σ^* is optimal: $v^* = v_{\sigma^*}$ ((b) $\Rightarrow$ (a)), and an optimal policy exists. We now show that part (a) implies (b). We write $M_\sigma g = r(x, \sigma(x)) + g(x, \sigma(x))$ for $x \in \mathsf{X}$, $g \in \mathcal{G}$ and g-greedy policy σ. Since $\bar{g} = Ev^*$, if σ^* is optimal, then we have

$$M_{\sigma^*}\bar{g} = M_{\sigma^*}Ev^* = T_{\sigma^*}v^* = T_{\sigma^*}v_{\sigma^*} = v_{\sigma^*} = v^* = Tv^* = MEv^* = M\bar{g}.$$

Hence, σ^* is $\bar{g}$-greedy. We next show that a policy is $\bar{g}$-greedy if and only if it is $M\bar{g}$-greedy. If $\sigma \in \Sigma$ is $\bar{g}$-greedy: $M_\sigma\bar{g} = M\bar{g}$, then $TM\bar{g} = MEM\bar{g} = MR\bar{g} = M\bar{g} = M_\sigma\bar{g} = M_\sigma R\bar{g} = M_\sigma EM\bar{g} = T_\sigma M\bar{g}$, whence σ is $M\bar{g}$-greedy. Conversely, if σ is $M\bar{g}$-greedy: $T_\sigma M\bar{g} = TM\bar{g}$, then $M_\sigma\bar{g} = M_\sigma R\bar{g} = M_\sigma EM\bar{g} = T_\sigma M\bar{g} = TM\bar{g} = MEM\bar{g} = MR\bar{g} = M\bar{g}$, whence σ is $\bar{g}$-greedy. Similarly, the same method shows that a policy is v^*-greedy if and only if it is Ev^*-greedy. □

Proof of Theorem 6.1. Let Condition A.1 and Assumption 6.1 hold. Part (a) follows from A.22. Part (b) follows from Lemma A.25. Part (e) follows from Part (b) and (c). Part (c), (d), (f), and (g) follow from Part (b), Lemma A.24, and A.27. □

Proof of Lemma 6.1. Let all the stated assumptions hold. We show the statement by induction. First, observe that

$$\begin{aligned} \hat{r}(x, a) &= \mathbb{E}_z u(R(\varepsilon')(w - c) + y(\varepsilon', z')) \geqslant \mathbb{E}_z u(y(z', \varepsilon')) > -\infty, \\ \bar{r}(x) &= \sup_{0 \leqslant c \leqslant w} u(c) = u(w) \leqslant pw + q =: \kappa(x). \end{aligned}$$

The last equation implies that we can set $d = 1$ in Assumption 6.1. Also, observe that

$$\mathbb{E}_{x,a}\kappa(x') = p\mathbb{E}_z(R(\varepsilon')(w - c) + y(z', \varepsilon')) + q$$

is continuous in $(x,a)=(w,z,c)$. Therefore, Assumption 6.1 (a) and (c) hold. We next show Assumption 6.1 (b). Note that

$$\begin{aligned}\frac{\mathbb{E}_{x,a}\beta(x,a,x')\kappa(x')}{\kappa(x)} &= \frac{\mathbb{E}_z\beta(z,z')(pw'+q)}{pw+q} \\ &= \frac{\mathbb{E}_z\beta(z,z')[p(R(\varepsilon')(w-c)+y(z',\varepsilon'))+q]}{pw+q} \\ &\leqslant \frac{\mathbb{E}_z\beta(z,z')[p(R(\varepsilon')w+y(z',\varepsilon'))+q]}{pw+q}.\end{aligned}$$

Since the right-hand side is a monotone function of w and then achieves the supremum at either $w=0$ or $w=\infty$, we have

$$\frac{\mathbb{E}_{x,a}\beta(x,a,x')\kappa(x')}{\kappa(x)} \leqslant \max\left\{\frac{\mathbb{E}_z\beta(z,z')[py(z',\varepsilon')+q]}{q}, \mathbb{E}_z\beta(z,z')\mathbb{E}R(\varepsilon')\right\}.$$

where the last inequality follows from that ε_t is IID. Then, since $q>1$ can be arbitrarily large, (17) implies

$$\frac{\mathbb{E}_z\beta(z,z')[py(z',\varepsilon')+q]}{q} \to \mathbb{E}_z\beta(z,z')$$

as $q\to\infty$. Since $\mathbb{E}R(\varepsilon')>1$, we have

$$\frac{\mathbb{E}_{x,a}\beta(x,a,x')\kappa(x')}{\kappa(x)} \leqslant \mathbb{E}_z\beta(z,z')\mathbb{E}R(\varepsilon') =: L\mathbb{1}(x)$$

for large enough $q>1$. Hence, we have $\mathbb{E}_{x,a}\beta(x,a,x')\kappa(x')L^0\mathbb{1}(x') \leqslant \kappa(x)L\mathbb{1}(x)$ for large enough $q>1$. Assume the induction hypothesis that $\mathbb{E}_{x,a}\beta(x,a,x')\kappa(x')L^n\mathbb{1}(x') \leqslant \kappa(x)L^{n+1}\mathbb{1}(x)$ for some $n\in\mathbb{N}$. Then, iteration yields

$$\begin{aligned}&\frac{\mathbb{E}_{x,a}\beta(x,a,x')\kappa(x')L^{n+1}\mathbb{1}(x')}{\kappa(x)} \\ &\leqslant \frac{\mathbb{E}_z\beta(z,z')[p(R(\varepsilon')w+y(z',\varepsilon'))+q]\mathbb{E}_{z'}\beta_1\beta_2\ldots\beta_{n+1}(\mathbb{E}R(\varepsilon'))^{n+1}}{pw+q} \\ &\leqslant \max\left\{\frac{\mathbb{E}_z\beta(z,z')[py(z',\varepsilon')+q]\mathbb{E}_{z'}\beta_1\beta_2\ldots\beta_{n+1}(\mathbb{E}R(\varepsilon'))^{n+1}}{q},\right. \\ &\qquad\qquad \left.\mathbb{E}_z\beta(z,z')\beta_1\beta_2\cdots\beta_{n+1}(\mathbb{E}R(\varepsilon'))^{n+2}\right\} \\ &\to \max\{\mathbb{E}_z\beta(z,z')\beta_1\beta_2\cdots\beta_{n+1}(\mathbb{E}R(\varepsilon'))^{n+1}, \\ &\qquad\qquad \mathbb{E}_z\beta(z,z')\beta_1\beta_2\cdots\beta_{n+1}(\mathbb{E}R(\varepsilon'))^{n+2}\} \\ &\leqslant L^{n+2}\mathbb{1}(x),\end{aligned}$$

where the first inequality uses the definitions of κ and $L^{n+1}\mathbb{1}$, the second inequality follows from that the supremum is at either $w = 0$ or $w = \infty$, the limit follows from taking q arbitrarily large, and the last inequality uses $\mathbb{E}R(\varepsilon') > 1$. Therefore, induction implies that there is large enough $q(n) > 1$ such that

$$\mathbb{E}_{x,a}\beta(x,a,x')\kappa(x')L^n\mathbb{1}(x') \leqslant \kappa(x)L^{n+1}\mathbb{1}(x)$$

for all $x \in \mathsf{X}$ and all $n \in \mathbb{N}$. □

A.5. **Proofs in Section 7.**

Lemma A.28. *If Condition 7.1 and , then* $R\mathcal{G} \subset \mathcal{G}$.

Proof of Lemma A.28. Let Condition 7.1 and Assumption 7.2 hold. Let $g \in \mathcal{G}$. We first show that $\|Rg\| < \infty$. Since $g \geqslant \|g\|$, we have

$$\begin{aligned} Rg(x,a) &\geqslant -\frac{\beta(x,a)}{\theta}\ln \mathbb{E}_{x,a}\exp\left(-\theta \sup_{a'\in\Gamma(x')}\{r(x',a') - \|g\|\}\right) \\ &= -\frac{\beta(x,a)}{\theta}\ln \mathbb{E}_{x,a}\exp\left(-\theta(\bar{r}(x') - \|g\|)\right) = \beta(x,a)(\hat{r}(x,a) - \|g\|). \end{aligned}$$

Hence, since $\hat{r}$ is bounded below, Rg is also bounded below. Similarly, we can show that $Rg(x,a) \leqslant \beta(x,a)(\hat{r}(x,a) + \|g\|)$ for all $(x,a) \in \mathsf{G}$. Since β is bounded and $\bar{r}$ is bounded above, $\hat{r}$ is bounded above and then Rg is bounded above. Therefore, we have $\|Rg\| < \infty$.

Next, since g and r are u.s.c., and Γ is compact-valued and u.s.c., the map $x \mapsto h(x) := \sup_{a\in\Gamma(x)}\{r(x,a) + g(x,a)\}$ is u.s.c. by the Maximum theorem. Since $\bar{r}$ and $\|g\|$ are bounded above, function h is bounded above, whence there exists a sequence of bounded continuous functions h_n such that $h_n \downarrow h$. Then, we can show by Condition 7.1 that h is u.s.c..[19]. Now, note that if $x_1 \leqslant x_2$ in X, then $x_1' = f(x_1,a,\varepsilon') \leqslant f(x_2,a,\varepsilon') = x_2'$ and then $\Gamma(x_1') \subset \Gamma(x_2')$. Since r and g are increasing in x, we have

$$h(x_1') = \sup_{a'\in\Gamma(x_1')}\{r(x_1',a') + g(x_1',a')\} \leqslant \sup_{a'\in\Gamma(x_2')}\{r(x_2',a') + g(x_2',a') = h(x_2').$$

Therefore, we have $-(1/\theta)\ln \mathbb{E}_{x_1,a}\exp(-\theta h(x_1')) \leqslant -(1/\theta)\ln \mathbb{E}_{x_2,a}\exp(-\theta h(x_2'))$. Since in addition β is increasing in x, we see that Rg is increasing in x. We conclude that $R\mathcal{G} \subset \mathcal{G}$. □

[19] See the proof of Lemma 2.1 or Chapter 3.3 of Hernández-Lerma and Lasserre (2012a).

Lemma A.29. *If Condition 7.1, Assumption 7.1, 7.2, and 7.3 hold, then R is eventually contracting on $(\mathcal{G}, \|\cdot\|)$ and has a unique fixed point in $\mathcal{G}$.*

Proof of Lemma A.29. Let Condition 7.1, Assumption 7.1, 7.2, and 7.3 hold. We show that there is an $n \in \mathbb{N}$ such that R^n satisfies Blackwell's condition. Let $g \in \mathcal{G}$ and $K \geqslant 0$. Fix $(x, a) \in \mathsf{G}$. Observe that

$$\begin{aligned} R(g+K)(x,a) &= \frac{-\beta(x,a)}{\theta} \ln \mathbb{E}_{x,a} \exp\left(-\theta \sup_{a' \in \Gamma(x')} \{r(x',a') + g(x',a') + K\} \right) \\ &= Rg(x,a) + \beta(x,a)K. \end{aligned}$$

Define the function $\varphi(t) = -1/\theta \ln \mathbb{E}_{x,a} \exp(-\theta t)$ for random variable $t = t(x')$. Since φ is monotonically increasing, iteration implies

$$\begin{aligned} R^2(g+K)(x,a) &= \beta(x,a)\varphi\left(\sup_{a' \in \Gamma(x')} \{r(x',a') + R(g+K)(x',a')\} \right) \\ &= \beta(x,a)\varphi\left(\sup_{a' \in \Gamma(x')} \{r(x',a') + Rg(x',a') + \beta(x',a')K\} \right) \\ &\leqslant \beta(x,a)\varphi\left(\sup_{a' \in \Gamma(x')} \{r(x',a') + Rg(x',a')\} + K \sup_{a' \in \Gamma(x')} \{\beta(x',a')\} \right). \end{aligned}$$

Let $X = \sup_{a' \in \Gamma(x')} \{r(x',a') + Rg(x',a')\}$ and $Y = K \sup_{a' \in \Gamma(x')} \{\beta(x',a')\}$. Since $Rg(x',a')$, $r(x',x')$, $\beta(x',a')$, and $\Gamma(x')$ are increasing in x', and $x' = f(x,a,\varepsilon')$ is increasing in x and ε' for independent ε', we have $\operatorname{Cov}(e^{-\theta X}, e^{-\theta Y} | (x,a)) \geqslant 0$. Therefore, since $\mathbb{E}_{x,a} e^{-\theta X} e^{-\theta Y} \geqslant \mathbb{E}_{x,a} e^{-\theta X} \mathbb{E}_{x,a} e^{-\theta Y}$, we have[20]

$$\begin{aligned} \varphi(X+Y) &= \frac{-1}{\theta} \ln \mathbb{E}_{x,a} e^{-\theta X - \theta Y} \\ &\leqslant \frac{-1}{\theta} \ln \left(\mathbb{E}_{x,a} e^{-\theta X} \mathbb{E}_{x,a} e^{-\theta Y} \right) = \varphi(X) + \varphi(Y). \end{aligned}$$

We then have

$$\begin{aligned} &R^2(g+K)(x,a) \\ &\quad \leqslant \beta(x,a)\varphi\left(\sup_{a' \in \Gamma(x')} \{r(x',a') + Rg(x',a')\} \right) + \beta(x,a)\varphi\left(K \sup_{a' \in \Gamma(x')} \{\beta(x',a')\} \right) \\ &\quad = R^2 g(x,a) + \beta(x,a)\varphi\left(K \sup_{a' \in \Gamma(x')} \{\beta(x',a')\} \right). \end{aligned}$$

[20] See also Lemma 3 of Bäuerle and Jaśkiewicz (2018).

Now, since $t \mapsto \exp(-\theta t)$ is convex, Jensen inequality implies

$$\varphi(Y) = \frac{-1}{\theta} \ln \mathbb{E}_{x,a} e^{-\theta Y} \leqslant \frac{-1}{\theta} \ln e^{-\theta \mathbb{E}_{x,a} Y} = \mathbb{E}_{x,a} Y.$$

Therefore, we obtain

$$\begin{aligned} R^2(g+K)(x,a) &\leqslant R^2 g(x,a) + K\beta(x,a)\mathbb{E}_{x,a} \sup_{a' \in \Gamma(x'a)} \beta(x',a') \\ &\leqslant R^2 g(x,a) + KL\bar{\beta}(x), \end{aligned}$$

where L and $\bar{\beta}$ are defined in Assumption 7.3. With the same argument, the induction shows that $R^n(g+K)(x,a) \leqslant R^n g(x,a) + KL^{n-1}\beta(x)$ for all $n \in \mathbb{N}$. Since Assumption 7.3 implies that there is $m \in \mathbb{N}$ such that $\|L^{m-1}\bar{\beta}\| = \sup_x L^{m-1}\bar{\beta}(x) < 1$, we have

$$R^m(g+K)(x,a) \leqslant R^m g(x,a) + K\|L^{m-1}\bar{\beta}\|.$$

Since (x,a) is arbitrary, we have $R^m(g+K) \leqslant R^m g + K\|L^{m-1}\bar{\beta}\|$, whence R^m satisfies the Blackwell's condition. Hence, R^m is a contracting map. Then, since $\mathcal{G}$ is a Banach space, R admits a unique fixed point in $\mathcal{G}$ by the generalized Contracting Mapping theorem. □

Lemma A.30. *If Condition 7.1, Assumption 7.1, 7.2, and 7.3 hold, then for all constant $c \in \mathbb{R}$ we have $v_\sigma(x) = \limsup_{n\to\infty} T_\sigma^n \bar{r}(x) = \limsup_{n\to\infty} T_\sigma^n(\bar{r}+c)(x)$ for all $x \in \mathsf{X}$ and $\sigma \in \Sigma$.*

Proof of Lemma A.30. Let Condition 7.1, Assumption 7.1, 7.2, and 7.3 hold. Fix $\sigma \in \Sigma$ and constant $K \geqslant 0$. Since $\bar{r} + K \geqslant \bar{r}$, the monotonicity of T_σ implies $\limsup_{n\to\infty} T_\sigma^n(\bar{r}+K) \geqslant \limsup_{n\to\infty} T_\sigma^n \bar{r}$. Next, similar to the iteration in Lemma A.29, we have, for all $x \in \mathsf{X}$,

$$\begin{aligned} T_\sigma(\bar{r}+K)(x) &= r_\sigma(x) - \frac{\beta_\sigma(x)}{\theta} \ln \mathbb{E}_x^\sigma e^{-\theta(\bar{r}+K)(x')} \\ &= r_\sigma(x) - \frac{\beta_\sigma(x)}{\theta} \ln e^{-\theta K} \mathbb{E}_x^\sigma e^{-\theta \bar{r}(x')} \\ &= r_\sigma(x) - \frac{\beta_\sigma(x)}{\theta} \ln \mathbb{E}_x^\sigma e^{-\theta \bar{r}(x')} + \beta_\sigma(x)K = T_\sigma \bar{r}(x) + \beta_\sigma(x)K, \end{aligned}$$

where $r_\sigma(x) = r(x,\sigma(x))$ and $\beta_\sigma(x) = \beta(x,\sigma(x))$ for all $x \in \mathsf{X}$. Since r_σ, β_σ, $x' = f(x,\sigma(x),\varepsilon')$ are increasing in x, we see that $T_\sigma \bar{r}(x)$ is increasing in x. Then, it follows

from the argument in Lemma A.29 and Assumption 7.1 that $\mathbb{E}_x^\sigma e^{-\theta T_\sigma \bar{r}(x')} e^{-\theta \beta_\sigma(x') K} \geqslant \mathbb{E}_x^\sigma e^{-\theta T_\sigma \bar{r}(x')} \mathbb{E}_x^\sigma e^{-\theta \beta_\sigma(x') K}$, so iteration yields that for all $x \in \mathsf{X}$

$$\begin{aligned}
T_\sigma^2(\bar{r}+K)(x) &= r_\sigma(x) - \frac{\beta_\sigma(x)}{\theta} \ln \mathbb{E}_x^\sigma e^{-\theta T_\sigma(\bar{r}+K)(x')} \\
&= r_\sigma(x) - \frac{\beta_\sigma(x)}{\theta} \ln \mathbb{E}_x^\sigma e^{-\theta (T_\sigma \bar{r}(x') + \beta_\sigma(x') K)} \\
&= r_\sigma(x) - \frac{\beta_\sigma(x)}{\theta} \ln \mathbb{E}_x^\sigma e^{-\theta T_\sigma \bar{r}(x')} e^{-\theta \beta_\sigma(x') K} \\
&\leqslant r_\sigma(x) - \frac{\beta_\sigma(x)}{\theta} \ln \mathbb{E}_x^\sigma e^{-\theta T_\sigma \bar{r}(x')} \mathbb{E}_x^\sigma e^{-\theta \beta_\sigma(x') K} \\
&= T_\sigma^2 \bar{r}(x) - \frac{\beta_\sigma(x)}{\theta} \ln \mathbb{E}_x^\sigma e^{-\theta \beta_\sigma(x') K} \\
&\leqslant T_\sigma^2 \bar{r}(x) + K \beta_\sigma(x) \mathbb{E}_x^\sigma \beta_\sigma(x') \\
&\leqslant T_\sigma^2 \bar{r}(x) + K L \bar{\beta}(x)
\end{aligned}$$

where the second inequality follows from Jensen inequality as iteration in Lemma A.29, and the last inequality follows from Assumption 7.3. Then, using the same induction, we have

$$T_\sigma^n(\bar{r}+K)(x) \leqslant T_\sigma^n \bar{r}(x) + K L^{n-1} \bar{\beta}(x).$$

Hence, since $\rho(L) < 1$, there exists $m \in \mathbb{N}$ such that $\|L^m \bar{b}\| < 1$, which implies that $\limsup_{n\to\infty} L^n \bar{\beta}(x) \to 0$ for all $x \in \mathsf{X}$. Therefore, we have $\limsup_{n\to\infty} T_\sigma^n(\bar{r} + K)(x) \leqslant \limsup_{n\to\infty} T_\sigma^n \bar{r}(x)$ for all $x \in \mathsf{X}$. Then, we have $\limsup_{n\to\infty} T_\sigma^n(\bar{r}+K)(x) = \limsup_{n\to\infty} T_\sigma^n \bar{r}(x)$ for all $x \in \mathsf{X}$ when $K \geqslant 0$. Similarly, we can show the statement when $K < 0$, where the above inequalities are reversed. □

Proof of Theorem 7.1. Let Condition 7.1, Assumption 7.1, 7.2, and 7.3 hold. Part (a) and (b) follow from A.28 and A.29. Part (c), (d), and (e) of optimality follow from the proof of Theorem 5.3 of Ma et al. (2022) and Lemma A.30. □

References

Albuquerque, R., M. Eichenbaum, V. X. Luo, and S. Rebelo (2016): "Valuation risk and asset pricing," *The Journal of Finance*, 71, 2861–2904.

Arellano, C. (2008): "Default risk and income fluctuations in emerging economies," *American economic review*, 98, 690–712.

Backus, D., A. Ferriere, and S. Zin (2015): "Risk and ambiguity in models of business cycles," *Journal of Monetary Economics*, 69, 42–63.

Balbus, L., W. Olszewski, K. Reffett, and L. P. Wozny (2022): "Iterative Monotone Comparative Statics," Available at SSRN: https://ssrn.com/abstract=4039543 or http://dx.doi.org/10.2139/ssrn.4039543.

Bäuerle, N. and A. Jaśkiewicz (2018): "Stochastic optimal growth model with risk sensitive preferences," *Journal of Economic Theory*, 173, 181–200.

Becker, G. S. and C. B. Mulligan (1997): "The endogenous determination of time preference," *The Quarterly Journal of Economics*, 112, 729–758.

Bertsekas, D. (2022): *Abstract dynamic programming*, Athena Scientific.

Bloise, G., C. Le Van, and Y. Vailakis (2021): "Do not blame Bellman: It is Koopmans' fault," Available at SSRN: https://ssrn.com/abstract=4039543 or http://dx.doi.org/10.2139/ssrn.4039543.

Bodenstein, M. (2011): "Closing large open economy models," *Journal of International Economics*, 84, 160–177.

——— (2013): "Equilibrium stability in open economy models," *Journal of Macroeconomics*, 35, 1–13.

Borovička, J. and J. Stachurski (2020): "Necessary and sufficient conditions for existence and uniqueness of recursive utilities," *The Journal of Finance*, 75, 1457–1493.

Bühler, T. and D. A. Salamon (2018): *Functional analysis*, vol. 191, American Mathematical Soc.

Campbell, J. Y. and J. Ammer (1993): "What moves the stock and bond markets? A variance decomposition for long-term asset returns," *The journal of finance*, 48, 3–37.

Cheney, E. W., E. Cheney, and W. Cheney (2001): *Analysis for applied mathematics*, vol. 1, Springer.

Choi, H., N. C. Mark, and D. Sul (2008): "Endogenous discounting, the world saving glut and the US current account," *Journal of international Economics*, 75, 30–53.

Christensen, T. M. (2022): "Existence and uniqueness of recursive utilities without boundedness," *Journal of Economic Theory*, 200, 105413.

Cochrane, J. (2009): *Asset pricing: Revised edition*, Princeton university press.

Cochrane, J. H. (2011): "Presidential address: Discount rates," *The Journal of finance*, 66, 1047–1108.

Cohen, J., K. M. Ericson, D. Laibson, and J. M. White (2020): "Measuring time preferences," *Journal of Economic Literature*, 58, 299–347.

Durdu, C. B., E. G. Mendoza, and M. E. Terrones (2009): "Precautionary demand for foreign assets in Sudden Stop economies: An assessment of the New Mercantilism," *Journal of development Economics*, 89, 194–209.

Dutta, D. and Y. Yang (2013): "Endogenous time preference: evidence from Australian households' behaviour," School of Economics, The University of Sydney, Available at http://hdl.handle.net/2123/9265.

Epstein, L. G. and J. A. Hynes (1983): "The rate of time preference and dynamic economic analysis," *Journal of Political Economy*, 91, 611–635.

Erol, S., C. Le Van, and C. Saglam (2011): "Existence, optimality and dynamics of equilibria with endogenous time preference," *Journal of Mathematical Economics*, 47, 170–179.

Guest, R. S. and I. M. McDonald (2001): "How Uzawa preferences improve the simulation properties of the small open economy model," *Journal of Macroeconomics*, 23, 417–440.

Hansen, L. P. and E. Renault (2010): "Pricing kernels," *Encyclopedia of Quantitative Finance*.

Hansen, L. P. and T. J. Sargent (1995): "Discounted linear exponential quadratic gaussian control," *IEEE Transactions on Automatic control*, 40, 968–971.

Harrison, J. M. and D. M. Kreps (1978): "Speculative investor behavior in a stock market with heterogeneous expectations," *The Quarterly Journal of Economics*, 92, 323–336.

Hashimzade, N., O. Kirsanov, and T. Kirsanova (2023): "Distributional effects of endogenous discounting," *Mathematical Social Sciences*, 122, 1–6.

Hatchondo, J. C., L. Martinez, and H. Sapriza (2009): "Heterogeneous borrowers in quantitative models of sovereign default," *International Economic Review*, 50, 1129–1151.

Hatchondo, J. C., L. Martinez, and C. Sosa-Padilla (2016): "Debt dilution and sovereign default risk," *Journal of Political Economy*, 124, 1383–1422.

Helpman, E. and A. Razin (1982): "Dynamics of a floating exchange rate regime," *Journal of political Economy*, 90, 728–754.

Hernández-Lerma, O. and J. B. Lasserre (2012a): *Discrete-time Markov control processes: basic optimality criteria*, vol. 30, Springer Science & Business Media.

——— (2012b): *Further topics on discrete-time Markov control processes*, vol. 42, Springer Science & Business Media.

Hills, T. S. and T. Nakata (2018): "Fiscal multipliers at the zero lower bound: the role of policy inertia," *Journal of Money, Credit and Banking*, 50, 155–172.

Hills, T. S., T. Nakata, and S. Schmidt (2019): "Effective lower bound risk," *European Economic Review*, 120, 103321.

Hubmer, J., P. Krusell, and A. A. Smith Jr (2021): "Sources of US wealth inequality: Past, present, and future," *NBER Macroeconomics Annual*, 35, 391–455.

Huffman, D., R. Maurer, and O. S. Mitchell (2019): "Time discounting and economic decision-making in the older population," *The Journal of the Economics of Ageing*, 14, 100121.

Izadi, H. and M. S. Lamsoo (2022): "The role of the Frisch elasticity on households' behavior using the endogenous discount factor model," *Çankırı Karatekin Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 12, 251–267.

Jasso-Fuentes, H., R. R. López-Martínez, and J. A. Minjárez-Sosa (2022): "Some advances on constrained Markov decision processes in Borel spaces with random state-dependent discount factors," *Optimization*, 1–27.

Justiniano, A. and G. E. Primiceri (2008): "The time-varying volatility of macroeconomic fluctuations," *American Economic Review*, 98, 604–641.

Krasnosel'skii, M. A., G. M. Vainikko, R. Zabreyko, Y. B. Rutticki, and V. V. Stet'senko (2012): *Approximate solution of operator equations*, Springer Science & Business Media.

Lawrance, E. C. (1991): "Poverty and the rate of time preference: evidence from panel data," *Journal of Political economy*, 99, 54–77.

Lucas Jr, R. E. (1978): "Asset prices in an exchange economy," *Econometrica: journal of the Econometric Society*, 1429–1445.

Ma, Q., J. Stachurski, and A. A. Toda (2022): "Unbounded dynamic programming via the Q-transform," *Journal of Mathematical Economics*, 100, 102652.

Maeda, A. and M. Nagaya (2023): "Exhaustible resource use under endogenous time preference," *International Journal of Economic Policy Studies*, 17, 223–248.

Marinacci, M. and L. Montrucchio (2010): "Unique solutions for stochastic recursive utilities," *Journal of Economic Theory*, 145, 1776–1804.

Mendoza, E. G. (1991): "Real business cycles in a small open economy," *The American Economic Review*, 797–818.

Minjárez-Sosa, J. A. (2015): "Markov control models with unknown random state–action-dependent discount factors," *Top*, 23, 743–772.

Nakata, T. (2016): "Optimal fiscal and monetary policy with occasionally binding zero bound constraints," *Journal of Economic Dynamics and control*, 73, 220–240.

Obstfeld, M. (1990): "Intertemporal dependence, impatience, and dynamics," *Journal of Monetary Economics*, 26, 45–75.

Olszewski, W. (2021): "On sequences of iterations of increasing and continuous mappings on complete lattices," *Games and Economic Behavior*, 126, 453–459.

Ren, G. and J. Stachurski (2021): "Dynamic programming with value convexity," *Automatica*, 130, 109641.

Rosenberg, J. V. and R. F. Engle (2002): "Empirical pricing kernels," *Journal of Financial Economics*, 64, 341–372.

Samwick, A. A. (1998): "Discount rate heterogeneity and social security reform," *Journal of Development Economics*, 57, 117–146.

Sargent, T. J. and J. Stachurski (2023): "Dynamic programming volume 1," QuantEcon, Available at https://dp.quantecon.org/ or https://github.com/QuantEcon/book-dp1.

Schmitt-Grohé, S. and M. Uribe (2003): "Closing small open economy models," *Journal of international Economics*, 61, 163–185.

Stachurski, J., O. Wilms, and J. Zhang (2022a): "Asset Pricing Models with Preference Shocks: Existence and Uniqueness," Available at SSRN: https://ssrn.com/abstract=4041393 or http://dx.doi.org/10.2139/ssrn.4041393.

——— (2022b): "Unique Solutions to Power-Transformed Affine Systems," *arXiv preprint arXiv:2212.00275*.

Stachurski, J. and J. Zhang (2021): "Dynamic programming with state-dependent discounting," *Journal of Economic Theory*, 192, 105190.

Stern, M. L. (2006): "Endogenous time preference and optimal growth," *Economic Theory*, 29, 49–70.

Stokey, N. L. (1989): *Recursive methods in economic dynamics*, Harvard University Press.

Toda, A. A. (2021): "Perov's contraction principle and dynamic programming with stochastic discounting," *Operations Research Letters*, 49, 815–819.

——— (2023): "Unbounded Markov Dynamic Programming with Weighted Supremum Norm Perov Contractions," *arXiv preprint arXiv:2310.04593*.

Uzawa, H. (1968): "Time preference, the consumption function, and optimum asset holdings," *Value, capital and growth: papers in honor of Sir John Hicks. The University of Edinburgh Press, Edinburgh*, 485–504.

Vasilev, A. (2022a): "A Real-Business-Cycle Model with Endogenous Discounting and a Government Sector," *Notas Económicas*, 73–86.

——— (2022b): "Um Modelo de Ciclos Económicos Reais com Taxa de Desconto Endógena e Estado," *Notas Económicas*, 71–84.

Wei, Q. and X. Guo (2011): "Markov decision processes with state-dependent discount factors and unbounded rewards/costs," *Operations Research Letters*, 39, 369–374.

Weil, P. (1993): "Precautionary savings and the permanent income hypothesis," *The Review of Economic Studies*, 60, 367–383.

Wu, X. and J. Zhang (2016): "Finite approximation of the first passage models for discrete-time Markov decision processes with varying discount factors," *Discrete Event Dynamic Systems*, 26, 669–683.

Wu, X., X. Zou, and X. Guo (2015): "First passage Markov decision processes with constraints and varying discount factors," *Frontiers of Mathematics in China*, 10, 1005–1023.

Yue, V. Z. (2010): "Sovereign default and debt renegotiation," *Journal of international Economics*, 80, 176–187.

Zhou, L. (1994): "The set of Nash equilibria of a supermodular game is a complete lattice," *Games and economic behavior*, 7, 295–300.