

# Unbounded Dynamic Programming via the Q-Learning Transform<sup>1</sup>

Qingyin Ma<sup>a</sup>, John Stachurski<sup>b</sup>, Alexis Akira Toda<sup>c</sup>

<sup>a</sup>ISEM, Capital University of Economics and Business

<sup>b</sup>Research School of Economics, Australian National University

<sup>c</sup>Department of Economics, University of California San Diego

December 1, 2020

**ABSTRACT.** We propose a new approach to solving dynamic decision problems with unbounded rewards, based on an application of the Q-learning transform to the Bellman equation. Q-learning is a technique from the reinforcement learning literature with strong convergence properties. In our case, the objective of the transform is to convert an unbounded dynamic program into a bounded one. The approach is general enough to handle problems for which existing methods struggle, and yet is accessible to applied researchers and Ph.D. students. We show by example that many common decision problems satisfy our conditions.

*JEL Classifications:* C61, C65

*Keywords:* Dynamic programming, Optimality, Reinforcement learning

## 1. INTRODUCTION

Dynamic programming forms the backbone of modern economics. Every year, thousands of students in graduate programs around the world learn the standard methodology for infinite horizon problems with discounting. Constructed primarily by Blackwell (1962, 1965), this theory uses contraction mappings over spaces of bounded functions metrized by the supremum norm. The approach is elegant, powerful in terms of deriving theoretical results and, when applicable, generates globally convergent

---

<sup>1</sup>We thank Takashi Kamihigashi and Yiannis Vailakis for valuable feedback and suggestions, as well as audience members at the Econometric Society meeting in Auckland in 2018 and the 2nd Conference on Structural Dynamic Models in Copenhagen. Financial supports from ARC Discovery Grant DP120100321 and NSFC No.72003138 are gratefully acknowledged.

*Email addresses:* qingyin.ma@cueb.edu.cn, john.stachurski@anu.edu.au, atoda@ucsd.edu

algorithms. Standard textbook treatments can be found in [Stokey et al. \(1989\)](#) and [Bertsekas \(2017\)](#).

Unfortunately, after investing in these techniques, students soon learn that the approach is not in fact applicable in the vast majority of concrete economic problems, due to the fact that almost all reward functions used in applications are unbounded. For example, in quantitative work, the most commonly used flow utility function is the constant relative risk aversion (CRRA) specification

$$u(x) = \begin{cases} \frac{x^{1-\gamma}}{1-\gamma} & \text{if } \gamma > 0 \text{ and } \gamma \neq 1, \\ \log x & \text{if } \gamma = 1, \end{cases} \quad (1)$$

where  $\gamma > 0$  is the risk aversion coefficient. The function  $u$  is unbounded above if  $0 < \gamma < 1$ , is unbounded below if  $\gamma > 1$ , and is unbounded both from above and below if  $\gamma = 1$ . Unbounded reward functions violate Blackwell's conditions.

The need to deal with unbounded reward functions has led researchers to build various extensions of Blackwell's theory. These extensions typically involve either (a) recovering contractivity by modifying the metric that measures distance between candidate value functions, or (b) introducing a weaker form of contractivity that preserves at least some of Blackwell's optimality results. The former approach is exemplified by the weighted supremum norm method, introduced by [Wessels \(1977\)](#) and applied to economic problems by [Boyd \(1990\)](#), [Alvarez and Stokey \(1998\)](#), [Bäuerle and Jaśkiewicz \(2018\)](#) and several other authors.<sup>2</sup> The second approach can be seen in the work of [Rincón-Zapatero and Rodríguez-Palmero \(2003\)](#) and [Martins-da-Rocha and Vailakis \(2010\)](#) for the deterministic case and [Matkowski and Nowak \(2011\)](#) for the stochastic case, who apply local contractions on successively larger subsets of the state space.

While these techniques are ingenious, and certainly important from a theoretical perspective, their direct impact on quantitative applications in economics has, as yet, been limited. Weighted supremum norms work well with certain problems but struggle with others, such as when rewards are unbounded below. Local contraction methods are broadly applicable under reasonable assumptions but require testing of technical conditions involving increasing sequences of compact sets that exhaust the

---

<sup>2</sup>Modern summaries of the method can be found in [Hernández-Lerma and Lasserre \(1999\)](#), [Bäuerle and Rieder \(2011\)](#) and [Bertsekas \(2018\)](#).

state space.<sup>3</sup> Proofs of convergence properties are significantly more complex than the bounded case, and the statements of the theorems are more challenging to interpret.

In this paper we take an alternative route. Rather than transforming the standard contraction mapping theory of Blackwell to handle unbounded dynamic programs, we transform the unbounded dynamic programs into bounded ones so that standard contraction mapping theory can be applied. The transformation that we use maps value functions into the “action-value” functions used in Q-learning, which is a popular reinforcement learning algorithm that allows online updating by a controller in an incremental fashion.<sup>4</sup> It has been shown that the algorithm has strong global convergence properties and our results are in this spirit.

While our methodology differs from the standard Q-learning algorithm (which focuses on online learning in the presence of model uncertainty), it uses the same transform. The core idea is as follows: In standard dynamic programs, the Bellman operator is defined by composing the following three operations: given a candidate value function  $v$ , (a) compute the discounted expectation  $g := \beta \mathbb{E}v$  over current states and actions, (b) add current reward  $r$  and obtain  $r + g$ , and (c) maximize over current feasible actions and update as  $v = \max \{r + g\}$ . Alternatively, given  $g$ , which corresponds to the action-value function from Q-learning, one can define the transformed Bellman operator by applying operations  $(b) \rightarrow (c) \rightarrow (a)$  to update  $g$ . We show that, under relatively weak and easily testable conditions, the action-value function is bounded, the transformed Bellman operator is a contraction with unique fixed point  $g^*$ , and the true value function  $v^*$  can be recovered as  $v^* = \max \{r + g^*\}$ .

There are significant advantages to the transformation-based approach to unbounded programs adopted in this paper. One is that the methodology fits well with a case where the weighted supremum norm approach struggles: maximization problems where rewards are unbounded below.<sup>5</sup> Such optimization problems are commonplace in quantitative applications, such as those involving CRRA flow utility with

---

<sup>3</sup>It is required that the decision problem satisfies contractivity on each of these compact sets, and some control over the way that contractivity fades must also be imposed (see, e.g., [Matkowski and Nowak \(2011\)](#), Assumptions A1–A5, C1–C2 and D1–D2).

<sup>4</sup>Q-learning was introduced by [Watkins and Dayan \(1992\)](#). This methodology has had some striking successes in recent years, as part of a suite of artificial intelligence tools that typically includes deep neural networks to approximate the action-value function. A recent exposition and summary of the theoretical literature on Q-learning and other reinforcement learning algorithms can be found in [Szepesvári \(2010\)](#).

<sup>5</sup>For background, see [Le Van and Vailakis \(2005\)](#) or [Jaśkiewicz and Nowak \(2011\)](#).

$\gamma > 1$ , which is the empirically relevant case. We show that, for many canonical applications from this class of problems, the Q-learning transform converts unbounded value functions into bounded action-value functions. Standard contraction mapping theory can then be applied.

A second advantage of the Q-learning transform method is that it has no difficulty handling stochastic dynamic programs. In fact, the action-value functions associated with the Q-learning transform tend to be well-behaved and regular in the stochastic case, due to the fact that conditional expectations operators have a smoothing effect on functions. In contrast, many existing papers on unbounded dynamic programming either focus on the deterministic case or restrict the support of shocks.<sup>6</sup>

A third advantage is that, despite its generality, the Q-learning transform approach is accessible to a general audience. In particular, it requires only the standard contraction mapping theorem. In addition, we provide many canonical examples such as optimal savings, optimal default, job search, and optimal portfolio, all with unbounded rewards and in stochastic environments.

On a technical level, the contribution of our paper is twofold. First, we identify general sufficient conditions under which unbounded dynamic programs can be transformed into bounded ones. Second, we prove that, when such a transformation is available, the solution to the transformed problem is equal to that to the original problem. To the best of our knowledge, this is the first research in which the Q-learning transform has been used to convert unbounded reward dynamic programs into bounded ones.<sup>7</sup>

There are connections between our work and the study of unbounded dynamic programming in [Kamihigashi \(2014\)](#). Assuming the Bellman operator maps an order interval of functions into itself and some transversality-like conditions hold, [Kamihigashi \(2014\)](#) shows the existence and uniqueness of the fixed point of the Bellman equation and obtains optimality properties. The relative advantages of the approach

---

<sup>6</sup>[Alvarez and Stokey \(1998\)](#) handle certain homogeneous problems using weighted supremum norm methods, although they focus on the deterministic case. A generalization to the stochastic case requires bounds on the maximum growth rate. Assumptions D1-D2 of [Matkowski and Nowak \(2011\)](#) also require the state not to jump too much. These assumptions are strong from an applied perspective.

<sup>7</sup>Researchers in economics have used alternative transformations of the Bellman equation when studying dynamic programming problems, including [Rust \(1987\)](#), [Jovanovic \(1982\)](#), [Abbring et al. \(2018\)](#) and [Ma and Stachurski \(2020\)](#). These transformations are typically aimed at improving economic intuition, estimation properties or computational efficiency.

presented here include treating stochastic decision problems ([Kamihigashi \(2014\)](#) restricts attention to the deterministic case) and obtaining uniform geometric rates of convergence in value function iteration, rather than pointwise convergence.

Our work is also related to results in [Van Der Wal \(1980\)](#) and [Jaśkiewicz and Nowak \(2011\)](#), which explicitly admit problems with rewards that are unbounded below. In this setting, [Jaśkiewicz and Nowak \(2011\)](#) show that the value function of a Markov decision process is a solution to the Bellman equation. The methodology developed here strengthens their results by adding uniqueness and proving that value function iteration leads to an optimal policy. In an extension section, we combine our methodology with the weighted supremum norm approach, allowing us to handle problems that are both unbounded above and unbounded below.

Some studies have approached dynamic programming with unbounded rewards via an Euler equation method, as seen for example in [Li and Stachurski \(2014\)](#) and [Ma et al. \(2020\)](#). This methodology can be powerful but is limited in scope. For example, in Section 4, we show how the Q-learning transform method can be applied to the kinds of optimal savings problem with endogenous labor choice that are common in applied quantitative work (see, e.g., [Castañeda et al. \(2003\)](#)). The Euler equation method of [Li and Stachurski \(2014\)](#) and [Ma et al. \(2020\)](#) is not applicable in this setting because the choice variable (consumption and labor) is multi-dimensional. Similarly, the Euler equation method of [Ma et al. \(2020\)](#) is not applicable in the savings problem in Section 4.5, due to nontrivial portfolio choice. Optimal consumption-portfolio problems have been mostly studied in the literature under special homogeneity assumptions ([Samuelson, 1969](#); [Toda, 2014](#)) or finite horizon ([He and Pearson, 1991](#)). Our framework shows that the problem can be studied in an infinite-horizon environment when the utility function is unbounded below.

The Q-learning transformation is not limited to dynamic programs that are additively separable. In a recent paper, [Bäuerle and Jaśkiewicz \(2018\)](#) study an optimal growth model in the presence of risk-sensitive preference, in which the agent is risk averse in future utility (in addition to being risk averse in future consumption).<sup>8</sup> They provide valuable optimality results, although these results cannot treat many common

---

<sup>8</sup>This is in comparison with the classical additively separable preference model, where the agent is risk neutral in future utility. The additively separable preference is a special limiting case by letting  $\gamma \rightarrow 0$  and using the property  $\lim_{\gamma \rightarrow 0} -\frac{1}{\gamma} \log \mathbb{E} e^{-\gamma X} = \mathbb{E} X$  for a random variable  $X$ . Further comments on risk-sensitive preference can be found in [Föllmer and Schied \(2004\)](#), [Bäuerle and Rieder \(2011\)](#), [Bäuerle and Jaśkiewicz \(2018\)](#) and references cited therein. Models with risk-sensitive preference are also related to robust control problems, as discussed in [Hansen and Sargent \(2008\)](#).

period utility functions, such as CRRA with relative risk aversion at least 1, because they exclude all utility functions that are unbounded below. Furthermore, an optimal growth model is a rather special dynamic program. In an extension section we present a general theory of dynamic programming with risk-sensitive preferences via transformation of the Bellman equation.

The rest of our paper is structured as follows. Section 2 starts the exposition with typical examples. Section 3 presents the general theory when rewards are bounded above (though potentially unbounded below). Section 4 provides additional applications. Section 5 extends the general theory to the case when rewards are unbounded both from above and below using the weighted supremum norm approach, and also considers the case with recursive (risk-sensitive) preferences. Most proofs are deferred to the appendix.

## 2. EXAMPLE APPLICATIONS

We first illustrate the methodology for converting unbounded problems to bounded ones in some relatively simple settings. More sophisticated applications are deferred to Section 4 after the presentation of the theory.

**2.1. Application 1: Optimal Savings.** Consider an optimal savings problem where a borrowing constrained agent solves

$$\begin{aligned} \text{maximize} \quad & \mathbb{E} \sum_{t=0}^{\infty} \beta^t u(c_t) \end{aligned} \tag{2a}$$

$$\text{subject to} \quad 0 \leq c_t \leq w_t, \tag{2b}$$

$$w_{t+1} = R(w_t - c_t) + y_{t+1}, \tag{2c}$$

with  $(w_0, y_0)$  given. Here  $\beta \in (0, 1)$  is the discount factor,  $c_t, w_t, y_t \geq 0$  are, respectively, consumption, wealth and non-financial income at time  $t$ ,  $R \geq 0$  is the gross rate of return on financial income,<sup>9</sup> and  $u: \mathbb{R}_+ \rightarrow \mathbb{R} \cup \{-\infty\}$  is a utility function, which is increasing and continuous. For now, suppose that  $u$  is bounded above but

---

<sup>9</sup>By convention, the financial wealth  $w_t$  in the budget constraint in (2c) includes the current non-financial income  $y_t$ . One can modify the budget constraint to an alternative timing such as  $w_{t+1} = R(w_t - c_t + y_t)$ , where the time  $t$  financial wealth  $w_t$  excludes current income  $y_t$ , and the arguments below still go through after suitable modifications. An application along these lines is given in Section 4.4.

unbounded below, with  $u(0) = -\infty$ . This is the case for, say, the constant relative risk aversion (CRRA) specification (1) with  $\gamma > 1$  (as in much of the literature).

Assume that  $\{y_t\}$  satisfies  $y_t = y(z_t, \xi_t)$ , where  $z_t$  is a Markov process with state space  $Z$ ,  $\xi_t$  is an IID shock of arbitrary dimension, and  $y$  is a nonnegative measurable function. The Bellman equation of this problem is

$$v(w, z) = \sup_{0 \leq c \leq w} \{u(c) + \beta \mathbb{E}_z v(R(w - c) + y', z')\}, \quad (3)$$

where  $y' = y(z', \xi')$ . The value function is unbounded below, and the classical arguments in Blackwell (1965) cannot be applied.<sup>10</sup>

Consider, however, the following line of argument. Suppose that

$$\inf_z \mathbb{E}_z u(y(z', \xi')) > -\infty, \quad (4)$$

which is a relatively mild restriction.<sup>11</sup> Let

$$g(w, z, c) := \beta \mathbb{E}_z v(R(w - c) + y', z'). \quad (5)$$

The function  $g$  is called the *action-value function*, since it returns the value of the state after committing to a given action in the current period (and using continuation values dictated by  $v$  thereafter). The Q-learning approach begins by rewriting the Bellman equation in terms of the action-value function alone. Our approach slightly differs, in that we do not include current rewards and actions in the action-value function, but a similar idea can be applied.

To show this, we combine (3) and (5) to give

$$v(w, z) = \sup_{0 \leq c \leq w} \{u(c) + g(w, z, c)\}. \quad (6)$$

We eliminate the function  $v$  from (6) by using the definition of  $g$  in (5). The first step is to evaluate  $v$  in (6) at  $(R(w - c) + y', z')$ , which gives

$$v(R(w - c) + y', z') = \sup_{0 \leq c' \leq R(w - c) + y'} \{u(c') + g(R(w - c) + y', z', c')\}.$$

---

<sup>10</sup>To confirm this, suppose to the contrary that  $v$  is the value function and  $|v| \leq M < \infty$ . Then  $-M \leq v(0, z) \leq u(0) + \beta M = -\infty$ . Contradiction.

<sup>11</sup>The expectation should be understood as  $\mathbb{E}[u(y(z_{t+1}, \xi_{t+1})) \mid z_t = z]$ . The condition in (4) holds if, say,  $y(z, \xi) = z$  for all  $\xi$  and  $Z$  is finite and positive (Aiyagari, 1994; Açıkgöz, 2018; Cao, 2020), or if income has a persistent-transitory representation (Ejrnæs and Browning, 2014) such as  $\log y(z, \xi) = \mu(z) + \sigma(z)\xi$ , with suitable distributional assumptions (e.g.,  $u$  is CRRA,  $z$  is a finite state Markov chain, and  $\xi$  has a finite moment generating function).

Taking the conditional expectation of both sides with respect to  $z$ , multiplying by  $\beta$  and using (5) again now gives

$$g(w, z, c) = \beta \mathbb{E}_z \sup_{0 \leq c' \leq R(w-c)+y'} \{u(c') + g(R(w-c) + y', z', c')\}, \quad (7)$$

which is a functional equation in  $g$ . Consider a transformed Bellman operator  $S$  such that  $Sg(w, z, c)$  is equal to the right hand side of (7). By construction, any solution  $g$  of (7) is a fixed point of  $S$  and vice versa. Let  $\mathcal{G}$  be the space of bounded measurable functions on the set  $\mathbf{D}$  defined by

$$\mathbf{D} := \{(w, z, c) \in \mathbb{R}_+ \times \mathbf{Z} \times \mathbb{R}_+ : c \leq w\}$$

equipped with the supremum norm  $\|\cdot\|$ . The set  $\mathcal{G}$  can be understood as the family of candidate action-value functions for this problem. We claim that  $S$  maps  $\mathcal{G}$  into itself and, moreover, is a contraction of modulus  $\beta$  with respect to the supremum norm.

To see that this is so, pick any  $g \in \mathcal{G}$ . Then  $Sg$  is bounded above, since

$$Sg(w, z, c) \leq \beta(\sup u + \|g\|) < \infty.$$

More importantly,  $Sg$  is bounded below. Indeed, using  $g(w', z', c') \geq -\|g\|$  and the monotonicity of  $u$ , we obtain

$$\begin{aligned} Sg(w, z, c) &\geq \beta \mathbb{E}_z \sup_{0 \leq c' \leq R(w-c)+y'} \{u(c') - \|g\|\} \\ &= \beta \mathbb{E}_z \{u(R(w-c) + y') - \|g\|\} \\ &\geq \beta \mathbb{E}_z u(y') - \beta \|g\|. \end{aligned}$$

The last term is finite by (4). Hence  $S$  is a self map on  $\mathcal{G}$ . To show that  $S$  is a contraction mapping, we verify Blackwell (1965)'s sufficient conditions. From (7) we see that  $g_1 \leq g_2$  implies  $Sg_1 \leq Sg_2$ , so monotonicity holds. If  $M \geq 0$  is any constant, then for any  $g \in \mathcal{G}$  we have

$$\begin{aligned} S(g + M)(w, z, c) &= \beta \mathbb{E}_z \sup_{0 \leq c' \leq R(w-c)+y'} \{u(c') + g(R(w-c) + y', z', c') + M\} \\ &= \beta \mathbb{E}_z \sup_{0 \leq c' \leq R(w-c)+y'} \{u(c') + g(R(w-c) + y', z', c')\} + \beta M \\ &= Sg(w, z, c) + \beta M, \end{aligned}$$

so the discounting property holds. We have now shown that  $S$  is a contractive self-map on  $\mathcal{G}$ . Moreover,  $\mathcal{G}$  is a space of bounded functions. By Banach's contraction mapping theorem,  $S$  has a unique fixed point  $g^*$  in  $\mathcal{G}$ .



It is natural to guess that we can now insert  $g^*$  into the right hand side of (6), maximize at each state, and obtain the optimal consumption policy. We show that this conjecture is correct and, more generally, that Bellman's principle of optimality vis-à-vis the transformed Bellman equation also holds. The arguments are not trivial, since the transformation in (5) that maps  $v$  to  $g$  is not bijective. Full details are provided in Section 3.<sup>12</sup>

**2.2. Application 2: Optimal Default.** Consider an infinite horizon optimal savings problem with default, in the spirit of [Arellano \(2008\)](#) and a large related literature.<sup>13</sup> A country with current assets  $w_t$  chooses between continuing to participate in international financial markets and defaulting. Output  $y_t = y(z_t, \xi_t)$  is a function of a Markov process  $\{z_t\}$  and an IID shock  $\{\xi_t\}$ , as in (8). To simplify the exposition, we assume that default leads to permanent exclusion from financial markets, with lifetime value

$$v^d(y, z) = \mathbb{E} \sum_{t=0}^{\infty} \beta^t u(y_t).$$

The utility function  $u$  has the same properties as Section 2.1. The value of continued participation in financial markets is

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

where  $b > 0$  is a constant borrowing constraint and  $v$  is the value function satisfying

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

The function  $v$  is unbounded below because  $u(0) = -\infty$ . However, we can convert this into a bounded problem, as the following analysis shows.

Let  $i$  be a discrete choice variable taking values in  $\{0, 1\}$ , with 0 indicating default and 1 indicating continued participation. We introduce the action-value function

$$g(z, w', i) := \begin{cases} \beta \mathbb{E}_z v^d(y', z') & \text{if } i = 0, \\ \beta \mathbb{E}_z v(w', y', z') & \text{if } i = 1, \end{cases}$$

---

<sup>12</sup>For the savings problem treated above, one can also use Euler equation methods, which circumvent some of the issues associated with unbounded rewards (see, e.g., [Li and Stachurski \(2014\)](#) and [Ma et al. \(2020\)](#)). However, for many practical applications, these Euler equation arguments cannot be used, due to features such as recursive preferences or discrete or multi-dimensional choices (see below). Moreover, our detailed treatment of optimal savings in Section 4.1 shows that, even when Euler equation methods are available, the assumptions needed for the theory in this paper are significantly weaker, at least in some dimensions.

<sup>13</sup>Recent examples include [Aguilar and Amador \(2019\)](#) and [Aguilar et al. \(2019\)](#).

so that for  $-b \leq w' \leq R(w + y)$ , we have

$$v(w, y, z) = \max \left\{ u(y) + g(z, w', 0), \sup_{w'} \{ u(w + y - w'/R) + g(z, w', 1) \} \right\}.$$

Eliminating the value function  $v$  yields

$$g(z, w', 0) = \beta \mathbb{E}_z \{ u(y') + g(z', w', 0) \} \quad \text{and}$$

$$g(z, w', 1) = \beta \mathbb{E}_z \max \left\{ u(y') + g(z', w', 0), \sup_{w''} \{ u(w' + y' - w''/R) + g(z', w'', 1) \} \right\},$$

where  $-b \leq w'' \leq R(w' + y')$ . We can then define the fixed point operator  $S$  corresponding to these functional equations.

If  $g$  is bounded above by some constant  $K$ , then  $Sg \leq \sup_c u(c) + K$ . More importantly, if  $g$  is bounded below by some constant  $M$ , we obtain

$$\begin{aligned} Sg(z, w', 0) &\geq \beta \mathbb{E}_z u(y') + \beta M, \\ Sg(z, w', 1) &\geq \beta \mathbb{E}_z \max \{ u(y') + M, u(w' + y' + b/R) + M \} \\ &= \beta \mathbb{E}_z \max \{ u(y'), u(w' + y' + b/R) \} + \beta M. \end{aligned}$$

Hence,  $Sg$  is bounded below by a finite constant if (4) holds. An argument similar to the one in Section 2.1 now proves that  $S$  is a contraction with respect to the supremum norm. (Section 4.2 gives details.)

**2.3. Application 3: Job Search.** Following McCall (1970), consider a search problem where an unemployed worker can either accept the current job offer and work at that wage forever or choose an outside option (e.g., work in the informal sector) and continue to the next period. Letting  $z_t$  be the worker's productivity at time  $t$ , which is a Markov process, the job offer  $w_t$  and outside option  $c_t$  satisfy

$$w_t = w(z_t, \xi_t) \quad \text{and} \quad c_t = c(z_t, \xi_t), \tag{8}$$

where  $w, c$  are nonnegative measurable functions and  $\xi_t$  is an IID shock that could be vector-valued.<sup>14</sup> Letting  $u$  be the utility function and  $\beta \in (0, 1)$  be the discount factor, a worker that accepts a job offer  $w$  enjoys lifetime utility  $\sum_{t=0}^{\infty} \beta^t u(w) = \frac{u(w)}{1-\beta}$ . Therefore, the worker's value function satisfies the Bellman equation

$$v(w, c, z) = \max \left\{ \frac{u(w)}{1-\beta}, u(c) + \beta \mathbb{E}_z v(w', c', z') \right\}. \tag{9}$$

---

<sup>14</sup>For instance, if the job offer  $w_t$  and outside option  $c_t$  are independent conditional on  $z_t$ , then we may write  $\xi = (\xi_1, \xi_2)$ , where  $\xi_1$  and  $\xi_2$  are independent,  $w(z, \xi)$  depends only on  $z$  and  $\xi_1$ , and  $c(z, \xi)$  depends only on  $z$  and  $\xi_2$ .

For now, let  $u$  be bounded above. In addition, analogous to (4), assume

$$\text{either } \inf_z \mathbb{E}_z u(w') > -\infty \text{ or } \inf_z \mathbb{E}_z u(c') > -\infty. \quad (10)$$

The value function  $v(w, c, z)$  is unbounded below, if, say  $u$  is CRRA as in (1) with  $\gamma > 1$  and the job offer and outside option in (8) can be arbitrarily small. To shift to a bounded problem, we can proceed in a similar vein to our manipulation of the Bellman equation in the optimal savings case. First we set

$$g(z) := \beta \mathbb{E}_z v(w', c', z'),$$

so that (9) can be written as

$$v(w, c, z) = \max \left\{ \frac{u(w)}{1 - \beta}, u(c) + g(z) \right\}.$$

Next we use the definition of  $g$  to eliminate  $v$  from this last expression, which leads to the functional equation

$$g(z) = \beta \mathbb{E}_z \max \left\{ \frac{u(w')}{1 - \beta}, u(c') + g(z') \right\}. \quad (11)$$

Let  $S$  be an operator such that  $Sg(z)$  is equal to the right hand side of (11). It is clear that if  $g$  is bounded above then so is  $Sg$ . In addition, if  $g$  is bounded below then so is  $Sg$ . To show this, using the elementary bound

$$\mathbb{E} \max \{X, Y\} \geq \max \{\mathbb{E}X, \mathbb{E}Y\} \quad (12)$$

for arbitrary random variables  $X, Y$  and  $g \geq -\|g\|$ , we have

$$Sg(z) \geq \beta \max \left\{ \mathbb{E}_z \frac{u(w')}{1 - \beta}, \mathbb{E}_z u(c') - \|g\| \right\}.$$

Condition (10) then implies that  $Sg$  is also bounded below.

An argument similar to the one adopted in Sections 2.1–2.2 show that  $S$  is a contraction mapping with respect to the supremum norm on a space of bounded functions. Thus, we can proceed down the same path to establish optimality.

### 3. GENERAL FORMULATION

Section 2 showed how some unbounded problems can be converted to bounded problems by transforming the Bellman equation. The next step is to confirm the validity of such a transformation in terms of the connection between the transformed Bellman equation and optimal policies. We do this in a generic dynamic programming setting that contains the applications given above.

**3.1. Transformed Bellman Equation.** For a given set  $E$ , let  $\mathcal{B}(E)$  be the Borel subsets of  $E$ . For our purpose, a dynamic program consists of

- a nonempty set  $\mathsf{X}$  called the *state space*,
- a nonempty set  $\mathsf{A}$  called the *action space*,
- a nonempty correspondence  $\Gamma : \mathsf{X} \rightrightarrows \mathsf{A}$  called the *feasible correspondence*, along with the associated set of *state action pairs*

$$\mathsf{D} := \{(x, a) \in \mathsf{X} \times \mathsf{A} : a \in \Gamma(x)\},$$

- a measurable map  $r : \mathsf{D} \rightarrow \mathbb{R} \cup \{-\infty\}$  called the *reward function*,
- a constant  $\beta \in (0, 1)$  called the *discount factor*, and
- a *stochastic kernel*  $Q$  governing the evolution of states.<sup>15</sup>

Each period, an agent observes a state  $x_t \in \mathsf{X}$  and responds with an action  $a_t \in \Gamma(x_t) \subset \mathsf{A}$ . The agent then obtains a reward  $r(x_t, a_t)$ , moves to the next period with a new state  $x_{t+1}$ , and repeats the process by choosing  $a_{t+1}$  and so on. The state process updates according to  $x_{t+1} \sim Q(x_t, a_t, \cdot)$ .

Let  $\Sigma$  denote the set of *feasible policies*, which we assume to be nonempty and define as all measurable maps  $\sigma : \mathsf{X} \rightarrow \mathsf{A}$  satisfying  $\sigma(x) \in \Gamma(x)$  for all  $x \in \mathsf{X}$ .<sup>16</sup> Given any policy  $\sigma \in \Sigma$  and initial state  $x_0 = x \in \mathsf{X}$ , the  $\sigma$ -*value function*  $v_\sigma$  is defined by

$$v_\sigma(x) = \sum_{t=0}^{\infty} \beta^t \mathbb{E}_x r(x_t, \sigma(x_t)) \quad (13)$$

whenever the expectation and infinite sum are well-defined. We understand  $v_\sigma(x)$  as the lifetime value of following policy  $\sigma$  now and forever, starting from current state  $x$ .

The *value function* associated with this dynamic program is defined at each  $x \in \mathsf{X}$  by

$$v^*(x) = \sup_{\sigma \in \Sigma} v_\sigma(x). \quad (14)$$

A feasible policy  $\sigma^*$  is called *optimal* if  $v_{\sigma^*} = v^*$  on  $\mathsf{X}$ . The objective of the agent is to find an optimal policy that attains the maximum lifetime value.

<sup>15</sup>Here a *stochastic kernel* corresponding to our controlled Markov process  $\{(x_t, a_t)\}$  is a mapping  $Q : \mathsf{D} \times \mathcal{B}(\mathsf{X}) \rightarrow [0, 1]$  such that (1) for each  $(x, a) \in \mathsf{D}$ ,  $A \mapsto Q(x, a, A)$  is a probability measure on  $\mathcal{B}(\mathsf{X})$ , and (2) for each  $A \in \mathcal{B}(\mathsf{X})$ ,  $(x, a) \mapsto Q(x, a, A)$  is a measurable function.

<sup>16</sup>We can and do focus on stationary Markov policies in what follows since the value of any nonstationary policy can be obtained by a stationary Markov policy. See, e.g., Bertsekas (2018, Section 2.1).

The Bellman equation associated with the dynamic program is

$$v(x) = \sup_{a \in \Gamma(x)} \{r(x, a) + \beta \mathbb{E}_{x,a} v(x')\}, \quad (15)$$

where  $\mathbb{E}_{x,a}$  denotes the expectation with respect to the probability measure  $Q(x, a, \cdot)$ . As in the examples in Section 2, letting  $g$  be the action-value function defined by  $g(x, a) := \beta \mathbb{E}_{x,a} v(x')$ , we obtain

$$v(x) = \sup_{a \in \Gamma(x)} \{r(x, a) + g(x, a)\}.$$

Changing  $(x, a)$  to  $(x', a')$ , multiplying both sides by  $\beta$ , taking the conditional expectation with respect to  $(x, a)$ , and using the definition of  $g$ , we obtain the transformed Bellman equation

$$g(x, a) = \beta \mathbb{E}_{x,a} \sup_{a' \in \Gamma(x')} \{r(x', a') + g(x', a')\}. \quad (16)$$

Motivated by this derivation, given a real-valued measurable function  $g$  on  $\mathbf{D}$ , we define the *transformed Bellman operator*  $S$  by

$$Sg(x, a) := \beta \mathbb{E}_{x,a} \sup_{a' \in \Gamma(x')} \{r(x', a') + g(x', a')\}. \quad (17)$$

A feasible policy  $\sigma$  is called  *$g$ -greedy* if

$$\sigma(x) \in \operatorname{argmax}_{a \in \Gamma(x)} \{r(x, a) + g(x, a)\} \quad \text{for all } x \in \mathbf{X}. \quad (18)$$

At each  $x \in \mathbf{X}$  and  $(x, a) \in \mathbf{D}$ , we define

$$\bar{r}(x) := \sup_{a \in \Gamma(x)} r(x, a) \quad \text{and} \quad \hat{r}(x, a) := \mathbb{E}_{x,a} \bar{r}(x'). \quad (19)$$

The function  $\bar{r}$  can be interpreted as the maximum reward given the current state  $x \in \mathbf{X}$ . The function  $\hat{r}$  can be interpreted as its expectation conditional on the previous state and action. We make the following assumption.

**Assumption 3.1.** The function  $\bar{r}$  in (19) is bounded above and  $\hat{r}$  is bounded below.

Let  $\mathcal{G}$  be the set of bounded measurable functions on  $\mathbf{D}$  and  $\|\cdot\|$  be the supremum norm. Despite of the potentially unbounded below rewards, the following result illustrates that  $S$  maps elements of  $\mathcal{G}$  into itself and the dynamic program can be solved via the standard contraction mapping theorem.

**Theorem 3.1.** *If Assumption 3.1 holds, then  $v^*$  in (14) is well-defined,*

- (1)  $S\mathcal{G} \subset \mathcal{G}$  and  $S$  is a contraction mapping on  $(\mathcal{G}, \|\cdot\|)$ ,

- (2)  $S$  admits a unique fixed point  $g^*$  in  $\mathcal{G}$ , and
- (3)  $S^k g$  converges to  $g^*$  at rate  $O(\beta^k)$  under  $\|\cdot\|$ .

Moreover, if there exists a closed subset  $\mathcal{G}_1$  of  $\mathcal{G}$  such that  $S\mathcal{G}_1 \subset \mathcal{G}_1$  and a  $g$ -greedy policy exists for each  $g \in \mathcal{G}_1$ , then

- (a)  $g^*$  is an element of  $\mathcal{G}_1$  and satisfies

$$g^*(x, a) = \beta \mathbb{E}_{x,a} v^*(x') \quad \text{and} \quad v^*(x) = \sup_{a \in \Gamma(x)} \{r(x, a) + g^*(x, a)\},$$

- (b) at least one optimal policy exists, and
- (c) a feasible policy is optimal if and only if it is  $g^*$ -greedy.

**3.2. Existence of Optimal Policy.** Theorem 3.1 states that under Assumption 3.1, which is satisfied in many applications, the transformed Bellman operator  $S$  is a contraction. However, it requires a high-level assumption to guarantee that a solution to the dynamic program exists.

We now discuss some general sufficient conditions for parts (a)–(c) of Theorems 3.1 to hold. To this end, we introduce an additional assumption.

**Assumption 3.2.** (1) The sets  $\mathbf{X}$  and  $\mathbf{A}$  are complete separable metric spaces, (2) the reward function  $r$  is upper semicontinuous, (3) the feasible correspondence  $\Gamma$  is compact-valued and upper hemicontinuous,<sup>17</sup> and (4) the stochastic kernel  $Q$  is Feller.<sup>18</sup>

In most applications Assumption 3.2 is satisfied.

Let  $\mathcal{G}_1$  be the set of upper semicontinuous functions in  $\mathcal{G}$ . The following theorem shows that the conclusions of Theorem 3.1 hold.

**Theorem 3.2.** *If Assumptions 3.1 and 3.2 hold, then  $\mathcal{G}_1$  is a closed subset of  $\mathcal{G}$ ,  $S\mathcal{G}_1 \subset \mathcal{G}_1$ , and a  $g$ -greedy policy exists for each  $g \in \mathcal{G}_1$ . Consequently, all the conclusions of Theorem 3.1 hold and  $g^*, v^*$  are upper semicontinuous.*

## 4. APPLICATIONS

Now we complete the discussion of all applications in Section 2. We also provide additional applications to optimal savings with endogenous labor choice and optimal consumption-portfolio choice.

<sup>17</sup>In other words, the set  $\{x \in \mathbf{X} : \Gamma(x) \subset U\}$  is open for each open subset  $U \subset \mathbf{A}$ . See Aliprantis and Border (2006, Lemma 17.4) for alternative characterizations of upper hemicontinuity.

<sup>18</sup>In other words,  $(x, a) \mapsto \int h(x')Q(x, a, dx')$  is bounded and continuous whenever  $h$  is.

**4.1. Optimal Savings (Continued).** Recall the optimal savings problem of Section 2.1. Following the setting in [Ma et al. \(2020\)](#), we allow for capital income risk in a Markov environment. The agent seeks to solve (2), except that the return  $R = R_{t+1}$  can also be stochastic.<sup>19</sup> For concreteness, suppose that

$$R_t = R(z_t, \xi_t) \quad \text{and} \quad y_t = y(z_t, \xi_t), \quad (20)$$

where  $R, y$  are nonnegative measurable functions,  $z_t$  is a finite state Markov chain, and  $\xi_t$  is an IID shock that could be vector-valued.

To apply the general theory in Section 3, we assume that the utility function  $u$  is upper semicontinuous, increasing, bounded above, and

$$\inf_z \mathbb{E}u(y(z, \xi)) > -\infty. \quad (21)$$

Let us verify that the assumptions in Section 3 are satisfied. The state  $x = (w, z)$  consists of the financial wealth  $w$  and the exogenous Markov state  $z$ . The action is consumption  $a = c$ . The feasible correspondence is  $\Gamma(x) = [0, w]$ , which is the borrowing constraint (2b). The reward function is  $r(x, a) = u(c)$ . The stochastic kernel  $Q$  is defined through the (exogenous) stochastic kernel  $P$  of the Markov state  $z$ , the distribution of the IID shock  $\xi$ , and the budget constraint (2c). The functions  $\bar{r}$  and  $\hat{r}$  in (19) are defined by

$$\begin{aligned} \bar{r}(x) &= \sup_{0 \leq c \leq w} u(c) = u(w), \\ \hat{r}(x, a) &= \mathbb{E}_z u(R'(w - c) + y') \\ &\geq \mathbb{E}_z u(y') = \mathbb{E}_z u(y(z', \xi')) > -\infty, \end{aligned}$$

where we have used the monotonicity of  $u$  and (21). Since by assumption  $u$  is bounded above, so is  $\bar{r}$ . Therefore, Assumption 3.1 holds. Since  $u$  is upper semicontinuous,  $\Gamma$  is nonempty compact valued, and  $\{z_t\}$  is a finite state Markov chain, Assumption 3.2 is also satisfied. Therefore, the conclusions of Theorem 3.1 hold.

**Remark 4.1.** [Ma et al. \(2020\)](#) (henceforth MST) solve the optimal savings problem using the Euler equation iteration. Our approach is different because it uses the (transformed) value function iteration under different assumptions. While we require

---

<sup>19</sup>The importance of capital income risk for wealth dynamics is highlighted in [Toda \(2014\)](#), [Benhabib et al. \(2015\)](#), [Cao and Luo \(2017\)](#), [Stachurski and Toda \(2019\)](#), [Fagereng et al. \(2020\)](#) and [Hubmer et al. \(2020\)](#), among others.

that the utility function is bounded above, MST does not require it. On the other hand, MST requires the utility function to be concave, differentiable, and satisfy

$$\sup_z \mathbb{E}_z u'(y) < \infty. \quad (22)$$

The following argument shows that our assumptions are weaker.<sup>20</sup> Since  $u$  is concave and differentiable under the assumptions of MST, we obtain

$$u(1) - u(y) \leq u'(y)(1 - y) \leq u'(y),$$

where we have used  $u' \geq 0$  and  $y \geq 0$ . Taking the conditional expectation on  $z$ , we obtain

$$u(1) - \mathbb{E}_z u(y) \leq \mathbb{E}_z u'(y) < \infty$$

by (22), implying (21).

More importantly, MST requires the condition  $G_{\beta R} < 1$ , where  $G_{\beta R}$  is the long run geometric average of  $\beta R_t$ . Using our approach, we do not require any assumption (other than nonnegativity and measurability) on the returns  $R_t$ .

**4.2. Optimal Default (Continued).** Recall the optimal default problem studied in Section 2.2. This setting is a special case of our framework. In particular,

$$x := (w, y, z), \quad a := (w', i), \quad \mathbf{X} := [-b, \infty) \times \mathbf{Y} \times \mathbf{Z} \quad \text{and} \quad \mathbf{A} := [-b, \infty) \times \{0, 1\},$$

where  $i$  is a discrete choice variable taking values in  $\{0, 1\}$ , and  $\mathbf{Y}$  and  $\mathbf{Z}$  are respectively the range spaces of  $\{y_t\}$  and  $\{z_t\}$ . The reward function  $r$  reduces to

$$r(w, y, w', i) := \begin{cases} u(y) & \text{if } i = 0, \\ u(w + y - w'/R) & \text{if } i = 1. \end{cases}$$

We have shown that  $S\mathcal{G} \subset \mathcal{G}$ , where  $\mathcal{G}$  is the set of bounded measurable functions on  $\mathbf{Z} \times [-b, \infty) \times \{0, 1\}$ . Moreover,  $\hat{r}$  satisfies

$$\hat{r}(z, w') = \mathbb{E}_z \max \{u(y'), u(w' + y' + b/R)\} \geq \mathbb{E}_z u(y'),$$

which is bounded below by (4). Let  $\mathcal{G}_1$  be the set of functions in  $\mathcal{G}$  that is increasing in its second-to-last argument and upper semicontinuous. Through similar steps to the proof of Theorem 3.2, one can show that  $S\mathcal{G}_1 \subset \mathcal{G}_1$  and a  $g$ -greedy policy exists for each  $g \in \mathcal{G}_1$ . As a result, all the conclusions of Theorem 3.1 are true.

---

<sup>20</sup>Ma et al. (2020) allow the discount factor  $\beta$  to be random. It is straightforward to extend our theory to a setting with stochastic discounting.



**4.3. Job Search (Continued).** Recall the job search problem of Section 2.3. This problem fits into the framework of Section 3.1 if we let choice  $a$  take values in  $\{0, 1\}$ , where 0 represents the decision to stop and 1 represents continue,

$$x := (w, z, c), \quad \mathbf{X} := (0, \infty)^3, \quad \mathbf{A} := \{0, 1\}, \quad \Gamma(x) := \{0, 1\}, \quad \mathbf{D} := (0, \infty)^3 \times \{0, 1\}$$

and the reward function is  $r(x, a) := u(w)/(1 - \beta)$  if  $a = 0$  and  $r(x, a) := u(c)$  if  $a = 1$ . We have shown that  $S\mathcal{G} \subset \mathcal{G}$ , where  $\mathcal{G}$  is the set of bounded measurable functions on  $(0, \infty)$ . Note that, in this case, the function  $\hat{r}(x, a)$  reduces to  $\hat{r}(z) = \mathbb{E}_z \max \{u(w)/(1 - \beta), u(c')\}$ . Then  $\hat{r}$  is bounded below by the inequality (12) and (10). Since in addition the action set is finite, a  $g$ -greedy policy always exists for each  $g \in \mathcal{G}$ . Let  $\mathcal{G}_1 := \mathcal{G}$ . The analysis above implies that all the conclusions of Theorem 3.1 hold.

**4.4. Optimal Savings with Endogenous Labor Choice.** As another example application, consider the optimal savings problem with endogenous labor supply

$$\begin{aligned} & \text{maximize} && \mathbb{E} \sum_{t=0}^{\infty} \beta^t u(c_t, l_t) \\ & \text{subject to} && 0 \leq c_t \leq w_t + y_t l_t, \\ & && 0 \leq l_t \leq 1, \\ & && w_{t+1} = R_{t+1}(w_t - c_t + y_t l_t). \end{aligned}$$

Here  $c_t$  is consumption,  $l_t$  is labor supply,  $y_t$  is wage,  $w_t$  is financial wealth at time  $t$  excluding current labor income (see Footnote 9), and  $R_{t+1} \geq 0$  is the gross return on wealth between time  $t$  and  $t + 1$ . As before assume that  $R, y$  take the form (20), where  $z_t$  is a finite state Markov chain and  $\xi_t$  is an IID shock.

The state  $x = (w, y, z)$  consists of the financial wealth  $w$ , wage  $y$ , and the exogenous Markov state  $z$ . The action  $a = (c, l)$  consists of consumption and labor supply. The feasible correspondence is

$$\Gamma(x) = \{(c, l) \in \mathbb{R}^2 : 0 \leq c \leq w + yl \text{ and } 0 \leq l \leq 1\}.$$

Suppose that the utility function  $u$  is bounded above and increasing in its first argument. The function  $\bar{r}$  in (19) is defined by

$$\bar{r}(x) = \sup_{0 \leq l \leq 1} \sup_{0 \leq c \leq w + yl} u(c, l) = \sup_{0 \leq l \leq 1} u(w + yl, l),$$

which is bounded above. Noting we can bound  $\bar{r}$  from below as

$$\bar{r}(x) = \sup_{0 \leq l \leq 1} u(w + yl, l) \geq \sup_{0 \leq l \leq 1} u(yl, l),$$

the function  $\hat{r}$  in (19) becomes bounded below if

$$\inf_z \mathbb{E} \sup_{0 \leq l \leq 1} u(y(z, \xi)l, l) > -\infty, \quad (23)$$

which is analogous to (21). In summary, by a similar argument to Section 4.1, the conclusions of Theorem 3.1 hold if  $u$  is upper semicontinuous, bounded above, increasing in its first argument, and (23) holds.

**4.5. Optimal Consumption-Portfolio Problem.** As yet another example application, consider the optimal consumption-portfolio problem

$$\begin{aligned} & \text{maximize} && \mathbb{E} \sum_{t=0}^{\infty} \beta^t u(c_t) \\ & \text{subject to} && 0 \leq c_t \leq w_t, \\ & && \theta_t \in \Theta(z_t), \\ & && w_{t+1} = R(\theta_t, z_{t+1}, \xi_{t+1})(w_t - c_t) + y_{t+1}. \end{aligned}$$

Here  $c_t$  is consumption,  $z_t$  is an exogenous finite state Markov chain,  $\Theta(z_t) \subset \mathbb{R}^J$  is the set of admissible portfolios of financial assets  $j = 1, \dots, J$  in state  $z_t$  ( $\theta_t$  is a portfolio),  $y_t = y(z_t, \xi_t)$  is non-financial income ( $\xi_t$  is an IID shock that could be vector-valued),  $w_t$  is financial wealth at time  $t$  including current non-financial income, and  $R(\theta_t, z_{t+1}, \xi_{t+1})$  is the gross return on wealth between time  $t$  and  $t+1$  given the portfolio  $\theta_t$  and shocks  $(z_{t+1}, \xi_{t+1})$ .

This problem is a special case of our framework. The state  $x = (w, z)$  consists of the financial wealth  $w$  and the exogenous Markov state  $z$ . The action  $a = (c, \theta)$  consists of consumption and portfolio. The feasible correspondence is  $\Gamma(x) = [0, w] \times \Theta(z)$ . By the same argument as in Section 4.1, if the utility function  $u$  is increasing, bounded above, and (21) holds, then so does Assumption 3.1. Under additional regularity conditions ( $u$  is upper semicontinuous and the portfolio constraint  $\Theta(z)$  is nonempty and compact for each  $z$ ), Assumption 3.2 is satisfied and the conclusions of Theorem 3.1 hold.

## 5. EXTENSIONS

In this section, we extend our theory in two important directions. First, we illustrate how the idea of Bellman transformation could be extended to handle rewards that are potentially unbounded above as well as below. Second, we extend the theory of Section 3 to solve dynamic programs with risk-sensitive preferences.

**5.1. Unbounded Above Rewards.** In Section 3, we assume that the reward function is bounded above, although it could be unbounded below. To handle rewards that are potentially unbounded above and below, we extend our theory by introducing a weighting function  $\kappa$ , which is a continuous function mapping  $\mathbf{X}$  to  $[1, \infty)$ . Let  $\mathcal{G}$  be the set of measurable functions  $g : \mathbf{D} \rightarrow \mathbb{R}$  such that  $g$  is bounded below and

$$\|g\|_\kappa := \sup_{(x,a) \in \mathbf{D}} \frac{|g(x,a)|}{\kappa(x)} < \infty.$$

The pair  $(\mathcal{G}, \|\cdot\|_\kappa)$  is a Banach space (see, e.g., Bertsekas (2018)). We make the following assumption.

**Assumption 5.1.** (1) There exist constants  $d \in \mathbb{R}_+$  and  $\alpha \in (0, 1/\beta)$  such that  $\bar{r}(x) \leq d\kappa(x)$  and  $\mathbb{E}_{x,a}\kappa(x') \leq \alpha\kappa(x)$  for all  $(x,a) \in \mathbf{D}$ , and (2)  $\hat{r}$  in (19) is bounded below.

Note that Assumption 3.1 is a special case of Assumption 5.1 by setting  $\kappa(x) \equiv 1$  and  $\alpha = 1$ . Assumption 5.1 relaxes the standard weighted supremum norm assumptions (see, e.g., Wessels (1977) or Bertsekas (2018)), in the sense that ratio of the reward function to the weighting function is allowed to be unbounded from below.

Although rewards are potentially unbounded above and below, the dynamic program can be solved by the operator  $S$ , as the following theorem shows.

**Theorem 5.1.** *If Assumption 5.1 holds, then  $v^*$  in (14) is well-defined,*

- (1)  $S\mathcal{G} \subset \mathcal{G}$  and  $S$  is a contraction mapping on  $(\mathcal{G}, \|\cdot\|_\kappa)$ ,
- (2)  $S$  admits a unique fixed point  $g^*$  in  $\mathcal{G}$ , and
- (3)  $S^k g$  converges to  $g^*$  at rate  $O((\alpha\beta)^k)$  under  $\|\cdot\|_\kappa$ .

Moreover, if there exists a closed subset  $\mathcal{G}_1$  of  $\mathcal{G}$  such that  $S\mathcal{G}_1 \subset \mathcal{G}_1$  and a  $g$ -greedy policy exists for each  $g \in \mathcal{G}_1$ , then

(a)  $g^*$  is an element of  $\mathcal{G}_1$  and satisfies

$$g^*(x, a) = \beta \mathbb{E}_{x,a} v^*(x') \quad \text{and} \quad v^*(x) = \sup_{a \in \Gamma(x)} \{r(x, a) + g^*(x, a)\},$$

(b) at least one optimal policy exists, and

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

Let  $\mathcal{G}_1$  be the set of upper semicontinuous functions in  $\mathcal{G}$  and

$$\hat{\kappa}(x, a) := \mathbb{E}_{x,a} \kappa(x'). \quad (24)$$

In most applications Assumption 3.2 is satisfied. The following theorem shows that the continuity of  $\hat{\kappa}$  is sufficient for the conclusions of Theorem 5.1 to hold.

**Theorem 5.2.** *If Assumptions 5.1 and 3.2 hold and  $\hat{\kappa}$  in (24) is continuous, then  $\mathcal{G}_1$  is a closed subset of  $\mathcal{G}$ ,  $S\mathcal{G}_1 \subset \mathcal{G}_1$ , and a  $g$ -greedy policy exists for each  $g \in \mathcal{G}_1$ . Consequently, all the conclusions of Theorem 5.1 hold and  $g^*, v^*$  are upper semicontinuous.*

As an example application of Theorem 5.2, consider the optimal savings problem (2), where the utility function  $u$  can now be unbounded both from above and below. Suppose that  $u$  is upper semicontinuous, increasing, satisfies (21), and there exist constants  $p > 0$  and  $q \in \mathbb{R}$  such that

$$u(c) \leq pc + q \quad \text{for all } c > 0. \quad (25)$$

This condition trivially holds if  $u$  is concave, and we can choose  $q$  arbitrarily large.

Suppose that asset return and income take the form

$$R_t = R(\xi_t) \quad \text{and} \quad y_t = y(z_t, \xi_t), \quad (26)$$

where  $R, y$  are nonnegative measurable functions,  $z_t$  is a finite state Markov chain, and  $\xi_t$  is an IID shock that could be vector-valued.<sup>21</sup> In addition, assume

$$Y := \sup_z \mathbb{E}_z y(z', \xi') < \infty, \quad \beta < 1, \quad \text{and} \quad \beta \mathbb{E} R < 1. \quad (27)$$

As in Section 4.1, the state is  $x = (w, z)$  and the action is  $a = c$ . To apply Theorem 5.2, define the weighting function by  $\kappa(x) = pw + q$ , where  $q > 1$ . Since

$$\bar{r}(x) = \sup_{0 \leq c \leq w} u(c) = u(w) \leq pw + q = \kappa(x)$$

---

<sup>21</sup>Unlike the setting in Section 4.1, the return is permitted to depend only on the IID shock  $\xi$ . Treating the general case requires generalizing Theorem 5.1 further such that  $\alpha$  depends on  $x$ .

by (25), we can set  $d = 1$  in Assumption 5.1. As we have seen in Section 4.1, the condition (21) implies that  $\hat{r}$  in (19) is bounded below. Therefore, to satisfy Assumption 5.1, it remains to verify  $\mathbb{E}_{x,a}\kappa(x') \leq \alpha\kappa(x)$  for some  $\alpha \in (0, 1/\beta)$ . To this end, note that

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

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

$$\sup_{(x,a) \in \mathcal{D}} \frac{\mathbb{E}_{x,a}\kappa(x')}{\kappa(x)} \leq \sup_z \max \left\{ \frac{p\mathbb{E}_zy' + q}{q}, \mathbb{E}_zR' \right\} \leq \max \left\{ \frac{pY + q}{q}, \mathbb{E}R \right\}, \quad (28)$$

where we have used the fact that  $R$  does not depend on  $z$  (by (26)) and (27). Since  $q > 1$  can be taken arbitrarily large and  $\frac{pY+q}{q} \rightarrow 1$  as  $q \rightarrow \infty$ , the right hand side of (28) can be made arbitrarily close to  $\max\{1, \mathbb{E}R\}$ , which is strictly smaller than  $1/\beta$  by (27). Therefore, we can indeed choose  $\alpha \in (0, 1/\beta)$  such that Assumption 5.1 holds with  $d = 1$  and  $\kappa(x) = pw + q$  for large enough  $q > 1$ . Assumption 3.2 trivially holds. Finally,

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

is clearly continuous in  $(x, a) = (w, z, c)$ . Therefore, all the assumptions of Theorem 5.2 are satisfied.

**Remark 5.1.** Under the assumption that discounting is constant and the asset return depends only on the IID shock as in (26), the assumptions in Ma et al. (2020) are strictly stronger than ours, since they assume (27) and the concavity of  $u$  (which implies (25)). (See also Remark 4.1.)

**5.2. Risk-Sensitive Preferences.** We consider the general setting in Sections 3. Unlike the additively separable case, in order to define the value function and optimality in the recursive case, let  $\sigma \in \Sigma$  be any feasible policy and  $\gamma > 0$  be the agent's risk-sensitive coefficient. A measurable function  $v_\sigma : \mathcal{X} \rightarrow \mathbb{R} \cup \{-\infty\}$  is called a  $\sigma$ -value function if

$$v_\sigma(x) = r(x, \sigma(x)) - \frac{\beta}{\gamma} \log \mathbb{E}_{x, \sigma(x)} e^{-\gamma v_\sigma(x')}$$

for all  $x \in \mathcal{X}$ . Setting aside the issue of existence and uniqueness for now,  $v_\sigma(x)$  can be interpreted as the lifetime value of following policy  $\sigma$  forever, starting from current state  $x$ . The value function  $v^*$  and optimal policy  $\sigma^*$  are defined as in (14). The

Bellman equation associated with the dynamic program with risk-sensitive coefficient  $\gamma > 0$  is

$$v(x) = \sup_{a \in \Gamma(x)} \left\{ r(x, a) - \frac{\beta}{\gamma} \log \mathbb{E}_{x,a} e^{-\gamma v(x')} \right\}. \quad (29)$$

Letting  $g(x, a) := -\frac{\beta}{\gamma} \log \mathbb{E}_{x,a} e^{-\gamma v(x')}$ , analogous to the derivation of (16), we obtain the transformed Bellman equation

$$g(x, a) = -\frac{\beta}{\gamma} \log \mathbb{E}_{x,a} \exp \left( -\gamma \sup_{a' \in \Gamma(x')} \{ r(x', a') + g(x', a') \} \right). \quad (30)$$

As in Section 3, we consider the case with rewards that are bounded above. Let  $\mathcal{G}$  be the set of bounded measurable functions on  $\mathbf{D}$  and  $\|\cdot\|$  be the supremum norm. We define the *transformed Bellman operator*  $S$  on  $\mathcal{G}$  by letting  $Sg(x, a)$  be the right hand side of (30) for each  $g \in \mathcal{G}$ . A feasible policy  $\sigma \in \Sigma$  is called  *$g$ -greedy* if (18) holds. The following theorem generalizes Theorems 3.1 and 3.2 to dynamic programs with risk-sensitive preferences.

**Theorem 5.3.** *If Assumption 3.1 holds for  $\hat{r}$  defined by*

$$\hat{r}(x, a) := -\frac{1}{\gamma} \log \mathbb{E}_{x,a} e^{-\gamma \bar{r}(x')}, \quad (31)$$

- (1)  $S\mathcal{G} \subset \mathcal{G}$  and  $S$  is a contraction mapping on  $(\mathcal{G}, \|\cdot\|)$ ,
- (2)  $S$  admits a unique fixed point  $g^*$  in  $\mathcal{G}$ , and
- (3)  $S^k g$  converges to  $g^*$  at rate  $O(\beta^k)$  under  $\|\cdot\|$ .

Moreover, if Assumption 3.2 holds, then  $v^*$  is well-defined and

- (a)  $g^*, v^*$  are upper semicontinuous and satisfy

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

- (b) at least one optimal policy exists, and
- (c) a feasible policy is optimal if and only if it is  $g^*$ -greedy.

## 6. PROOFS

Since Theorems 3.1, 3.2 are special cases of Theorems 5.1, 5.2 by setting  $\kappa(x) \equiv 1$  and  $\alpha = 1$ , we only prove the latter.

We first show that the  $\sigma$ -value function  $v_\sigma$  in (13) and the value function  $v^*$  in (14) are well-defined.

**Lemma 6.1.** *If Assumption 5.1(1) holds, then for any feasible policy  $\sigma \in \Sigma$  and initial state  $x_0 = x \in \mathbf{X}$ , the quantities  $v_\sigma(x)$  in (13) and  $v^*(x)$  in (14) are well-defined in  $\mathbb{R} \cup \{-\infty\}$ .*

*Proof.* Using the definition of  $\bar{r}$  in (19) and Assumption 5.1(1), the  $t$ -th term on the right hand side of (13) can be bounded above as

$$\beta^t \mathbb{E}_x r(x_t, \sigma(x_t)) \leq \beta^t \mathbb{E}_x \bar{r}(x_t) \leq \beta^t \mathbb{E}_x d\kappa(x_t) \leq \beta^t \alpha^t d\kappa(x) = d\kappa(x)(\alpha\beta)^t.$$

Since by assumption  $0 < \alpha\beta < 1$ , summing over  $t$ , we obtain

$$v_\sigma(x) = \sum_{t=0}^{\infty} \beta^t \mathbb{E}_x r(x_t, \sigma(x_t)) \leq \sum_{t=0}^{\infty} d\kappa(x)(\alpha\beta)^t = \frac{d\kappa(x)}{1 - \alpha\beta} < \infty.$$

Therefore,  $v_\sigma(x)$  in (13) is well-defined in  $\mathbb{R} \cup \{-\infty\}$ . Taking the supremum over  $\sigma \in \Sigma$ , we obtain

$$v^*(x) = \sup_{\sigma \in \Sigma} v_\sigma(x) \in \left[ -\infty, \frac{d\kappa(x)}{1 - \alpha\beta} \right],$$

so  $v^*(x)$  in (14) is also well-defined.  $\square$

*Proof of Theorem 5.1.* To see claim (1) holds, we first show that  $S\mathcal{G} \subset \mathcal{G}$ . Fix  $g \in \mathcal{G}$ . By the definition of  $\mathcal{G}$ , there is a lower bound  $\underline{g} \in \mathbb{R}$  such that  $g \geq \underline{g}$ . Then

$$\begin{aligned} Sg(x, a) &\geq \beta \mathbb{E}_{x,a} \sup_{a' \in \Gamma(x')} \{r(x', a') + \underline{g}\} = \beta \left[ \mathbb{E}_{x,a} \sup_{a' \in \Gamma(x')} r(x', a') + \underline{g} \right] \\ &= \beta [\mathbb{E}_{x,a} \bar{r}(x') + \underline{g}] = \beta [\hat{r}(x, a) + \underline{g}]. \end{aligned}$$

Since by assumption  $\hat{r}$  is bounded below, so is  $Sg$ . Moreover, by Assumption 5.1,

$$\begin{aligned} Sg(x, a) &\leq \beta \mathbb{E}_{x,a} \left\{ \bar{r}(x') + \sup_{a' \in \Gamma(x')} g(x', a') \right\} \\ &\leq \beta \mathbb{E}_{x,a} (d + \|g\|_\kappa) \kappa(x') \leq \alpha\beta (d + \|g\|_\kappa) \kappa(x) \end{aligned}$$

for all  $(x, a) \in \mathbf{D}$ . Hence,  $Sg/\kappa$  is bounded above. Since in addition  $Sg$  is bounded below and  $\kappa \geq 1$ , we have  $\|Sg\|_\kappa < \infty$ , implying  $Sg \in \mathcal{G}$ .

Obviously,  $S$  is a monotone operator, i.e.,  $Sg_1 \leq Sg_2$  whenever  $g_1 \leq g_2$ . To see that  $S$  is a contraction mapping on  $(\mathcal{G}, \|\cdot\|_\kappa)$  of modulus  $\alpha\beta$ , it suffices to show that<sup>22</sup>

$$S(g + M\kappa)(x, a) \leq Sg(x, a) + \alpha\beta M\kappa(x) \quad \text{for all } M \in \mathbb{R}_+. \quad (32)$$

<sup>22</sup>For all  $g_1, g_2 \in \mathcal{G}$ , we have  $g_1(x, a) \leq g_2(x, a) + \|g_1 - g_2\|_\kappa \kappa(x)$ . The monotonicity of  $S$  and (32) then imply that  $Sg_1(x, a) \leq Sg_2(x, a) + \alpha\beta \|g_1 - g_2\|_\kappa \kappa(x)$ . Switching the roles of  $g_1$  and  $g_2$  yields  $\|Sg_1 - Sg_2\|_\kappa \leq \alpha\beta \|g_1 - g_2\|_\kappa$ .

Condition (32) obviously holds, because by Assumption 5.1, we have

$$\begin{aligned}
S(g + M\kappa)(x, a) &= \beta \mathbb{E}_{x,a} \sup_{a' \in \Gamma(x')} \{r(x', a') + g_1(x', a') + M\kappa(x')\} \\
&= \beta \mathbb{E}_{x,a} \sup_{a' \in \Gamma(x')} \{r(x', a') + g_1(x', a')\} + \beta M \mathbb{E}_{x,a} \kappa(x') \\
&\leq Sg(x, a) + \alpha \beta M \kappa(x).
\end{aligned}$$

Hence  $S$  is a contraction mapping on  $(\mathcal{G}, \|\cdot\|_\kappa)$  and claim (1) is verified. Claims (2) and (3) follow immediately from claim (1) and the Banach contraction mapping theorem.

To show claims (a)–(c), Let  $\mathcal{V}$  (resp.,  $\mathcal{V}_1$ ) be the set of measurable functions  $v : \mathbf{X} \rightarrow \mathbb{R} \cup \{-\infty\}$  such that  $(x, a) \mapsto \beta \mathbb{E}_{x,a} v(x')$  is in  $\mathcal{G}$  (resp.,  $\mathcal{G}_1$ ), and let  $\mathcal{H}$  (resp.,  $\mathcal{H}_1$ ) be the set of measurable functions  $h : \mathbf{D} \rightarrow \mathbb{R} \cup \{-\infty\}$  such that  $h = r + g$  for some  $g$  in  $\mathcal{G}$  (resp.,  $\mathcal{G}_1$ ). Define the operators  $W_0$ ,  $W_1$  and  $M$  respectively on  $\mathcal{V}$ ,  $\mathcal{G}$  and  $\mathcal{H}$  as

$$W_0 v(x, a) := \beta \mathbb{E}_{x,a} v(x'), \quad W_1 g(x, a) := r(x, a) + g(x, a), \quad Mh(x) := \sup_{a \in \Gamma(x)} h(x, a).$$

Then the original Bellman operator  $T$  (the right hand side of (15)) and the transformed Bellman operator  $S$  in (17) satisfy  $T = MW_1 W_0$  on  $\mathcal{V}$  and  $S = W_0 M W_1$  on  $\mathcal{G}$ , respectively.

Since  $\mathcal{G}_1$  is a closed subset of  $\mathcal{G}$  and  $S\mathcal{G}_1 \subset \mathcal{G}_1$ ,  $S$  is also a contraction mapping on  $(\mathcal{G}_1, \|\cdot\|_\kappa)$  and the unique fixed point  $g^*$  of  $S$  is indeed in  $\mathcal{G}_1$ . By Lemmas 6.1 and 6.2 of Ma and Stachurski (2020), the Bellman operator  $T := MW_1 W_0$  maps elements of  $\mathcal{V}_1$  into itself and has a unique fixed point  $\bar{v}$  in  $\mathcal{V}_1$  that satisfies  $\bar{v} = MW_1 g^*$  and  $g^* = W_0 \bar{v}$ .

To verify claim (a), it remains to show that  $\bar{v}$  equals the value function  $v^*$  in (14). For all  $x_0 \in \mathbf{X}$  and  $\sigma \in \Sigma$ , we have

$$\begin{aligned}
\bar{v}(x_0) &\geq r(x_0, \sigma(x_0)) + \beta \mathbb{E}_{x_0, \sigma(x_0)} \bar{v}(x_1) \\
&\geq r(x_0, \sigma(x_0)) + \beta \mathbb{E}_{x_0, \sigma(x_0)} \{r(x_1, \sigma(x_1)) + \beta \mathbb{E}_{x_1, \sigma(x_1)} \bar{v}(x_2)\} \\
&= r(x_0, \sigma(x_0)) + \beta \mathbb{E}_{x_0, \sigma(x_0)} r(x_1, \sigma(x_1)) + \beta^2 \mathbb{E}_{x_0, \sigma(x_0)} \mathbb{E}_{x_1, \sigma(x_1)} \bar{v}(x_2) \\
&\geq \sum_{t=0}^T \beta^t \mathbb{E}_{x_0, \sigma(x_0)} \cdots \mathbb{E}_{x_{t-1}, \sigma(x_{t-1})} r(x_t, \sigma(x_t)) + \beta^{T+1} \mathbb{E}_{x_0, \sigma(x_0)} \cdots \mathbb{E}_{x_T, \sigma(x_T)} \bar{v}(x_{T+1}) \\
&= \sum_{t=0}^T \beta^t \mathbb{E}_{x_0} r(x_t, \sigma(x_t)) + \beta^T \mathbb{E}_{x_0, \sigma(x_0)} \cdots \mathbb{E}_{x_{T-1}, \sigma(x_{T-1})} g^*(x_T, \sigma(x_T)). \tag{33}
\end{aligned}$$



Notice that, by Assumption 5.1, we have

$$\begin{aligned}
& \left| \beta^T \mathbb{E}_{x_0, \sigma(x_0)} \cdots \mathbb{E}_{x_{T-1}, \sigma(x_{T-1})} g^*(x_T, \sigma(x_T)) \right| \\
& \leq \beta^T \mathbb{E}_{x_0, \sigma(x_0)} \cdots \mathbb{E}_{x_{T-1}, \sigma(x_{T-1})} |g^*(x_T, \sigma(x_T))| \\
& \leq \beta^T \mathbb{E}_{x_0, \sigma(x_0)} \cdots \mathbb{E}_{x_{T-1}, \sigma(x_{T-1})} \|g^*\|_\kappa \kappa(x_T) \\
& \leq \beta^T \alpha^T \|g^*\|_\kappa \kappa(x_0) = (\alpha\beta)^T \|g^*\|_\kappa \kappa(x_0) \rightarrow 0 \quad \text{as } T \rightarrow \infty.
\end{aligned}$$

Letting  $T \rightarrow \infty$  in (33), Lemma 6.1 implies that  $\bar{v}(x_0) \geq v_\sigma(x_0)$ . Since  $x_0 \in \mathbf{X}$  and  $\sigma \in \Sigma$  are arbitrary, we have  $\bar{v} \geq v^*$ . Moreover, since  $g^* = W_0 \bar{v}$  and there exists a  $g^*$ -greedy policy  $\sigma^*$  by assumption, all the inequalities in (33) hold with equality once we let  $\sigma = \sigma^*$ . In other words, we have  $\bar{v} = v_{\sigma^*} \leq v^*$ . In summary, we have shown that  $\bar{v} = v^*$ . Hence,  $g^* = W_0 v^*$  and  $v^* = MW_1 g^*$ , and claim (a) holds.

Since we have shown that  $v^*$  is the unique fixed point of  $T$  in  $\mathcal{V}_1$ , by Theorem 1 of Ma and Stachurski (2020), the set of optimal policies is nonempty, and a feasible policy is optimal if and only if it is  $v^*$ -greedy. Since in addition  $g^* = W_0 v^*$ , claims (b) and (c) hold.  $\square$

*Proof of Theorem 5.2.* To apply Theorem 5.1, it suffices to prove that  $\mathcal{G}_1$  is a closed subset of  $\mathcal{G}$ ,  $S\mathcal{G}_1 \subset \mathcal{G}_1$ , and that a  $g$ -greedy policy exists for each  $g \in \mathcal{G}_1$ .

To show that  $\mathcal{G}_1$  is a closed subset, let  $\{g_n\}$  be a sequence in  $\mathcal{G}_1$  such that  $\|g_n - g_0\|_\kappa \rightarrow 0$  for some  $g_0 \in \mathcal{G}$ . Because  $(\mathcal{G}, \|\cdot\|_\kappa)$  is complete, it suffices to show that  $g_0$  is upper semicontinuous. For all  $(x_0, a_0) \in \mathbf{D}$  and  $y > g_0(x_0, a_0)$ . Let  $\varepsilon := y - g_0(x_0, a_0)$ . Since  $\kappa$  is continuous and  $g_n$  is upper semicontinuous for all  $n$ , there exist  $N \in \mathbb{N}$  and a neighborhood  $B$  of  $(x_0, a_0)$  such that for all  $(x, a) \in B$ ,

$$|g_N(x, a) - g_0(x, a)| < \varepsilon/3 \quad \text{and} \quad g_N(x, a) < g_N(x_0, a_0) + \varepsilon/3.$$

Hence,  $g_0(x, a) < g_N(x, a) + \varepsilon/3 < g_N(x_0, a_0) + 2\varepsilon/3 < g_0(x_0, a_0) + \varepsilon = y$  for each  $(x, a) \in B$ , implying that  $g_0$  is upper semicontinuous.

Fix  $g \in \mathcal{G}_1$ . Note that  $r + g$  is upper semicontinuous because both  $r$  and  $g$  are upper semicontinuous. Since, in addition,  $\Gamma$  is compact-valued and upper hemicontinuous, Lemma 1 of Jaśkiewicz and Nowak (2011) implies that a  $g$ -greedy policy exists, and that  $x \mapsto h_g(x) := \sup_{a \in \Gamma(x)} \{r(x, a) + g(x, a)\}$  is upper semicontinuous.

Since  $Sg \in \mathcal{G}$  by Theorem 5.1, to see that  $S\mathcal{G}_1 \subset \mathcal{G}_1$ , it remains to show that  $Sg$  is upper semicontinuous. Assumption 5.1 and the definition of  $\mathcal{G}_1$  yield  $h_g \leq (d + \|g\|_\kappa)\kappa$ , so Lemma 7 of Jaśkiewicz and Nowak (2011) implies  $Sg(x, a) = \beta \mathbb{E}_{x, a} h_g(x')$  is upper semicontinuous. Hence,  $S\mathcal{G}_1 \subset \mathcal{G}_1$ .

All the claims of Theorem 5.2 then follow from Theorem 5.1.  $\square$

*Proof of Theorem 5.3.* To see claims (1)–(3) hold, we first show that  $S\mathcal{G} \subset \mathcal{G}$ . Fix  $g \in \mathcal{G}$ . Since  $g \geq -\|g\|$ , we obtain

$$\begin{aligned} Sg(x, a) &\geq -\frac{\beta}{\gamma} \log \mathbb{E}_{x,a} \exp \left( -\gamma \sup_{a' \in \Gamma(x')} \{r(x', a') - \|g\|\} \right) \\ &= -\frac{\beta}{\gamma} \log \mathbb{E}_{x,a} e^{-\gamma(\hat{r}(x') - \|g\|)} = \beta[\hat{r}(x, a) - \|g\|], \end{aligned}$$

where the last equality uses (31). Since by assumption  $\hat{r}$  is bounded below, so is  $Sg$ . A similar argument yields  $Sg(x, a) \leq \beta[\hat{r}(x, a) + \|g\|]$ , so  $Sg$  is bounded above. This shows that  $S$  is a self map on  $\mathcal{G}$ .

To show that  $S$  is a contraction mapping, we verify Blackwell (1965)’s sufficient conditions.  $S$  is clearly monotone. Let  $h(x, a) := r(x, a) + g(x, a)$ . If  $M \geq 0$  is any constant, then for any  $g \in \mathcal{G}$  we have

$$\begin{aligned} S(g + M)(x, a) &= -\frac{\beta}{\gamma} \log \mathbb{E}_{x,a} \exp \left( -\gamma \sup_{a' \in \Gamma(x')} \{h(x', a') + M\} \right) \\ &= -\frac{\beta}{\gamma} \log \mathbb{E}_{x,a} \exp \left( -\gamma \sup_{a' \in \Gamma(x')} h(x', a') \right) + \beta M = Sg(x, a) + \beta M, \end{aligned}$$

so the discounting property holds. Therefore, claims (1)–(3) hold.

To show that  $v^*$  is well-defined and that claims (a)–(c) hold, let  $\mathcal{G}_1$  be the set of upper semicontinuous functions in  $\mathcal{G}$ . A similar argument to the proof of Theorem 5.2 shows  $S\mathcal{G}_1 \subset \mathcal{G}_1$ . For any feasible policy  $\sigma \in \Sigma$ , define the map  $S_\sigma$  on  $\mathcal{G}$  by

$$S_\sigma g(x, a) := -\frac{\beta}{\gamma} \log \mathbb{E}_{x,a} \exp \left( -\gamma \{r(x', \sigma(x')) + g(x', \sigma(x'))\} \right).$$

By the same argument as above, we can show that  $S_\sigma$  is a contraction mapping on  $\mathcal{G}$  of modulus  $\beta$ . Combining Lemmas 6.1, 6.2, and Proposition 3.4 of Ma and Stachurski (2020), it follows that  $v_\sigma$  and  $v^*$  are well-defined and claims (a)–(c) hold.  $\square$

## REFERENCES

ABBRING, J. H., J. R. CAMPBELL, J. TILLY, AND N. YANG (2018): “Very Simple Markov-Perfect Industry Dynamics: Theory,” *Econometrica*, 86, 721–735.

- AÇIKGÖZ, Ö. T. (2018): “On the Existence and Uniqueness of Stationary Equilibrium in Bewley Economies with Production,” *Journal of Economic Theory*, 173, 18–55.
- AGUIAR, M. AND M. AMADOR (2019): “A Contraction for Sovereign Debt Models,” *Journal of Economic Theory*, 183, 842–875.
- AGUIAR, M., M. AMADOR, H. HOPENHAYN, AND I. WERNING (2019): “Take the Short Route: Equilibrium Default and Debt Maturity,” *Econometrica*, 87, 423–462.
- AIYAGARI, S. R. (1994): “Uninsured Idiosyncratic Risk and Aggregate Saving,” *Quarterly Journal of Economics*, 109, 659–684.
- ALIPRANTIS, C. D. AND K. C. BORDER (2006): *Infinite Dimensional Analysis*, Springer, third ed.
- ALVAREZ, F. AND N. L. STOKEY (1998): “Dynamic Programming with Homogeneous Functions,” *Journal of Economic Theory*, 82, 167–189.
- ARELLANO, C. (2008): “Default Risk and Income Fluctuations in Emerging Economies,” *American Economic Review*, 98, 690–712.
- BÄUERLE, N. AND A. JAŚKIEWICZ (2018): “Stochastic Optimal Growth Model with Risk Sensitive Preferences,” *Journal of Economic Theory*, 173, 181–200.
- BÄUERLE, N. AND U. RIEDER (2011): *Markov Decision Processes with Applications to Finance*, Springer Science & Business Media.
- BENHABIB, J., A. BISIN, AND S. ZHU (2015): “The Wealth Distribution in Bewley Economies with Capital Income Risk,” *Journal of Economic Theory*, 159, 489–515.
- BERTSEKAS, D. P. (2017): *Dynamic Programming and Optimal Control*, vol. 2, Athena Scientific, fourth ed.
- (2018): *Abstract Dynamic Programming*, Athena Scientific, second ed.
- BLACKWELL, D. (1962): “Discrete Dynamic Programming,” *Annals of Mathematical Statistics*, 33, 719–726.
- (1965): “Discounted Dynamic Programming,” *Annals of Mathematical Statistics*, 36, 226–235.
- BOYD, III, J. H. (1990): “Recursive Utility and the Ramsey Problem,” *Journal of Economic Theory*, 50, 326–345.
- CAO, D. (2020): “Recursive Equilibrium in Krusell and Smith (1998),” *Journal of Economic Theory*, 186, 104978.
- CAO, D. AND W. LUO (2017): “Persistent Heterogeneous Returns and Top End Wealth Inequality,” *Review of Economic Dynamics*, 26, 301–326.
- CASTAÑEDA, A., J. DÍAZ-GIMÉNEZ, AND J.-V. RÍOS-RULL (2003): “Accounting for the U.S. Earnings and Wealth Inequality,” *Journal of Political Economy*, 111,

818–857.

- EJRNÆS, M. AND M. BROWNING (2014): “The Persistent-Transitory Representation for Earnings Processes,” *Quantitative Economics*, 5, 555–581.
- FAGERENG, A., L. GUIISO, D. MALACRINO, AND L. PISTAFERRI (2020): “Heterogeneity and Persistence in Returns to Wealth,” *Econometrica*, 88, 115–170.
- FÖLLMER, H. AND A. SCHIED (2004): *Stochastic Finance: An Introduction in Discrete Time*, De Gruyter, Berlin.
- HANSEN, L. P. AND T. J. SARGENT (2008): *Robustness*, Princeton University Press.
- HE, H. AND N. D. PEARSON (1991): “Consumption and Portfolio Policies with Incomplete Markets and Short-Sale Constraints: The Finite-Dimensional Case,” *Mathematical Finance*, 1, 1–10.
- HERNÁNDEZ-LERMA, O. AND J. B. LASSERRE (1999): *Further Topics on Discrete-Time Markov Control Processes*, vol. 42 of *Applications of Mathematics*, Springer.
- HUBMER, J., P. KRUSELL, AND A. A. SMITH, JR. (2020): “Sources of US Wealth Inequality: Past, Present, and Future,” in *NBER Macroeconomics Annual*, ed. by M. Eichenbaum and E. Hurst, Chicago: University of Chicago Press, vol. 35, chap. 6.
- JAŚKIEWICZ, A. AND A. S. NOWAK (2011): “Discounted Dynamic Programming with Unbounded Returns: Application to Economic Models,” *Journal of Mathematical Analysis and Applications*, 378, 450–462.
- JOVANOVIĆ, B. (1982): “Selection and the Evolution of Industry,” *Econometrica*, 50, 649–670.
- KAMIHIGASHI, T. (2014): “Elementary Results on Solutions to the Bellman Equation of Dynamic Programming: Existence, Uniqueness, and Convergence,” *Economic Theory*, 56, 251–273.
- LE VAN, C. AND Y. VAILAKIS (2005): “Recursive Utility and Optimal Growth with Bounded or Unbounded Returns,” *Journal of Economic Theory*, 123, 187–209.
- LI, H. AND J. STACHURSKI (2014): “Solving the Income Fluctuation Problem with Unbounded Rewards,” *Journal of Economic Dynamics and Control*, 45, 353–365.
- MA, Q. AND J. STACHURSKI (2020): “Dynamic Programming Deconstructed: Transformations of the Bellman Equation and Computational Efficiency,” *Operations Research*, forthcoming.
- MA, Q., J. STACHURSKI, AND A. A. TODA (2020): “The Income Fluctuation Problem and the Evolution of Wealth,” *Journal of Economic Theory*, 187, 105003.
- MARTINS-DA-ROCHA, V. F. AND Y. VAILAKIS (2010): “Existence and Uniqueness of a Fixed Point for Local Contractions,” *Econometrica*, 78, 1127–1141.

- MATKOWSKI, J. AND A. S. NOWAK (2011): “On Discounted Dynamic Programming with Unbounded Returns,” *Economic Theory*, 46, 455–474.
- MCCALL, J. J. (1970): “Economics of Information and Job Search,” *Quarterly Journal of Economics*, 84, 113–126.
- RINCÓN-ZAPATERO, J. P. AND C. RODRÍGUEZ-PALMERO (2003): “Existence and Uniqueness of Solutions to the Bellman Equation in the Unbounded Case,” *Econometrica*, 71, 1519–1555.
- RUST, J. (1987): “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher,” *Econometrica*, 55, 999–1033.
- SAMUELSON, P. A. (1969): “Lifetime Portfolio Selection by Dynamic Stochastic Programming,” *Review of Economics and Statistics*, 51, 239–246.
- STACHURSKI, J. AND A. A. TODA (2019): “An Impossibility Theorem for Wealth in Heterogeneous-agent Models with Limited Heterogeneity,” *Journal of Economic Theory*, 182, 1–24.
- STOKEY, N., R. LUCAS, AND E. PRESCOTT (1989): *Recursive Methods in Economic Dynamics*, Harvard University Press.
- SZEPESVÁRI, C. (2010): “Algorithms for Reinforcement Learning,” *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 4, 1–103.
- TODA, A. A. (2014): “Incomplete Market Dynamics and Cross-Sectional Distributions,” *Journal of Economic Theory*, 154, 310–348.
- VAN DER WAL, J. (1980): *Stochastic Dynamic Programming: Successive Approximations and Nearly Optimal Strategies for Markov Decision Processes and Markov Games*, Stichting Mathematisch Centrum.
- WATKINS, C. J. C. H. AND P. DAYAN (1992): “Q-Learning,” *Machine Learning*, 8, 279–292.
- WESSELS, J. (1977): “Markov Programming by Successive Approximations with Respect to Weighted Supremum Norms,” *Journal of Mathematical Analysis and Applications*, 58, 326–335.