Unbounded Dynamic Programming via the Q-Transform[1]

Qingyin Ma[a], John Stachurski[b], Alexis Akira Toda[c]

[a]ISEM, Capital University of Economics and Business
[b]Research School of Economics, Australian National University
[c]Department of Economics, University of California San Diego

March 17, 2021

Abstract. We propose a new approach to solving dynamic decision problems with unbounded rewards based on the transformations used in Q-learning. 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 simple relative to other techniques and accessible for applied work. We show by example that many common decision problems satisfy our conditions.

JEL Classifications: C61, C65
Keywords: Dynamic programming, Optimality, Q-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 algorithms. Standard textbook treatments can be found in Stokey et al. (1989) and Bertsekas (2017).

[1]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

Unfortunately, the approach is not actually applicable in the vast majority of concrete economic problems. This is 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

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

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.[2] 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.[3] Local contraction methods are broadly applicable under reasonable assumptions but require verifying technical conditions involving increasing sequences of compact sets that exhaust the

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

[3]To be specific, the weighted supremum norm approach can be applied broadly when the reward function is unbounded above and bounded below. However, when the reward function is unbounded below, it is generally hard to construct a proper weighting function and thus a contraction mapping via the weighted supremum norm. Therefore, it is challenging to characterize the value function as the unique fixed point of the Bellman operator. For more background, see, e.g., Le Van and Vailakis (2005) and Jaśkiewicz and Nowak (2011).

state space.[4] 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, a popular reinforcement learning algorithm that allows online updating by a controller in an incremental fashion (see, e.g., Watkins and Dayan (1992) or Szepesvári (2010)). It has been shown that the algorithm has strong global convergence properties and our results are in this spirit.

The core idea is as follows: In standard dynamic programs, the Bellman operator is defined by composing the following two operations: given a candidate value function v, (a) compute the discounted expectation $g := \beta \mathbb{E} v$ over current states and actions, and (b) add current reward r, maximize over current feasible actions, and then update the candidate value function as $v = \max\{r + g\}$. Instead of aiming at updating v, one can transform the Bellman operator by applying operations $(b) \to (a)$, which is the Q-transform, and focus on updating the candidate "action-value" function g. We show that, under relatively weak and easily testable conditions, the transformed Bellman operator is a contraction with unique fixed point g^* that is bounded, and the true value function v^* can be recovered as $v^* = \max\{r + g^*\}$.

One advantage of the transformation-based approach to unbounded programs adopted in this paper is that the methodology fits well with the case where the weighted supremum norm approach struggles: maximization problems where rewards are unbounded below (see Footnote 3). Such optimization problems are commonplace in quantitative applications, such as those involving CRRA flow utility with $\gamma > 1$, which is the empirically relevant case. We show that, for many canonical applications from this class of problems, the Q-transform converts unbounded value functions into bounded action-value functions. Standard contraction mapping theory can then be applied.

A second advantage of the Q-transform method is that it has no difficulty handling stochastic dynamic programs. In fact, the action-value functions associated with the

[4] 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).

Q-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. Therefore, it requires less restriction on the primitive setup compared with many existing methods. In contrast, many existing papers on unbounded dynamic programming either focus on the deterministic case or restrict the support of shocks.[5]

A third advantage is that, despite its generality, the Q-transform approach is accessible to a general audience. In particular, it relies mainly on the standard contraction mapping theorem, and can be conveniently generalized to handle dynamic programs where rewards are both unbounded above and unbounded below. 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-transform has been used to convert unbounded reward dynamic programs into bounded ones.[6]

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 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

[5]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.

[6]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 (2021). These transformations are typically aimed at improving economic intuition, estimation properties or computational efficiency.

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 Kuhn (2013) 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-transform method can be applied to the kinds of optimal savings problem with endogenous labor choice that are common in both theoretic and applied works (see, e.g., Castañeda et al. (2003) and Zhu (2020)). The Euler equation method of Kuhn (2013) 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 is not applicable in the optimal 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-transform 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).[7] They provide valuable optimality results, although these results cannot treat many common period utility functions, such as CRRA with relative risk aversion at least one or logarithm utility, 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 the Q-transform.

[7]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 $\gamma \to 0$ and using the property $\lim_{\gamma\to 0} -\frac{1}{\gamma}\log \mathbb{E}e^{-\gamma X} = \mathbb{E}X$ 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).

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, and also considers the case with recursive (risk-sensitive) preferences. Section 6 concludes. Main 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 presentation of the theory.

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

$$\text{maximize} \qquad \mathbb{E}\sum_{t=0}^{\infty}\beta^t u(c_t) \tag{2a}$$

$$\text{subject to} \qquad 0 \leqslant c_t \leqslant w_t, \tag{2b}$$

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

with (w_0, y_0) given. Here $\beta \in (0,1)$ is the discount factor, $c_t, w_t, y_t \geqslant 0$ are, respectively, consumption, wealth and non-financial income at time t, $R \geqslant 0$ is the gross rate of return on financial income, and $u\colon \mathbb{R}_+ \to \mathbb{R}\cup\{-\infty\}$ is a utility function, which is increasing and continuous.[8] For now, suppose that u is bounded above but 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, ξ_t is an IID shock of arbitrary dimension, and y is a nonnegative measurable function. The Bellman equation of this problem is

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

[8]The optimal savings problem represents one of the fundamental workhorses of modern macroeconomics. 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.

where $y' = y(z', \xi')$. The value function is unbounded below, and the classical arguments in Blackwell (1965) cannot be applied.[9]

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

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

which is a relatively mild restriction.[10] Let

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

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). Following the spirit of Q-learning, we begin by rewriting the Bellman equation in terms of the action-value function alone.[11]

In particular, we combine (3) and (5) to give

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

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

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

Taking the conditional expectation of both sides with respect to z, multiplying by β and using (5) again, we get

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

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

[9] To confirm this, suppose to the contrary that v is the value function and $|v| \leqslant M < \infty$. Then $-M \leqslant v(0, z) \leqslant u(0) + \beta M = -\infty$. Contradiction.

[10] The expectation in (4) should be understood as $\mathbb{E}[u(y(z_{t+1}, \xi_{t+1})) \mid z_t = z]$. Condition (4) holds if, say, $y(z, \xi) = z$ for all ξ and Z is finite and positive (Aiyagari, 1994; Cao, 2020), or if income has a persistent-transitory representation (Heathcote et al., 2010; 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 ξ has a finite moment generating function).

[11] Our approach is slightly different from the Q-learning approach, where the action-value function is defined as $u+g$ instead. We will show that eliminating the current reward (which can be unbounded below) can help us transform the action-value function into a bounded function and, as a result, the classical dynamic programming theory applies. This is the essence of our approach.

of (7) is a fixed point of S and vice versa. Let $\mathcal{G}$ be the space of bounded measurable functions on the set D defined by

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

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 β with respect to the supremum norm.

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

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

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

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

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 $g_1 \leqslant g_2$ implies $Sg_1 \leqslant Sg_2$, so monotonicity holds. If $M \geqslant 0$ is any constant, then for any $g \in \mathcal{G}$ we have

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

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.[12]

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.[13] A country with current assets w_t chooses between continuing to participate in international financial markets and defaulting. As in Section 2.1, output $y_t = y(z_t, \xi_t)$ is a function of a Markov process $\{z_t\}$ and an IID shock $\{\xi_t\}$. 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

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

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

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

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

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

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

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

[12]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., Kuhn (2013) 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.

[13]See, e.g., Hatchondo et al. (2009), Hatchondo et al. (2016) and Aguiar et al. (2019).

Eliminating the value function v yields

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

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

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

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

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

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 ξ_t is an IID shock that could be vector-valued.[14] 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

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

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

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

[14]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 ξ_1 and ξ_2 are independent, $w(z, \xi)$ depends only on z and ξ_1, and $c(z, \xi)$ depends only on z and ξ_2.

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

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

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

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

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

$$\mathbb{E} \max \{X, Y\} \geqslant \max \{\mathbb{E}X, \mathbb{E}Y\} \tag{12}$$

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

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

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

An argument similar to the one adopted in Sections 2.1–2.2 shows 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 and Theory

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. **Problem Formulation.** For a given topological space E, let $\mathscr{B}(E)$ be the Borel subsets of E. For our purpose, a dynamic program consists of

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

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

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

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 $x_{t+1} \sim P(x_t, a_t, \cdot)$.

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

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

whenever the expectation and infinite sum are well-defined. We understand $v_\sigma(x)$ as the lifetime value of following policy σ 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

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

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

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

[16] 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

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

where $\mathbb{E}_{x,a}$ denotes the expectation with respect to the probability measure $P(x,a,\cdot)$.

3.2. **The Q-Transform.** As in the examples in Section 2, we define the action-value function g as $g(x,a) := \beta \mathbb{E}_{x,a} v(x')$. Similar to the Q-learning approach, we rewrite the Bellman equation in terms of the action-value function, which gives

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

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

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

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

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

A feasible policy σ is called *g-greedy* if

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

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

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

The function $\bar{r}$ can be interpreted as the maximum reward given the current state $x \in \mathsf{X}$. 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 D and $\|\cdot\|$ be the supremum norm. In spite 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.

3.3. **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 Theorem 3.1 to hold. To this end, we introduce an additional assumption.

Assumption 3.2. (1) The sets X and A are complete separable metric spaces, (2) the reward function r is upper semicontinuous, (3) the feasible correspondence Γ is compact-valued and upper hemicontinuous,[17] and (4) the stochastic kernel P is Feller.[18]

In most applications of interest, 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.*

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

[18] In other words, $(x,a) \mapsto \int h(x') P(x,a,\mathrm{d}x')$ is bounded and continuous whenever h is.

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.

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.[19] For concreteness, suppose that

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

where R, y are nonnegative measurable functions, z_t is a finite state Markov chain, and ξ_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

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

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 P is defined through the (exogenous) stochastic kernel of the Markov state z, the distribution of the IID shock ξ, and the budget constraint (2c). The functions $\bar{r}$ and $\hat{r}$ in (19) are defined by

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

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, Γ 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.

[19] 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.

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 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

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

The following argument shows that our assumptions are weaker.[20] Since u is concave and differentiable under the assumptions of MST, we obtain

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

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

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

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 βR_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,

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

where i is a discrete choice variable taking values in $\{0, 1\}$, and Y and 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 $\mathsf{Z} \times [-b, \infty) \times \{0, 1\}$. Moreover, $\hat{r}$ satisfies

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

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

[20] Ma et al. (2020) allow the discount factor β to be random. It is straightforward to extend our theory to a setting with stochastic discounting.

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.

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,

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

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

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

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 8), and $R_{t+1} \geqslant 0$ 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 ξ_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

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

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

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

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

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

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

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

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

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

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 $j = 1, \ldots, J$ in state z_t (θ_t is a portfolio), $y_t = y(z_t, \xi_t)$ is non-financial income (ξ_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 θ_t and shocks (z_{t+1}, ξ_{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 Q-transform 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 κ, which is a continuous function mapping X to $[1, \infty)$. Let $\mathcal{G}$ be the set of measurable functions $g : \mathsf{D} \to \mathbb{R}$ such that g is bounded below and

$$\|g\|_\kappa := \sup_{(x,a)\in\mathsf{D}} \frac{|g(x,a)|}{\kappa(x)} < \infty. \tag{24}$$

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 $\bar{r}(x) \leqslant d\kappa(x)$ and $\mathbb{E}_{x,a}\kappa(x') \leqslant \alpha\kappa(x)$ for all $(x, a) \in \mathsf{D}$, and (2) $\hat{r}$ in (19) is bounded below.

Remark 5.1. Note that Assumption 3.1 is a special case of Assumption 5.1 by setting $\kappa(x) \equiv 1$ and $\alpha = 1$. More importantly, Assumption 5.1 relaxes the classical weighted supremum norm assumptions greatly (see, e.g., Wessels (1977) or Bertsekas (2018)), in the sense that we allow the ratio of the reward function to the weighting function $\bar{r}/\kappa$ to be unbounded from below, a case where the classical weighted supremum norm approach struggles (recall Footnote 3).

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

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

(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

$$\hat{\kappa}(x,a) \coloneqq \mathbb{E}_{x,a} \kappa(x'). \tag{25}$$

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 (25) 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.*

Example 5.1. 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

$$u(c) \leqslant pc + q \quad \text{for all } c > 0. \tag{26}$$

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

Suppose that asset return and income take the form

$$R_t = R(\xi_t) \quad \text{and} \quad y_t = y(z_t, \xi_t), \tag{27}$$

where R, y are nonnegative measurable functions, z_t is a finite state Markov chain, and ξ_t is an IID shock that could be vector-valued.[21] In addition, assume

$$Y \coloneqq \sup_z \mathbb{E}_z y(z', \xi') < \infty, \quad \beta < 1, \quad \text{and} \quad \beta \mathbb{E} R < 1. \tag{28}$$

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

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

[21] Unlike the setting in Section 4.1, the return is permitted to depend only on the IID shock ξ. Treating the general case requires generalizing Theorem 5.1 further such that α depends on x.

by (26), 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 $\mathbb{E}_{x,a}\kappa(x') \leqslant \alpha\kappa(x)$ for some $\alpha \in (0, 1/\beta)$. To this end, note that

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

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

$$\sup_{(x,a)\in\mathsf{D}} \frac{\mathbb{E}_{x,a}\kappa(x')}{\kappa(x)} \leqslant \sup_z \max\left\{\frac{p\mathbb{E}_z y'+q}{q}, \mathbb{E}_z R'\right\} \leqslant \max\left\{\frac{pY+q}{q}, \mathbb{E}R\right\}, \tag{29}$$

where we have used the fact that R does not depend on z (by (27)) and (28). Since $q > 1$ can be taken arbitrarily large and $\frac{pY+q}{q} \to 1$ as $q \to \infty$, the right hand side of (29) can be made arbitrarily close to $\max\{1, \mathbb{E}R\}$, which is strictly smaller than $1/\beta$ by (28). 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,

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

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

Remark 5.2. Under the assumption that discounting is constant and the asset return depends only on the IID shock as in (27), the assumptions in Ma et al. (2020) are strictly stronger than ours, since they assume (28) and the concavity of u (which implies (26)). (See also Remark 4.1.)

5.2. **Risk-Sensitive Preferences.** We consider the general setting in Sections 3 but with recursive (non-additive) preferences. Unlike the additively separable case, in order to define the value function and optimality in the recursive case, let $\gamma > 0$ be the agent's risk-sensitive coefficient. Given any feasible policy $\sigma \in \Sigma$ and measurable function $v : \mathsf{X} \to \mathbb{R} \cup \{-\infty\}$, let

$$T_\sigma v(x) \coloneqq r(x, \sigma(x)) - \frac{\beta}{\gamma} \log \mathbb{E}_{x,\sigma(x)} \mathrm{e}^{-\gamma v(x')} \tag{30}$$

for all $x \in \mathsf{X}$ whenever the expectation is well-defined. Recall $\bar{r}$ defined in (19). In this case, the *σ-value function* v_σ is defined at each state $x \in \mathsf{X}$ by

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

Setting aside the issue of existence and uniqueness for now, $v_\sigma(x)$ can be interpreted as the lifetime value of following policy σ forever, starting from current state x. The value function v^* and optimal policy σ^* are defined as in (14). The Bellman equation associated with the dynamic program with risk-sensitive coefficient $\gamma > 0$ is

$$v(x) = \sup_{a \in \Gamma(x)} \left\{ r(x,a) - \frac{\beta}{\gamma} \log \mathbb{E}_{x,a} \mathrm{e}^{-\gamma v(x')} \right\}. \tag{31}$$

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

$$g(x,a) = -\frac{\beta}{\gamma} \log \mathbb{E}_{x,a} \exp\left(-\gamma \sup_{a' \in \Gamma(x')} \{ r(x',a') + g(x',a') \} \right). \tag{32}$$

We define the *transformed Bellman operator* S by letting $Sg(x,a)$ be the right hand side of (32) for each g in the space of candidate action-value functions (to be specified below). A feasible policy $\sigma \in \Sigma$ is called *g-greedy* if (18) holds.

We first consider the case with rewards that are bounded above. Let $\mathcal{G}$ be the set of bounded measurable functions on D and $\|\cdot\|$ be the supremum norm. 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*

$$\hat{r}(x,a) := -\frac{1}{\gamma} \log \mathbb{E}_{x,a} \mathrm{e}^{-\gamma \bar{r}(x')}, \tag{33}$$

(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

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

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

Next, we study the case with rewards that are unbounded above and below. In what follows, X is a partially ordered set. Similar to Section 5.1, we introduce a weighting

function κ, which is continuous increasing and maps X to $[1, \infty)$. Let $\mathcal{G}_2$ be the set of measurable function $g : \mathsf{D} \to \mathbb{R}$ such that $g(x, a)$ is increasing in x, bounded below, and (24) holds.

Suppose the state process evolves according to

$$x_{t+1} = f(x_t, a_t, \varepsilon_{t+1}), \tag{34}$$

where f is a measurable function, and $\{\varepsilon_t\}$ is an IID innovation process taking values in $\mathbb{R}^m$. For each t, we write $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{mt})$ and make the following assumption.

Assumption 5.2. (1) $r(x, a)$ is increasing in x and $f(x, a, \varepsilon')$ is increasing in (x, ε'), (2) $\Gamma(x_1) \subset \Gamma(x_2)$ if $x_1 \leqslant x_2$, and (3) $\varepsilon_{1t}, \ldots, \varepsilon_{mt}$ are independent for each t.

The following theorem extends Theorem 5.1 and Theorem 5.2 to dynamic decision problems with risk-sensitive preference.

Theorem 5.4. *If Assumptions 5.1 and 5.2 hold for $\hat{r}$ defined in (33), then*

(1) $S\mathcal{G}_2 \subset \mathcal{G}_2$ and S is a contraction mapping on $(\mathcal{G}_2, \|\cdot\|_\kappa)$.
(2) S admits a unique fixed point g^* in $\mathcal{G}_2$, and
(3) $S^k g$ converges to g^* at rate $O((\alpha\beta)^k)$ under $\|\cdot\|_\kappa$.

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

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

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

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

Example 5.2. Bäuerle and Jaśkiewicz (2018) study an optimal growth model with risk-sensitive preference. In their setting,

$$\mathsf{X} = \mathsf{A} = \mathbb{R}_+, \quad \Gamma(x) = [0, x] \quad \text{and} \quad r(x, a) = u(x - a),$$

where x is capital, a is the amount of investment, and u is the utility function. Bäuerle and Jaśkiewicz (2018) assume that the utility function u is bounded below.

Consider, for example, $u(c) = \log c$ and

$$x_{t+1} = f(a_t, \varepsilon_{t+1}) = \eta a_t + \varepsilon_{t+1}, \quad \{\varepsilon_t\} \overset{\text{IID}}{\sim} LN(\mu, \sigma^2),$$

where $\eta > 0$. This setup is common in applied works and is not covered by Bäuerle and Jaśkiewicz (2018), because the logarithmic utility is unbounded below. However, Theorem 5.4 can be applied. Clearly, Assumptions 3.2 and 5.2 hold. Moreover, since $\bar{r}(x) = u(x)$ on X, for all $(x, a) \in \mathsf{D}$,

$$\mathbb{E}_{x,a} \mathrm{e}^{-\gamma \bar{r}(x')} = \mathbb{E} \mathrm{e}^{-\gamma u(\eta a + \varepsilon')} \leqslant \mathbb{E} \mathrm{e}^{-\gamma u(\varepsilon')} = \mathrm{e}^{-\gamma \mu + \gamma^2 \sigma^2 / 2} < \infty.$$

Hence $\hat{r}(x, a) \geqslant \mu - \gamma \sigma^2 / 2$ on D and Assumption 5.1(2) holds. Let $\bar{\varepsilon} \coloneqq \mathbb{E} \varepsilon_t$.

- If $\eta \leqslant 1$, then Assumption 5.1(1) holds for all $\beta \in (0, 1)$ by letting $\alpha \in (1, 1/\beta)$ and $\kappa(x) \coloneqq x + \bar{\varepsilon}/(\alpha - 1)$.
- If $\eta > 1$, then Assumption 5.1(1) holds for all $\beta \in (0, 1/\eta)$ by letting $\alpha \coloneqq \eta$ and $\kappa(x) \coloneqq x + \bar{\varepsilon}/(\alpha - 1)$.

Assumption 5.1 is now verified. Therefore, all the conclusions of Theorem 5.4 hold.

Remark 5.3. In the state evolution path (34), we impose $\{\varepsilon_t\}$ to be an IID innovation process. Indeed, this restriction can be relaxed. For example, we can set $\{\varepsilon_t\}$ to be a finite Markov chain, or $\log \varepsilon_t = \mu(z_t) + \sigma(z_t)\xi_t$, where $\{z_t\}$ is a finite Markov chain and $\{\xi_t\}$ is an IID innovation process with finite moment generating function, etc. By expanding the state vector to accommodate the exogenous state ε_t or z_t and adjusting Assumption 5.2 mildly, the conclusions of Theorem 5.4 still hold in these generalized settings.

6. Conclusion

We proposed a new approach to solving dynamic programs with unbounded rewards, based on Q-transforms. The essence of our approach lies in transforming an originally unbounded dynamic program into a bounded one. We demonstrated via a range of applications that our method fits well with stochastic dynamic decision problems with unbounded below rewards and can be extended to handle crucial dynamic programs that are both unbounded above and unbounded below. In particular, the Q-transform approach is not limited to solving dynamic programs that are additively separable. We showed that dynamic decision problems with risk-sensitive preference and unbounded (above and below) period reward functions are also covered by Q-transform. Although exploring further recursive preference decision problems via our approach goes beyond the scope of this study, the theory of Q-transform presented here should serve as a solid foundation for new work along these lines.

7. Appendix: Proof of Main Results

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 σ-value function v_σ in (13) and the value function v^* in (14) are well-defined.

Lemma 7.1. *If Assumption 5.1(1) holds, then for any feasible policy $\sigma \in \Sigma$ and initial state $x_0 = x \in \mathsf{X}$, 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

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

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

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

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. □

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 $L \in \mathbb{R}$ such that $g \geqslant L$. Then

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

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

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

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

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

$$S(g + K\kappa)(x,a) \leqslant Sg(x,a) + \alpha\beta K\kappa(x) \quad \text{for all } K \in \mathbb{R}_+. \tag{35}$$

Condition (35) obviously holds, because by Assumption 5.1, we have

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

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 see that claims (a)–(c) hold, let $\mathcal{V}$ be the set of measurable functions $v : \mathsf{X} \to \mathbb{R} \cup \{-\infty\}$ such that $(x,a) \mapsto \beta \mathbb{E}_{x,a} v(x')$ is in $\mathcal{G}$, and define the operators W on $\mathcal{V}$ and M on $\mathcal{G}$ respectively as

$$Wv(x,a) \coloneqq \beta \mathbb{E}_{x,a} v(x') \quad \text{and} \quad Mg(x) \coloneqq \sup_{a \in \Gamma(x)} \{r(x,a) + g(x,a)\}.$$

Then the original Bellman operator T (i.e., Tv equals the right hand side of (15) given $v \in \mathcal{V}$) and the transformed Bellman operator S in (16) can be written as

$$T = MW \quad \text{and} \quad S = WM.$$

In particular, for each $g \in \mathcal{G}$, because S maps $\mathcal{G}$ into itself as was shown, $W(Mg) = WMg = Sg \in \mathcal{G}$. Hence, $Mg \in \mathcal{V}$ by the definition of $\mathcal{V}$. As g is chosen arbitrarily, this implies that M maps $\mathcal{G}$ into $\mathcal{V}$, and thus T maps $\mathcal{V}$ into itself.

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 indeed lies in $\mathcal{G}_1$. Next, we show that $\bar{v} \coloneqq Mg^*$ is a fixed point of T in $\mathcal{V}$. Note that $\bar{v} \in \mathcal{V}$ because $M\mathcal{G} \subset \mathcal{V}$. Moreover,

$$T\bar{v} = MW\bar{v} = MWMg^* = MSg^* = Mg^* = \bar{v}.$$

Therefore, $\bar{v}$ is a fixed point of T in $\mathcal{V}$, as was to be shown.

[22]For all $g_1, g_2 \in \mathcal{G}$, we have $g_1(x,a) \leqslant g_2(x,a) + \|g_1 - g_2\|_\kappa \kappa(x)$. The monotonicity of S and (35) then imply that $Sg_1(x,a) \leqslant Sg_2(x,a) + \alpha\beta \|g_1 - g_2\|_\kappa \kappa(x)$. Switching the roles of g_1 and g_2 yields $\|Sg_1 - Sg_2\|_\kappa \leqslant \alpha\beta \|g_1 - g_2\|_\kappa$.

Since $\bar{v} = Mg^*$ and $g^* = Sg^*$ as were shown, we have $g^* = WMg^* = W\bar{v}$. To verify claim (a), it remains to show that $\bar{v}$ equals the value function v^* in (14). For all $x_0 \in \mathsf{X}$ and $\sigma \in \Sigma$, because $\bar{v} = T\bar{v}$, the definition of T implies that

$$\begin{aligned}
\bar{v}(x_0) &\geqslant r(x_0, \sigma(x_0)) + \beta \mathbb{E}_{x_0,\sigma(x_0)} \bar{v}(x_1) \\
&\geqslant r(x_0, \sigma(x_0)) + \beta \mathbb{E}_{x_0,\sigma(x_0)} \left\{ r(x_1, \sigma(x_1)) + \beta \mathbb{E}_{x_1,\sigma(x_1)} \bar{v}(x_2) \right\} \\
&= r(x_0, \sigma(x_0)) + \beta \mathbb{E}_{x_0,\sigma(x_0)} r(x_1, \sigma(x_1)) + \beta^2 \mathbb{E}_{x_0,\sigma(x_0)} \mathbb{E}_{x_1,\sigma(x_1)} \bar{v}(x_2) \\
&\geqslant \sum_{t=0}^{N} \beta^t \mathbb{E}_{x_0,\sigma(x_0)} \cdots \mathbb{E}_{x_{t-1},\sigma(x_{t-1})} r(x_t, \sigma(x_t)) + \beta^{N+1} \mathbb{E}_{x_0,\sigma(x_0)} \cdots \mathbb{E}_{x_N,\sigma(x_N)} \bar{v}(x_{N+1}) \\
&= \sum_{t=0}^{N} \beta^t \mathbb{E}_{x_0} r(x_t, \sigma(x_t)) + \beta^N \mathbb{E}_{x_0,\sigma(x_0)} \cdots \mathbb{E}_{x_{N-1},\sigma(x_{N-1})} g^*(x_N, \sigma(x_N)). \qquad (36)
\end{aligned}$$

Notice that, by Assumption 5.1(1), we have

$$\begin{aligned}
&\left| \beta^N \mathbb{E}_{x_0,\sigma(x_0)} \cdots \mathbb{E}_{x_{N-1},\sigma(x_{N-1})} g^*(x_N, \sigma(x_N)) \right| \\
&\leqslant \beta^N \mathbb{E}_{x_0,\sigma(x_0)} \cdots \mathbb{E}_{x_{N-1},\sigma(x_{N-1})} \left| g^*(x_N, \sigma(x_N)) \right| \\
&\leqslant \beta^N \mathbb{E}_{x_0,\sigma(x_0)} \cdots \mathbb{E}_{x_{N-1},\sigma(x_{N-1})} \left\| g^* \right\|_\kappa \kappa(x_N) \\
&\leqslant (\alpha\beta)^N \left\| g^* \right\|_\kappa \kappa(x_0) \to 0 \quad \text{as } N \to \infty.
\end{aligned}$$

Letting $N \to \infty$ in (36), Lemma 7.1 implies that $\bar{v}(x_0) \geqslant v_\sigma(x_0)$. Since $x_0 \in \mathsf{X}$ and $\sigma \in \Sigma$ are arbitrary, we have $\bar{v} \geqslant v^*$. Moreover, because $g^* = W\bar{v}$ and $g^* \in \mathcal{G}_1$ implies that a g^*-greedy policy σ^* exists, all the inequalities in (36) hold with equality once we let $\sigma = \sigma^*$. In other words, we have $\bar{v} = v_{\sigma^*} \leqslant v^*$. In summary, we have shown that $\bar{v} = v^*$. Hence, $g^* = Wv^*$, $v^* = Mg^*$, and claim (a) holds.

Moreover, the above arguments also imply that σ^* is an optimal policy (i.e., $v^* = v_{\sigma^*}$) if it is g^*-greedy. Because a g^*-greedy policy exists by assumption, the set of optimal policies is nonempty and claim (b) holds. To see that claim (c) holds, it remains to show that any optimal policy σ^* is g^*-greedy. Note that

$$\begin{aligned}
r(x, \sigma^*(x)) + \beta g^*(x, \sigma^*(x)) &= r(x, \sigma^*(x)) + \beta \mathbb{E}_{x,\sigma^*(x)} v^*(x') \\
&= r(x, \sigma^*(x)) + \beta \mathbb{E}_{x,\sigma^*(x)} v_{\sigma^*}(x') \\
&= v_{\sigma^*}(x) = v^*(x) = Tv^*(x) = MWv^*(x) = Mg^*(x),
\end{aligned}$$

where the first and last equalities hold since $g^* = Wv^*$, the second and fourth equalities hold because σ^* is optimal, the third equality equality holds by the definition of v_{σ^*}, and the fifth equality holds because v^* is a fixed point of T as was shown above. Therefore, σ^* is a g^*-greedy policy. We have now shown that claim (c) holds. □

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 $\|g_n - g_0\|_\kappa \to 0$ 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 $(x_0, a_0) \in \mathsf{D}$ and $y > g_0(x_0, a_0)$. Let $\varepsilon := y - g_0(x_0, a_0)$. Since κ 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, Γ 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 $h_g \leqslant (d + \|g\|_\kappa)\kappa$, 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. □

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 $g \geqslant -\|g\|$, we obtain

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

where the last equality uses (33). Since by assumption $\hat{r}$ is bounded below, so is Sg. A similar argument yields $Sg(x,a) \leqslant \beta[\hat{r}(x,a) + \|g\|]$, 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 $K \geqslant 0$ is any

constant, then for any $g \in \mathcal{G}$ we have

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

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

Let $\mathcal{V}$ be all measurable maps $v : \mathsf{X} \to \mathbb{R}\cup\{-\infty\}$ such that $(x,a) \mapsto -\frac{\beta}{\gamma}\log \mathbb{E}_{x,a}\mathrm{e}^{-\gamma v(x')}$ is in $\mathcal{G}$. Define the operators W on $\mathcal{V}$ and M on $\mathcal{G}$ respectively as

$$Wv(x,a) \coloneqq -\frac{\beta}{\gamma}\log \mathbb{E}_{x,a}\mathrm{e}^{-\gamma v(x')} \quad \text{and} \quad Mg(x) \coloneqq \sup_{a\in\Gamma(x)}\{r(x,a)+g(x,a)\}. \tag{37}$$

The original Bellman operator T and the transformed Bellman operator S can then be written as $T = MW$ and $S = WM$. Let $\mathcal{G}_1$ be the set of upper semicontinuous functions in $\mathcal{G}$. Similar to the proof of Theorem 5.1, $\bar{v} \coloneqq Mg^*$ is a fixed point of T in $\mathcal{V}$ and $g^* = W\bar{v}$. Moreover, a similar argument to the proof of Theorem 5.2 shows that $\mathcal{G}_1$ is a closed subset of $\mathcal{G}$, S maps $\mathcal{G}_1$ into itself, the unique fixed point g^* of S is in $\mathcal{G}_1$, and a g-greedy policy exists for each $g \in \mathcal{G}_1$. To see that claim (a) holds, it remains to verify $v^* = \bar{v}$.

Fix $\sigma \in \Sigma$. The definition of T and the monotonicity of T_σ in (30) imply that

$$\bar{v} = T\bar{v} \geqslant T_\sigma\bar{v} \geqslant \ldots \geqslant T_\sigma^n\bar{v} \tag{38}$$

for all $n \in \mathbb{N}$. One can show that, for all constant $\ell \in \mathbb{R}$ and $x \in \mathsf{X}$,

$$T_\sigma^n(\bar{r}+\ell)(x) = T_\sigma^n\bar{r}(x) + \beta^n\ell. \tag{39}$$

Since $\beta \in (0,1)$, letting $n \to \infty$ yields

$$v_\sigma(x) = \limsup_{n\to\infty} T_\sigma^n\bar{r}(x) = \limsup_{n\to\infty} T_\sigma^n(\bar{r}+\ell)(x)$$

for all $x \in \mathsf{X}$ and $\ell \in \mathbb{R}$. Since $\bar{v} \in [\bar{r} - \|g^*\|, \bar{r} + \|g^*\|]$, by the monotonicity of T_σ,

$$\limsup_{n\to\infty} T_\sigma^n\bar{v}(x) = \limsup_{n\to\infty} T_\sigma^n\bar{r}(x) = v_\sigma(x) \tag{40}$$

for all $x \in \mathsf{X}$. Letting $n \to \infty$ in (38) then yields $\bar{v} \geqslant v_\sigma$. Since σ is chosen arbitrarily, this implies $\bar{v} \geqslant v^*$. To see conversely $\bar{v} \leqslant v^*$, we define M_σ at each $g \in \mathcal{G}$ by

$$M_\sigma g(x) \coloneqq r(x,\sigma(x)) + g(x,\sigma(x)). \tag{41}$$

Then $T_\sigma = M_\sigma W$. Since $\bar{v}$ is a fixed point of T, $g^* = W\bar{v}$, and a g^*-greedy policy σ^* exists as were shown, we have

$$\bar{v} = T\bar{v} = MW\bar{v} = Mg^* = M_{\sigma^*}g^* = M_{\sigma^*}W\bar{v} = T_{\sigma^*}\bar{v}.$$

Hence, $\bar{v} = T_{\sigma^*}^n \bar{v}$ for all $n \in \mathbb{N}$. Letting $n \to \infty$ and then using (40) yield $\bar{v} = v_{\sigma^*} \leqslant v^*$. In summary, we have shown that $\bar{v} = v^*$. Claim (a) is now verified.

The above arguments also imply that a policy is optimal if it is g^*-greedy, and an optimal policy exits (since there exists a g^*-greedy policy). Hence claim (b) holds. To see that claim (c) holds, it remains to show that any optimal policy σ^* must be g^*-greedy, equivalently, $v^* = v_{\sigma^*}$ implies $M_{\sigma^*}g^* = Mg^*$.

To see this, because $\bar{v} = v^*$ and $g^* = W\bar{v}$, we have

$$M_{\sigma^*}g^* = M_{\sigma^*}Wv^* = T_{\sigma^*}v^* \quad \text{and} \quad Mg^* = MWv^* = Tv^*.$$

Since $v^* = v_{\sigma^*}$ and $v_{\sigma^*} = \limsup_{n\to\infty} T_{\sigma^*}^n v^*$ as shown above,

$$M_{\sigma^*}g^* = T_{\sigma^*}\left(\limsup_{n\to\infty} T_{\sigma^*}^n v^*\right). \tag{42}$$

Because $\mathrm{e}^{-\gamma \limsup_{n\to\infty} x_n} = \liminf_{n\to\infty} \mathrm{e}^{-\gamma x_n}$ for a given sequence $\{x_n\}$, the Fatou's lemma implies that, for all $x \in \mathsf{X}$,

$$\mathbb{E}_{x,\sigma^*(x)}\mathrm{e}^{-\gamma \limsup_{n\to\infty} T_{\sigma^*}^n v^*(x')} \leqslant \liminf_{n\to\infty} \mathbb{E}_{x,\sigma^*(x)}\mathrm{e}^{-\gamma T_{\sigma^*}^n v^*(x')}.$$

So $W(\limsup_{n\to\infty} T_{\sigma^*}^n v^*) \geqslant \limsup_{n\to\infty} W(T_{\sigma^*}^n v^*)$. Applying M_σ^* on both sides and then using (42), $T_{\sigma^*} = M_{\sigma^*}W$, and the definition of M_{σ^*} yield

$$M_{\sigma^*}g^* = T_{\sigma^*}\left(\limsup_{n\to\infty} T_{\sigma^*}^n v^*\right) \geqslant \limsup_{n\to\infty} T_{\sigma^*}^{n+1} v^* = v_{\sigma^*} = v^* = Tv^* = Mg^*.$$

Because $M_{\sigma^*}g^* \leqslant Mg^*$ by definition, we must have $M_{\sigma^*}g^* = Mg^*$, equivalently, σ^* is g^*-greedy. Claim (c) is verified and the proof is now complete. □

Proof of Theorem 5.4. Let $\mathcal{V}_2$ be the set of measurable maps $v : \mathsf{X} \to \mathbb{R} \cup \{-\infty\}$ such that $(x,a) \mapsto -\frac{\beta}{\gamma}\log \mathbb{E}_{x,a}\mathrm{e}^{-\gamma v(x')}$ is in $\mathcal{G}_2$. Define the operators W on $\mathcal{V}_2$ and M on $\mathcal{G}_2$ as in (37). By Jensen's inequality and Assumption 5.1(1), for all $(x,a) \in \mathsf{D}$ and $K \in \mathbb{R}_+$, we have

$$W(K\kappa)(x,a) = -\frac{\beta}{\gamma}\log \mathbb{E}_{x,a}\mathrm{e}^{-\gamma K\kappa(x')} \leqslant \beta\mathbb{E}_{x,a}K\kappa(x') \leqslant \alpha\beta K\kappa(x). \tag{43}$$

We first show that $S\mathcal{G}_2 \subset \mathcal{G}_2$. Fix $g \in \mathcal{G}_2$. Assumption 5.1(1) implies that

$$Mg(x) \leqslant \bar{r}(x) + \|g\|_\kappa \kappa(x) \leqslant (d + \|g\|_\kappa)\kappa(x)$$

for all $x \in \mathsf{X}$. Using the monotonicity of W and (43) then gives

$$Sg(x,a) = WMg(x,a) \leqslant W((d + \|g\|_\kappa)\kappa)(x,a) \leqslant \alpha\beta(d + \|g\|_\kappa)\kappa(x)$$

for all $(x,a) \in \mathsf{D}$. Hence, Sg/κ is bounded above. Moreover, a similar argument to the proof of Theorem 5.3 shows that Sg is bounded below.

To see that $Sg(x,a)$ is increasing in x, let $x_1, x_2 \in \mathsf{X}$ with $x_1 \leqslant x_2$ and fix $a \in \Gamma(x_1)$. By Assumption 5.2(1)–(2), $x_1' := f(x_1,a,\varepsilon') \leqslant f(x_2,a,\varepsilon') =: x_2'$ and $\Gamma(x_1') \subset \Gamma(x_2')$. Since in addition $r(x,a)$ and $g(x,a)$ are increasing in x, we have

$$\begin{aligned} Mg(x_1') &= \sup_{a' \in \Gamma(x_1')} \{r(x_1',a') + g(x_1',a')\} \leqslant \sup_{a' \in \Gamma(x_2')} \{r(x_1',a') + g(x_1',a')\} \\ &\leqslant \sup_{a' \in \Gamma(x_2')} \{r(x_2',a') + g(x_2',a')\} = Mg(x_2'). \end{aligned}$$

Noting that $Mg(x_i') = Mg(f(x_i,a,\varepsilon'))$ for $i = 1,2$,

$$Sg(x_1,a) = -\frac{\beta}{\gamma} \log \mathbb{E} \mathrm{e}^{-\gamma Mg(f(x_1,a,\varepsilon'))} \leqslant -\frac{\beta}{\gamma} \log \mathbb{E} \mathrm{e}^{-\gamma Mg(f(x_2,a,\varepsilon'))} = Sg(x_2,a).$$

Hence, $Sg(x,a)$ is increasing in x. We have now shown that $S\mathcal{G}_2 \subset \mathcal{G}_2$.

Next, we show that S is a contraction on $(\mathcal{G}_2, \|\cdot\|_\kappa)$. Let $h_1, h_2 : \mathsf{X} \to \mathbb{R}$ be increasing functions on X. By Assumption 5.2(1), $\varepsilon' \mapsto h_i(f(x,a,\varepsilon'))$ is increasing for all $(x,a) \in \mathsf{D}$ and $i = 1,2$. Since Assumption 5.2(3) implies that ε' is independent across dimensions, applying the Fortuin–Kasteleyn–Ginibre inequality (Fortuin et al., 1971) gives

$$\begin{aligned} \mathbb{E}_{x,a} \mathrm{e}^{-\gamma(h_1+h_2)(x')} &= \mathbb{E} \mathrm{e}^{-\gamma h_1(f(x,a,\varepsilon'))} \mathrm{e}^{-\gamma h_2(f(x,a,\varepsilon'))} \\ &\geqslant \mathbb{E} \mathrm{e}^{-\gamma h_1(f(x,a,\varepsilon'))} \mathbb{E} \mathrm{e}^{-\gamma h_2(f(x,a,\varepsilon'))} = \mathbb{E}_{x,a} \mathrm{e}^{-\gamma h_1(x')} \mathbb{E}_{x,a} \mathrm{e}^{-\gamma h_2(x')} \end{aligned}$$

for all $(x,a) \in \mathsf{D}$ whenever the expectation is well-defined. Therefore,

$$W(h_1 + h_2)(x,a) \leqslant Wh_1(x,a) + Wh_2(x,a) \tag{44}$$

for all $(x,a) \in \mathsf{D}$ and real-valued increasing functions h_1, h_2 on X. By the definition of M, for all $x \in \mathsf{X}$,

$$M(g + K\kappa)(x) = Mg(x) + K\kappa(x).$$

Because both Mg and κ are increasing on X, applying (43) and (44) yields

$$\begin{aligned} S(g + K\kappa)(x,a) &= WM(g + K\kappa)(x,a) = W(Mg + K\kappa)(x,a) \\ &\leqslant WMg(x,a) + W(K\kappa)(x,a) \leqslant Sg(x,a) + \alpha\beta K\kappa(x) \end{aligned}$$

for all $x \in \mathsf{X}$. Since in addition S is monotone, S is a contraction mapping on $\mathcal{G}_2$ as was to be shown. Moreover, because $(\mathcal{G}_2, \|\cdot\|_\kappa)$ is a Banach space, claims (1)–(3) follow immediately by the Banach contraction mapping theorem.

Let $\mathcal{G}_3$ be the set of upper semicontinuous functions in $\mathcal{G}_2$. Similar to the proof of Theorem 5.1, we can show that $\bar{v} := Mg^*$ is a fixed point of T in $\mathcal{V}_2$ and $g^* = W\bar{v}$. A similar argument to the proof of Theorem 5.2 shows that $\mathcal{G}_3$ is a closed subset of $\mathcal{G}_2$, S maps $\mathcal{G}_3$ into itself, the unique fixed point g^* of S is in $\mathcal{G}_3$, and a g-greedy policy exists for each $g \in \mathcal{G}_3$. To see claim (a) holds, it remains to verify $v^* = \bar{v}$.

To that end, we first show that (40) holds in the current setting. For each $\sigma \in \Sigma$ and $g \in \mathcal{G}_2$, let $M_\sigma g$ be defined as in (41). By the monotonicity of M_σ, (43) and (44),

$$\begin{aligned} T_\sigma(\bar{r} + K\kappa)(x) &= M_\sigma W(\bar{r} + K\kappa)(x) \leqslant M_\sigma(W\bar{r} + W(K\kappa))(x) \\ &= M_\sigma W\bar{r}(x) + W(K\kappa)(x, \sigma(x)) \leqslant T_\sigma\bar{r}(x) + \alpha\beta K\kappa(x) \end{aligned}$$

for all $x \in \mathsf{X}$ and $K \in \mathbb{R}_+$. Induction shows that, for all $x \in \mathsf{X}$, $K \in \mathbb{R}_+$ and $n \in \mathbb{N}$,

$$T_\sigma^n(\bar{r} + K\kappa)(x) \leqslant T_\sigma^n\bar{r}(x) + (\alpha\beta)^n K\kappa(x). \tag{45}$$

Because $\bar{v} = Mg^*$ and g^* is in $\mathcal{G}_2$, g has a lower bound $L \in \mathbb{R}$ and

$$\bar{r} + L \leqslant \bar{v} \leqslant \bar{r} + \|g^*\|_\kappa \kappa,$$

where both $\bar{r}$ and κ are increasing functions on X. The monotonicity of T_σ, (39) and (45) then imply that, for all $x \in \mathsf{X}$ and $n \in \mathbb{N}$,

$$T_\sigma^n\bar{r}(x) + (\alpha\beta)^n L \leqslant T_\sigma^n\bar{v}(x) \leqslant T_\sigma^n\bar{r}(x) + (\alpha\beta)^n\|g^*\|_\kappa\kappa(x).$$

Letting $n \to \infty$ yields (40). Similar to the proof of Theorem 5.3, we can show that (38) holds in the current setting. Letting $n \to \infty$ in (38) then gives $\bar{v} \geqslant v_\sigma$ and thus $\bar{v} \geqslant v^*$ since σ is arbitrary. A similar argument to the proof of Theorem 5.3 shows that $\bar{v} \leqslant v^*$. In summary, $\bar{v} = v^*$. Claim (a) is verified.

The proof of claims (b)–(c) is same to the proof of claims (b)–(c) in Theorem 5.3 and thus omitted. Therefore, all the statements of the theorem hold. □

References

Abbring, J. H., J. R. Campbell, J. Tilly, and N. Yang (2018): "Very Simple Markov-Perfect Industry Dynamics: Theory," *Econometrica*, 86, 721–735.

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. Guiso, 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.

Fortuin, C. M., P. W. Kasteleyn, and J. Ginibre (1971): "Correlation Inequalities on Some Partially Ordered Sets," *Communications in Mathematical Physics*, 22, 89–103.

Hansen, L. P. and T. J. Sargent (2008): *Robustness*, Princeton University Press.

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.

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.

Heathcote, J., K. Storesletten, and G. L. Violante (2010): "The Macroeconomic Implications of Rising Wage Inequality in the United States," *Journal of Political Economy*, 118, 681–722.

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.

Jovanovic, 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.

Kuhn, M. (2013): "Recursive Equilibria in an Aiyagari-Style Economy with Permanent Income Shocks," *International Economic Review*, 54, 807–835.

Le Van, C. and Y. Vailakis (2005): "Recursive Utility and Optimal Growth with Bounded or Unbounded Returns," *Journal of Economic Theory*, 123, 187–209.

Ma, Q. and J. Stachurski (2021): "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.

Zhu, S. (2020): "Existence of Stationary Equilibrium in an Incomplete-Market Model with Endogenous Labor Supply," *International Economic Review*, 61, 1115–1138.