Real Options Models without Single-Investment Threshold Behavior

Faninam, Farzan; Huisman, Kuno; Kort, Peter; Vera, J. C.

Publication date:
2023

Document Version
Early version, also known as pre-print

CentER

Discussion Paper

No. 2023-029

REAL OPTIONS MODELS WITHOUT SINGLE-INVESTMENT THRESHOLD BEHAVIOR

By

Farzan Faninam, Kuno J.M. Huisman,
Peter M. Kort, Juan C. Vera

30 November 2023

ISSN 0924-7815
ISSN 2213-9532

TILBURG UNIVERSITY

Real Options Models without Single-Investment Threshold Behavior

Farzan Faninam[a,*], Kuno J.M. Huisman[a,b], Peter M. Kort[a], Juan C. Vera[a]

[a]*Department of Econometrics and Operations Research, Tilburg University, Tilburg, The Netherlands*
[b]*ASML Netherlands B.V., Veldhoven, The Netherlands*

Abstract

This paper investigates real options models that violate the assumption of positive persistence of uncertainty. Without this fundamental assumption, existing methodologies are inadequate to address the firm's investment problem. To tackle this issue, we introduce a discrete-time version of a real options model and employ reinforcement learning, specifically Q-learning, to derive the optimal solution. Our findings reveal that in scenarios where the assumption of positive persistence of uncertainty is violated, the firm's investment behavior can exhibit disconnected investment regions.

Keywords: Investment analysis, Real options, Reinforcement learning

1. Introduction

In a standard real options model, the value of a single investment project for a value-maximizing firm is considered where future cash flows are uncertain in an infinite continuous time setting and the investment problem is to derive the optimal investment timing ([16], [3], [18], [6]). Central to the existing methodologies are assumptions on the structure of the value function and the stochastic process. The first assumption corresponds to *monotonicity*, which dictates that the value function of waiting for one period is monotonic in the underlying state variable. The second fundamental assumption is the *positive persistence of uncertainty* of the probability distribution of the stochastic variable: *"There is positive persistence of uncertainty, in the sense that the cumulative probability distribution $\phi(x'|x)$ of future [e.g. demand] values x' shifts uniformly to the right when the current [e.g. demand] value x increases"* ([6]). In case these two assumptions are satisfied, the optimal investment decision is guaranteed to exhibit a single-threshold behavior, i.e. there exists a clear division of the range into low and high values separated by a threshold such that *not invest* is optimal for one side of the threshold and *invest* on the other.

In some investment problems, the assumption of positive persistence of uncertainty is violated when shocks in preferences create uncertainty. This means that an increase in demand at present could signify a decrease in demand that may occur in the future ([7]). Products such as fidget spinners, hoverboards, initial versions of virtual reality headsets, and Snap Inc's Spectacles (released in 2016) illustrate a distinctive market dynamic wherein an initial increase in demand does not result in an increase in the expected future demand. Fidget spinners, for instance, experienced a swift increase in demand, only to see an equally rapid decline in consumer interest. This trajectory of demand was mirrored by hoverboards and early iterations of virtual reality headsets. In the case of Snap Inc.'s Spectacles, the product was launched to significant public anticipation and a high initial consumer demand. However, despite this promising outlook, the demand for Spectacles diminished quickly. The declining consumer interest was not predominantly due to the emergence of superior market alternatives; instead, it can be attributed to the evanescent nature of consumer trends and a decrease in product appeal. Such scenarios exemplify that an

*Corresponding author
Email addresses: f.faninam@tilburguniversity.edu (Farzan Faninam), k.j.m.huisman@tilburguniversity.edu (Kuno J.M. Huisman), kort@tilburguniversity.edu (Peter M. Kort), j.c.veralizcano@tilburguniversity.edu (Juan C. Vera)

initial surge in product demand does not necessarily entail sustained future interest, thereby challenging the assumption of positive persistence of uncertainty in demand over time.

When the assumption of the positive persistence of uncertainty fails, the uniqueness of the optimal investment threshold is not guaranteed, and the current solution methods are inadequate to address the firm's investment problem. To address the latter, in this paper, we use a discrete Markov decision process to model the firm's investment problem. A firm faces a single-investment decision where it can only invest once. It has the choice to either invest or delay at each stage as long as the firm has not invested yet. We employ a model-based reinforcement learning approach, specifically Q-learning, to solve the problem formulated by the Markov decision process. The main advantage of our proposed solution method is its flexibility in handling complex problems without imposing assumptions on the probability distribution of the stochastic variable. Although [11] applies reinforcement learning within real options models to assess investment decisions related to upgrading hydropower plant capacity, their work still relies on the assumption of positive persistence of uncertainty (see also [5], [14]). To our knowledge, the present paper is the first effort to study real options models without this fundamental assumption on the probability distribution of the stochastic variable.

The literature on strategic investment strategy in a discrete-time context is not extensive. However, [22] integrates investment timing using the real options approach with fundamental game theory and industrial organization principles to demonstrate how competition can impact project valuation. In the case of [12], they offer a strategic justification for growth options amidst uncertainty and imperfect competition in a duopoly framework. They demonstrate that within a market characterized by strategic competition, greater uncertainty may prompt investment in growth options if a significant strategic advantage is present (see also [21] and [20]). In [9], the focus is on examining how technological competition influences the dynamics of value and returns for Research and Development (R&D) in a duopoly market. The authors reveal that the value of an R&D company in a race responds differently to successes and failures and that the risk premium is significantly impacted by losing a development stage. [17] tackles the investment timing problem, wherein firms encounter a range of distinct investment opportunities that may be viewed as a collection of real options. As for [19], it delves into bilateral deals among partnerships in uncertain conditions, but with downstream flexibility. The authors demonstrate how optionality impacts the synergies resulting from a partnership in their work. In [8], the investment strategies for capacity are developed using real options. The study compares binomial and Markov models, with the conclusion that the Markov model is more reliable and yields better decision policies. According to [8], Markov models offer greater flexibility due to their independence from any assumptions about the probability distribution of stochastic variables, non-constant probabilities of variation, and the ability to generalize the binomial distribution. We explore real options models while refraining from making any assumptions regarding the probability distribution of the stochastic variable.

The primary contribution of this paper lies in formulating a discrete and finite robust real options model using a Markov decision process capable of addressing a broad range of investment problems. Unlike existing real options methodologies, our methodology does not rely on the limiting assumption of the positive persistence of uncertainty for the stochastic variable's probability distribution. This flexibility allows us to delve into the complexities of a firm's decision-making process in markets where consistent growth is not expected. Notably, our findings indicate that without the positive persistence of uncertainty assumption, the firm's investment strategy might exhibit disconnected investment regions. This underscores that there may exist situations where the firm decides to invest at a certain demand level, but intriguingly, opts not to invest when the demand is marginally higher. We employ reinforcement learning, specifically $Q-$learning, as a solution method to determine the optimal investment decision for the firm. Additionally, we provide numerical examples and economic interpretations.

The paper is organized as follows. Section 2 presents the

general setup of our model. In Section 3 we develop the solution method. An example of disconnected investment regions is given in Section 4, and the paper is concluded in Section 5.

2. General Setup

We consider a value-maximizing firm with a single investment project in the monopoly market. The firm finds the optimal investment strategy at each moment by taking a binary action to *invest* or *not invest*. The firm can only invest once in the investment project in which it receives some immediate and future revenue and incurs an immediate irreversible investment cost, i.e. once the investment is made, recovering the cost of undertaking the project is not possible. On the other hand, there is no cost and immediate revenue if the firm decides not to invest at the current stage. In our framework, uncertainty arises from the demand level such that future revenues are not certain. Therefore, the investment problem for the firm is to decide on the optimal state to undertake an investment project under uncertainty.

The demand level is the only uncertain element of the model and is denoted by the stochastic random variable y. The firm receives information about the market in which the current demand level, y, is revealed. The firm evaluates the investment opportunity by calculating its immediate and discounted expected future revenues against incurring the irreversible investment cost. We denote the investment cost of the firm by I. The firm incurs the immediate irreversible investment cost once the investment is made. If the firm invests at the current stage or has already invested in the previous stage, it receives the immediate revenue of yD, where D is a constant demand factor and the discounted expected future revenue. Let $Y = \{y_0, y_1, \dots, y_n\}$ be a finite set of demand levels where the dynamics are determined by the transition function $P : Y \to [0, 1]$. The expected future revenue of the firm is discounted by $\gamma \in [0, 1)$.

In this paper, we model the described investment problem by employing finite Markov decision processes (MDP) ([1], [10]). MDPs are a class of stochastic sequential processes that have been applied in many fields ([13], [2], [26]). The essence of the model is that the decision maker (firm) inhabits an environment that changes the state randomly in response to the actions taken by the decision maker. The states embed information about the environment which affects the immediate reward obtained by the firm, and the probabilities of future transitions ([15]). The firm aims to select actions that maximize a long-term measure of total reward.

Formally, an MDP consists of a finite set of states $\mathcal{S}$ and a finite set of possible firm actions A_s. Depending on the action taken at state $s \in \mathcal{S}$, the system is transitioned to the next state $\bar{s} \in \mathcal{S}$ with respect to the transition function $\pi : \mathcal{S} \times A \to [0, 1]$. The transition map, $\pi(\bar{s}|s, a_s)$, represents the probability that the system jumps to the next state $\bar{s} \in \mathcal{S}$ if the action $a \in A_s$ is taken at the current state $s \in \mathcal{S}$. Hence, $\pi(\bar{s}|s, a) \geq 0$ and $\sum_{\bar{s}\in\mathcal{S}} \pi(\bar{s}|s, a) = 1$. After taking action $a \in A_s$, at each state $s \in \mathcal{S}$, the firm receives an immediate reward which is denoted by the function $R : \mathcal{S} \times A \to \mathbb{R}$.

For the monopoly market, the investment problem is modeled as follows. Let $\omega \in \Omega = \{0, 1\}$ denote whether the firm has already invested, $\omega = 1$, or not invested, $\omega = 0$. We call $\omega \in \Omega$ the internal state of the firm. Our MDP has a finite set of states of $\mathcal{S} = Y \times \Omega$. In each state, we keep track of the actual demand level, $y \in Y$, and the internal state of the firm, $\omega \in \Omega$. Given that we are at state $s = (y, \omega) \in \mathcal{S}$, the firm takes action $a \in A_s$ depending on the value of the internal state $\omega \in \Omega$. If at state $s \in \mathcal{S}$ the firm has already invested (i.e. $\omega = 1$) the only option for the firm is to do nothing which is denoted by $a = 0$. If at state $s \in \mathcal{S}$ the firm has not invested (i.e. $\omega = 0$) it needs to decide whether to *invest*, represented by "1" or *not invest*, represented by "0". The set of actions for the firm at state $s = (y, \omega)$ is given by

$$A_s := \begin{cases} \{0\} & \text{if } \omega = 1 \\ \{0, 1\} & \text{if } \omega = 0 \end{cases}. \tag{1}$$

The immediate reward for the firm by taking action $a \in A_s$ at state $s = (y, \omega) \in \mathcal{S}$ is represented by the function $R : \mathcal{S} \times A \to \mathbb{R}$. If the firm has already invested (i.e. $\omega = 1$), the firm does not take action (i.e. $a = 0$), and the firm receives $R(y, \omega, a) =$

yD. On the contrary, if the firm has not invested (i.e. $\omega = 0$), the firm must make a decision from the set of possible actions (i.e. $a \in A_s = \{0, 1\}$). Therefore, the immediate reward is $R(y, \omega, a) = yD - I$ when the firm invests, and zero in case of not investing in the market. Summing up, we obtain

$$R(y, \omega, a) := (\omega + a)yD - aI, \forall y \in Y, \omega \in \Omega, a \in A_{(y,\omega)}. \quad (2)$$

When the action is taken, the internal state of the firm is updated by $\bar{\omega} = \omega + a$, and the demand level is updated according to the dynamics of the system, $P(y)$. The relation between the transition function π and the demand dynamics $P(y)$ is given by

$$\pi(\bar{y}, \bar{\omega}|y, \omega, a) = \begin{cases} P(\bar{y}|y) & \text{if } \bar{\omega} = \omega + a \\ 0 & \text{otherwise} \end{cases}. \quad (3)$$

Equation (3) implies that the probability of transitioning to a subsequent state is contingent on the internal state $\omega \in \Omega$ and the action $a \in A_s$ executed at the current state $s \in \mathcal{S}$. If the forthcoming internal state $\bar{\omega} \in \Omega$ satisfies the update rule $\bar{\omega} = \omega + a$, then the state changes according to the system dynamics $P(\bar{y}|y)$.

3. Solution Methodology: Q-learning

In reinforcement learning, the state-action value function (also known as Q-function) estimates the expected total reward obtained by taking a particular action in a given state and following a specific policy thereafter. It takes into account the current state, the chosen action, and the possible future states and rewards that result from that action ([23]). To find the firm's optimal strategy, we use the state-action value function to compute the quality of undertaking a particular action $a \in A_s$ at each state $s \in \mathcal{S}$.

3.1. Strategies

A strategy is a set of rules that an agent uses to determine which action to take in a given state to maximize its expected cumulative reward. It is essentially a mapping from states to actions, and it is the firm's way of making decisions based on its environment. A strategy can be stochastic, meaning it selects actions with a certain probability distribution or it can be deterministic, meaning it assigns a probability one to an action at each state. The goal of reinforcement learning is to learn an optimal policy that maximizes the expected cumulative reward over time. By definition, every Markov decision process has a deterministic stationary optimal policy ([4]). Therefore, we define a strategy as $\sigma : \mathcal{S} \to A$ such that $\sigma(s) \in A_s$ for all $s \in \mathcal{S}$. To evaluate a strategy σ we use the Q-function. Let $Q^\sigma(s, a)$ be defined as the discounted expected future reward by taking action $a \in A_s$ at state $s \in \mathcal{S}$, and continuing according to the policy σ in the following states. Then by definition it follows that $Q^\sigma(s, a)$ satisfies the set of linear equations

$$Q^\sigma(s, a) = R(s, a) + \gamma \sum_{\bar{s} \in \mathcal{S}} \pi(\bar{s}|s, a) Q^\sigma(\bar{s}, \sigma(\bar{s})), \forall s \in \mathcal{S}. \quad (4)$$

Given an initial state $s \in \mathcal{S}$, the firm aims to find a policy σ that maximizes the total reward (i.e. $Q^\sigma(s, \sigma(a))$). [10] shows that there exists an optimal policy σ^* for any given initial state. The optimal Q-function, $Q^*(s, a)$, can be found as a set of nonlinear equations given by

$$Q^*(s, a) = R(s, a) + \gamma \sum_{\bar{s} \in \mathcal{S}} \pi(\bar{s}|s, a) \max_{\bar{a} \in A_s} \{Q^*(\bar{s}, \bar{a})\}, \quad (5)$$

where the policy that takes an action, $\arg\max_a Q^*(y, \omega, a)$, in state $s \in \mathcal{S}$ is optimal ([1]). The Bellman optimality equation (5) is a recursive relationship that specifies the optimal action-value function given the current state and optimal policy in subsequent states. It involves considering all possible actions that can be taken from the current state, calculating the immediate reward associated with each action, and adding the discounted expected value of the next state, which is determined by the optimal policy. This process is repeated until a terminal state is reached. The resulting Q-function represents the expected total reward that can be obtained by taking a specific action in a specific state $s \in \mathcal{S}$ and following the optimal strategy thereafter ([23]).

3.2. Q-values

Considering the explained investment problem in Section 2, if the firm at some state $s \in \mathcal{S}$ has already invested in the market

(i.e. $\omega = 1$) then the firm does not make a decision (i.e. $a = 0$). Therefore, from equation (5) the Bellman equation of the monopolistic firm is given by the linear system

$$Q^*(y, \omega, a) = Q^*(y, 1, 0) = yD + \gamma \mathbb{E}_P \left[Q^*(\bar{y}, 1, 0) | y \right]. \quad (6)$$

If at state $s \in \mathcal{S}$, the firm has not invested (i.e. $\omega = 0$), then for any given $y \in Y$, the Bellman equation of the monopolistic firm is given by the linear system

$$Q^*(y, 0, a) = \begin{cases} \gamma \mathbb{E}_P \left[\max_{\bar{a} \in \{0,1\}} Q^*(\bar{y}, 0, \bar{a}) | y \right] & \text{if } a = 0 \\ yD - I + \gamma \mathbb{E}_P \left[Q^*(\bar{y}, 1, 0) | y \right] & \text{if } a = 1 \end{cases}. \quad (7)$$

The firm's optimal action at state $s = (y, \omega) \in \mathcal{S}$ for any given value of demand $y \in Y$ is *not invest* if $Q(y, 0, a = 0) \geq Q(y, 0, a = 1)$, and *invest* if $Q(y, 0, a = 0) < Q(y, 0, a = 1)$.

3.3. Reinforcement Learning

Reinforcement learning is a powerful subfield of machine learning that focuses on enabling agents to learn how to make optimal decisions through interactions with an environment. The core idea behind reinforcement learning is that an agent learns by receiving feedback in the form of rewards or penalties for its actions and adjusts its decision-making policy accordingly. Q-learning is a popular and widely used algorithm in reinforcement learning that enables agents to learn an optimal policy for any given environment by iteratively improving their estimates of the expected rewards associated with different actions. The idea behind Q-learning is introduced in [25] as a simple way for agents to learn how to act optimally in controlled Markovian domains. The foundations for reinforcement learning using MDPs and proposing the Q-learning method as the solver was laid out by the work of [24]. The firm (agent) learns the environment by receiving information through its interaction with the environment by taking actions and obtaining rewards. Therefore, the firm's reasoning is over a learning process through its interaction with the environment with the aim of maximizing the rewards it receives over the planning horizon. The firm is not constrained to a particular course of action and must explore the quality of various actions through experimentation. Mostly, the taken action affects the immediate reward, the next state which the agent will fall into, and the subsequent rewards ([23]). To find the optimal investment strategy for the firm, we can use reinforcement learning algorithms where the environment dynamics are given, and the firm needs to learn the value of actions at different states. For that, we implement a specific class of algorithms in reinforcement learning called Q-learning to learn the value of an action in a particular state. We propose to use Q-learning as a solution method for the investment problem which gives the value of taking a particular action $a \in A_s$ at state $s \in \mathcal{S}$. In Q-learning, to solve equation (5), Q-values are iterated to update

$$Q_{t+1}(s, a) = \underbrace{Q_t(s, a)}_{\text{current value}} + \underbrace{\eta_t}_{\text{learning rate}} \overbrace{\Bigg(\underbrace{R(s, a) + \gamma \sum_{\bar{s} \in \mathcal{S}}^{n} \pi(\bar{s}|s, a) \cdot \max_{\bar{a} \in A_s} Q_t(\bar{s}, \bar{a})}_{\text{new estimation}} - Q_t(s, a) \Bigg)}^{\text{learned value}}, \quad (8)$$

where a learning rate is $\eta_t \in (0, 1)$. In the update rule (8), the current value $Q_t(s, a)$ is updated by a learned value which is adjusted by a constant learning rate through experimentation to fit our model-based Q-learning algorithm. The iterative process is based on the Bellman equation, which can be viewed as a fixed point equation. The fixed point theorem provides a mathematical framework for analyzing and proving the convergence of iterative algorithms like Q-learning. By applying the fixed point theorem, we can prove the existence and uniqueness of the solution to the Bellman equation, which is the optimal action-value function. This theoretical framework is essential for understanding the convergence properties of Q-learning and other iterative algorithms. Furthermore, it helps in proving the convergence of Q-learning to the optimal policy. The Q-learning algorithm converges with probability one if the set of states and the set of actions are finite, $\sum_t \eta_t = \infty$ and $\sum_t \eta_t^2 < \infty$, and the variance of the reward function is bounded ([24]). The pseudocode for the Q-learning algorithm is given by

Algorithm 1 Q-learning: Learn function $Q : \mathcal{S} \times A \to \mathbb{R}$

Require: Demand levels $Y = \{y_0, \ldots, y_n\}$, Internal state $\Omega = \{0, 1\}$, Sates $\mathcal{S} = Y \times \Omega$, Actions $A = \{0, 1\}$, Transition probabilities $\pi(\bar{s}|s, a)$, Discounting factor $\gamma \in [0, 1)$, learning rate $\eta \in (0, 1)$

procedure QLEARNING($\mathcal{S}$, A, π, R, γ)

Initialize $Q : \mathcal{S} \times A \to \mathbb{R}_0^+$ arbitrarily

while $\Delta < \theta$ (small positive number) **do**

for $s \in \mathcal{S}$ **do**

for $a \in A$ **do**

$\bar{Q}(s, a) \leftarrow Q(s, a) + \eta\left(R(s, a) + \gamma \sum_{\bar{s} \in \mathcal{S}}^{n} \pi(\bar{s}|s, a) \cdot \max_{\bar{a} \in A_s} Q(\bar{s}, \bar{a}) - Q(s, a)\right)$

$\Delta \leftarrow \max\{\Delta, |Q - \bar{Q}|\}$

$\bar{Q} \leftarrow Q$

return Q

4. Disconnected Investment Regions

The assumption of positive persistence of demand is that the current high demand is likely to continue into the future, shifting the future distribution of demand accordingly. While this assumption is valid for many investment problems, it fails in scenarios where a spike in demand is followed by a significant decline, as highlighted by [7]. There have been historical instances of products that experienced a sudden spike in demand, such as fidget spinners, hoverboards, early versions of virtual reality headsets, and Snap's Spectacles, failing to sustain this high demand over time. These products initially attracted significant consumer interest, only to see that interest fades quickly, not necessarily due to the introduction of superior alternatives but often due to the fleeting nature of consumer trends and diminished product appeal. Given these historical instances, it is imperative for companies to consider scenarios where initial surges in demand do not guarantee sustained future demand. Our methodology is designed to account for these complex scenarios, thereby providing firms with a more nuanced, reliable tool for understanding and navigating the unpredictable dynamics of market demand over time.

The discrete-time setting involves a system of difference equations rather than the partial differential equations used in continuous-time settings, and the optimal investment timing rules can be determined using recursive algorithms or other numerical methods. The monotonicity assumption, concerning the value function of waiting, is required in both discrete time and continuous time for the existence and uniqueness of solutions. The continuous-time setting requires the monotonicity assumption to ensure the regularity of the solutions to the Hamilton-Jacobi-Bellman (HJB) equation, which is essential for proving the existence and uniqueness of solutions. In our discrete model, given that the profit $P(y) = yD$ is monotonic in the underlying stochastic variable y, the assumption of positive persistence of uncertainty is sufficient to ensure the existence and uniqueness of the investment threshold.

In the following, we consider the introduced investment problem in Section 2 and provide numerical results using the Q-learning algorithm (1). The following example illustrates the case in which the real options models are without the restrictive assumption on the stochastic process. Thus, the results demonstrate different investment threshold behavior, namely disconnected investment regions, depending on the dynamics of the Markov decision process.

4.1. An Example of Disconnected Investment Regions

In this section, we illustrate the concept of disconnected investment regions through a concrete example. Let the matrix $P_Y^{\textit{Non-Standard}}$ in the following represent the market demand in terms of transition probabilities between different demand levels. The bolded values indicate high transition probabilities, suggesting a strong inclination towards retaining the current demand level or transitioning to an adjacent level. Conversely, gray-highlighted values indicate relatively low transition probabilities, depicting minor chances of those transitions. A standout feature of this matrix is the evident diagonal trend of high probabilities, implying that for demand levels y_0 to y_5, there is a strong likelihood of demand either staying consistent or transitioning to the immediately subsequent level. For example, given the current demand is at level y_0, there is a 92% chance that the demand will either remain the same or increase slightly.

However, the matrix also exhibits some anomalies. Particularly, for demand level y_6, rather than a heightened probability

of maintaining the current demand or a slight increase, the system shows a strong 97.4% inclination to revert dramatically to the demand level y_0. This deviation from the diagonal trend suggests that upon reaching a certain demand threshold (in this case, y_6), the system is highly likely to reset to its initial state. This behavior, which is evident at y_6, signifies that the assumption of the positive persistence of uncertainty is violated. The transition matrix $P_Y^{Non\text{-}Standard}$ is given as follows

$$
\begin{array}{c} \\ y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{array}
\begin{array}{c}
\begin{array}{ccccccccc} y_0 & y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 \end{array} \\
\begin{pmatrix}
\mathbf{0.92} & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\
0.001 & \mathbf{0.929} & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\
0.001 & 0.001 & \mathbf{0.938} & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\
0.001 & 0.001 & 0.001 & \mathbf{0.947} & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\
0.001 & 0.001 & 0.001 & 0.001 & \mathbf{0.956} & 0.01 & 0.01 & 0.01 & 0.01 \\
0.001 & 0.001 & 0.001 & 0.001 & 0.001 & \mathbf{0.965} & 0.01 & 0.01 & 0.01 \\
\mathbf{0.974} & 0.001 & 0.001 & 0.001 & 0.001 & 0.001 & 0.001 & 0.01 & 0.01 \\
0.001 & 0.001 & 0.001 & 0.001 & 0.001 & 0.001 & 0.001 & \mathbf{0.983} & 0.01 \\
0.001 & 0.001 & 0.001 & 0.001 & 0.001 & 0.001 & 0.001 & 0.001 & \mathbf{0.992}
\end{pmatrix}
\end{array}.
$$

Figure 1 underscores the absence of a distinct single-threshold behavior, indicating that there is not a single cut-off point separating optimal actions. Table 1 provides a detailed representation of the firm's action-value function for each state $s \in \mathcal{S}$, contingent on the chosen action $a \in A_s$. This function clearly deviates from the monotonic condition posited in [6].

Table 1: The Q-values are calculated using the Q-learning algorithm (1) for a monopoly market. The parameter values are, $D = 2$, $I = 4$, $\gamma = 0.8$, $y = \{0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2\}$, and $P_Y^{Non\text{-}Standard}$.

Disconnected investment regions										
ω	a	y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
0	0	1.6329	1.6329	2.1951	3.6919	5.2478	6.8665	1.5239	10.6127	12.4448
	1	-1.4391	0.4491	2.3901	4.3869	6.4428	8.5615	1.3291	13.3091	15.6412
1	0	2.5609	4.4491	6.3901	8.3869	10.4428	12.5614	5.3291	17.3091	19.6412
	1	-	-	-	-	-	-	-	-	-

Let $G : Y \to \mathbb{R}$ be defined as $G(y) := \max\{0, Q(y, \omega = 0, a = 0) - Q(y, \omega = 0, a = 1)\}$ for any given value of demand $y \in Y$. A positive value of $G(y)$ means that *not invest* is the optimal action, whereas zero corresponds to *invest* as the optimal action. When the firm operates in states s_0 and s_1, the optimal choice is *not invest*. This is evidenced by the inequalities $Q(y_0, \omega = 0, a = 0) = 1.6329 > -1.4391 = Q(y_0, \omega = 0, a = 1)$ and $Q(y_1, \omega = 0, a = 0) = 1.6329 > 0.4491 = Q(y_1, \omega = 0, a = 1)$. In contrast, if the firm finds itself in states s_2, *investing* becomes the optimal choice, as illustrated by $Q(y_2, \omega = 0, a = 0) = 2.1951 < 2.3901 = Q(y_2, \omega = 0, a = 1)$. Yet, when at state $s_6 = (y_6, \omega = 0)$, the optimal strategy reverts to *not invest*. The results highlight the presence of disconnected investment regions. Contrary to the conventional single-investment threshold, there may be scenarios where a firm decides to invest when demand is at a certain level, but decides not to invest when the demand increases slightly.

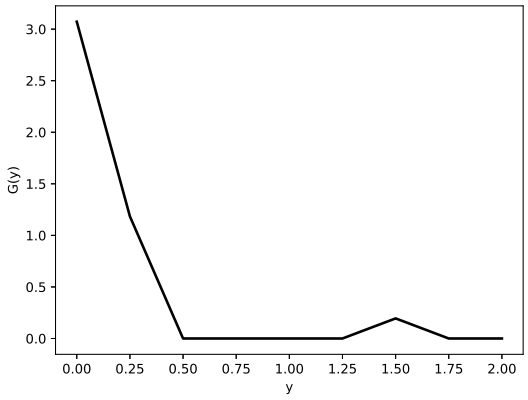

Figure 1: Disconnected investment regions. The parameter values for the function $G(y)$ are given by $D = 2$, $I = 4$, $\gamma = 0.8$, $Y = \{0, 0.25, 0.5, 0.75, 0.1, 1.25, 1.5, 1.75, 2\}$, and $P_Y^{Non\text{-}Standard}$.

5. Conclusion

Real options models can be used to assess investment projects in an uncertain economic environment. The current methodologies only tackle situations where two assumptions are met, resulting in single-investment threshold behavior. The first assumption is the monotonicity of the value function in the underlying state, and the second one is the positive persistence of uncertainty. However, investment problems may fail to satisfy the positive persistence of uncertainty. In this case, existing methods are insufficient to tackle the firm's investment problem. To overcome this limitation, we propose to model a single firm's investment problem as a finite and discrete Markov decision process and employ reinforcement learning, in particular Q-learning, to determine optimal strategies without relying on restrictive probability distribution assumptions. Our re-

search highlights the need for more flexible analytical methods in handling real-world investment complexities. When positive persistence of uncertainty is absent, a firm's optimal decisions may exhibit disconnected investment regions, diverging from conventional single-threshold behavior. Specifically, there may exist situations where the firm decides to invest at a particular demand level, but intriguingly, opts not to invest when the demand is marginally higher.

References

[1] Bellman, R., 1957. A markovian decision process. Journal of Mathematics and Mechanics , 679–684.

[2] Berninghaus, S., Seifert-Vogt, H.G., 1993. The role of the target saving motive in guest worker migration: A theoretical study. Journal of Economic Dynamics and Control 17, 181–205.

[3] Bertola, G., 1988. Adjustment costs and dynamic factor demands: investment and employment under uncertainty. Ph.D. thesis. Massachusetts Institute of Technology.

[4] Bertsekas, D.P., 1987. Dynamic Programming: Determinist. and Stochast. Models. Englewood Cliffs, NJ, US: Prentice-Hall.

[5] Caputo, C., Cardin, M.A., 2022. Analyzing real options and flexibility in engineering systems design using decision rules and deep reinforcement learning. Journal of Mechanical Design 144, 021705.

[6] Dixit, Pindyck, 1994. Investment under uncertainty. Princeton, NJ, US: Princeton University Press.

[7] Dixit, A., 1992. Investment and hysteresis. Journal of economic perspectives 6, 107–132.

[8] Fontes, D.B., Fontes, F.A., 2006. Valuing capacity investment decisions: Binomial vs. markov models. 10th Real Options:Theory Meets Practice .

[9] Garlappi, L., 2004. Risk premia and preemption in r&d ventures. Journal of Financial and Quantitative Analysis 39, 843–872.

[10] Howard, R.A., 1960. Dynamic Programming and Markov Processes. Cambridge, MA, US: The MIT Press.

[11] Kleiven, A., Nadarajah, S., Fleten, S.E., . Revisiting hierarchical planning for hydropower plant upgrades using semi-analytical policies and reinforcement learning .

[12] Kulatilaka, N., Perotti, E.C., 1998. Strategic growth options. Management Science 44, 1021–1031.

[13] Kydland, F.E., Prescott, E.C., 1980. Dynamic optimal taxation, rational expectations and optimal control. Journal of Economic Dynamics and control 2, 79–91.

[14] Lee, J.S., Chun, W., Roh, K., Heo, S., Lee, J., . Applying real options with reinforcement learning to assess commercial ccu deployment. Available at SSRN 4535371 .

[15] Littman, M.L., 2001. Value-function reinforcement learning in markov games. Cognitive systems research 2, 55–66.

[16] McDonald, R., Siegel, D., 1986. The value of waiting to invest. The Quarterly Journal of Economics 101, 707–727.

[17] Murto, P., Näsäkkälä, E., Keppo, J., 2004. Timing of investments in oligopoly under uncertainty: A framework for numerical analysis. European Journal of Operational Research 157, 486–500.

[18] Pindyck, R., 1988. Irreversible investment, capacity choice, and the value of the firm." american economic review, 78 (5), 969-985.(1991). Irreversibility, Uncertainty, and Investment," Journal of Economic Literature 29, 1110–1148.

[19] Savva, N., Scholtes, S., 2005. Real options in partnership deals: The perspective of cooperative game theory. Discussion Paper Presented at the Real Options Conference 2005, Paris .

[20] Smit, H., Trigeorgis, L., 2004. Strategic Investment: Real Options and Games. Princeton, NJ, US: Princeton University Press.

[21] Smit, H.T., 2003. Infrastructure investment as a real options game: the case of european airport expansion. Financial Management , 27–57.

[22] Smit, H.T., Ankum, L., 1993. A real options and game-theoretic approach to corporate investment strategy under competition. Financial Management , 241–250.

[23] Sutton, R.S., Barto, A.G., 2018. Reinforcement learning: An introduction. Cambridge, MA, US: MIT press.

[24] Watkins, C.J., Dayan, P., 1992. Q-learning. Machine learning 8, 279–292.

[25] Watkins, C.J.C.H., 1989. Learning from delayed rewards. King's College, Cambridge United Kingdom .

[26] White, D.J., 1993. A survey of applications of markov decision processes. Journal of the operational research society 44, 1073–1096.