Solving Schumpeterian Models of Business Process and Product Innovation with Reinforcement Learning

Max Guo
Economics 980KK
Pedro Degiovanni and Professor Robert Barro
May 2023

Abstract

The concepts of creative destruction and Schumpeterian growth implemented in recently developed mathematical models seek to interpret several phenomena related to economic growth. While these mathematical models can provide intuition and testable predictions, they are subject to constraints that limit their applicability for explaining certain observations. In particular, we demonstrate using Annual Business Survey (ABS) data that the probability of innovation across sectors follows a distribution that does not align with existing models. Moreover, the ABS also distinguishes between the notions of product and business process innovation, the latter of which is not traditionally included in Schumpeterian models. Finally, existing models often require simplifying assumptions and convenient functional specifications to obtain closed form solutions. Building off of the existing Schumpeterian models, we provide two Markov Decision Process (MDP) models that seek to endogenize both product and business innovation rate distributions across sectors. We solve the first model using dynamic programming techniques from RL, and we outline how to obtain a solution to the second model. The flexibility of MDPs and RL solution methods result in tractable solutions even with modifications to existing model assumptions.

Author's Note

I acknowledge that, due to the timeline, this project primarily represents preliminary work. I made a significant attempt at solving the second model, but ran out of time. Please see this Github repository for the code base, including brief ABS data analysis, code for specifying and solving the first model, preliminary work towards solving the second model, and miscellaneous simulation work of Barro and Sala-i Martin (2004) that did not make it into this paper.

1 Introduction

The concept of Schumpeterian growth and creative destruction is a prominent theory in economics that describes the dynamic process of economic growth and change. The ideas were first developed by the renowned economist Joseph Schumpeter in the first half of the 20th century in Schumpeter (1942). Schumpeter argued that economic growth is not just about the accumulation of capital, but also involves a continuous process of creative destruction. In Schumpeter's view, this process occurs when new and innovative technologies, products, or business models replace older ones, leading to the decline and eventual destruction of established industries and firms. Nevertheless, this destruction also creates opportunities for new firms and industries to emerge, generating innovation and growth.

Starting with Aghion and Howitt (1992) and Grossman and Helpman (1991), the idea of creative destruction and Schumpeterian growth is encoded in a number of mathematical models. These models can provide insights and predictions in a number of different areas of economics; Aghion et al. (2014) provides a survey of applications of Schumpeterian related models on several areas, including competition and innovation, firm dynamics, growth and development, and technological progress. We focus on the *quality ladder* model provided in Chapter 7 of Barro and Sala-i Martin (2004) and in Section 3 of Aghion et al. (2014). These models assume that the quality of products exist on a discrete ladder (tier list) across all sectors. The highest quality available at any time contributes to the quality frontier. Innovators, or R&D firms, may invest to improve the quality of a good on this ladder, and the arrival of innovations is modeled as a stochastic process. The incentives of innovation are given by the additional profit that comes from being at the quality frontier.

The mathematical models of Barro and Sala-i Martin (2004) and Aghion et al. (2014) provide closed form expressions for important endogenous quantities like amount of R&D investment and resulting growth rates. However, there are at least three ways in which these models may be limited. Firstly, the model in Barro and Sala-i Martin (2004) assumes a constant innovation probability across all sectors, and the models in Aghion et al. (2014) do not determine the innovation probability distribution across sectors. Based on 2015-2017 data from the Annual Business Survey (ABS), we demonstrate that the innovation probability is heterogeneous across different sectors. Adjusting either model in an attempt to describe this distribution of innovation probabilities across sectors is mathematically difficult.

Secondly, the data in ABS distinguishes product innovation from business process innovation. The relationship between business process innovation and product innovation is not investigated in recent Schumpeterian models. Intuitively, one might expect that business process innovation can follow a creative destruction process that works symbiotically with the product innovation process.

Finally, we note that it is generally difficult to solve the mathematical models in Barro and Sala-i Martin (2004) and Aghion et al. (2014) when specifying different assumptions. For example, when assuming a nonlinear relationship between expenditure and probability different than the ones in the text, the mathematical derivations become intractable. Ideally, a model could generate insights and predictions under a wide variety of functional specifications.

In this paper, we provide models and methods for solving the models that address all three aforementioned issues. We provide two Markov Decision Process (MDP) models that 1) include both product and business process innovations in a quality-ladder manners, 2) numerically endogenize both quantities to obtain innovation rate distributions that can be compared to the ABS data, and 3) are solvable under a variety of model functional specifications.

We solve both models using reinforcement learning techniques. The first model represents a single-agent reinforcement learning (RL) problem, describing the dynamics of a lagging firm attempting to catch up to the product quality frontier. The second model represents a multi-agent reinforcement learning (MARL) problem, modeling how two firms compete in a given sector. We solve the first model using dynamic programming, and we outline how to solve the second using MARL techniques.

The rest of this work is as follows. Section 2 discusses relevant background material. Section 2.1 reviews the high level ideas of Barro and Sala-i Martin (2004) and Aghion et al. (2014), while Section 2.2 introduces the field of RL and then provides some technical details on MDPs, dynamic programming, and MARL. Section 3 introduces the Annual Business Survey data and explores the distributions of product and business innovation rates. Section 4 provides our first model, its dynamic programming solution, and simulations that compare the solution with the ABS data. Section 5 provides our second model and discusses MARL techniques that could solve it. Section 6 concludes.

2 Background

2.1 Existing Models of Schumpeterian Growth

We provide a brief overview of Barro and Sala-i Martin (2004) and Aghion et al. (2014) to motivate the choices in our models. In Chapter 7 of Barro and Sala-i Martin (2004), there are three types of entities. R&D firms produce intermediate goods, final output producers buy these intermediate goods and make products, and consumers buy products from the final output producers. There are a fixed N number of sectors[1], each corresponding to one intermediate good. Intermediate goods exist on a discrete quality ladder, and the R&D firm that current produces the highest quality for a given intermediate retains a monopoly over that sector. R&D firms choose an amount to invest into research based on the expected net present value of the flow of monopoly profit. The arrival rate of quality improvements follows a Poisson process with rate parameter that depends on the amount of investment. Under a certain mathematical specification for this dependence, the resulting endogenous research investment results in a constant rate of innovation independent of the frontier quality level and sector.

In Section 3 of Aghion et al. (2014), there is a continuum of intermediate goods and a continuum of workers who can choose to work for final goods producers or work for R&D firms, which again produce the intermediate goods. We again assume a discrete quality ladder for each intermediate good. Each sector is assumed to be duopolistic and is characterized by the quality level of the leading firm and the gap between it and its competitor. The model in Aghion et al. (2014) forces the gap to be at most one, but other models (Acemoglu and Akcigit (2012) and Aghion et al. (2001)) allow for larger gaps. The model then solves for R&D expenditure for the two scenarios in which the R&D firms are at the same or different quality levels.

In particular, we note that neither model incorporates business process innovations, and neither solve for an endogenous, heterogeneous distribution of innovation rates across sectors. Moreover, the functional specifications for the dependence on expenditure (either linearly, as in Aghion et al. (2014), or based on a particular decreasing function of quality ladder level, as in Barro and Sala-i Martin (2004)) are necessary for the results, and modifications may result in mathematical intractability.

2.2 Reinforcement Learning

Now we introduce our method of choice for solving our models that address the above issues. Introduced to the machine learning community in the 1980s, reinforcement learning (RL) is a type of machine learning that allows one or more agents to learn through trial and error interactions with an environment in order to maximize a specified reward function. This is unlike common machine learning approaches which require labeled training data. Instead, RL ensures that the agent(s) update their internal policy based on the feedback it receives from the environment. RL has been successfully applied to a wide range of applications, such as game playing, robotics, and autonomous

[1] We use sector and industry interchangeably throughout this paper.

vehicles. The field of reinforcement learning is constantly evolving, with new algorithms and techniques being developed to address increasingly complex problems. Notably, the introduction of deep learning to reinforcement learning has dramatically improved the caliber of RL models, thus spawning the rapidly developing subfield of Deep Reinforcement Learning (DRL).

It is worth noting that RL models are not the same as large language models (LLM); the latter is generally categorized under supervised or semi-supervised learning instead. Thus, one must not conflate the success of LLMs such as ChatGPT and GPT-4 with the field of RL. However, the field of DRL stands on its own, *sui generis*, with several outstanding achievements in the past few decades, especially in game-playing. Two especially popular milestones include the defeat of then chess world champion Garry Kasparov against IBM's supercomputer Deep Blue in 1997 (Deep Blue won 3.5 to 2.5) and the defeat of top Go player Lee Sedol against Google Deepmind's AlphaGo program in 2016 (AlphaGo won 4-1). Both represented the first time a world-class player (champion or top-rated player) was defeated by a program. Go and chess have an enormous number of possible states, and human intuition was thought to be a critical component of masterful gameplay. Advances in RL and computational power have thus demonstrated the ability to find optimized policies that, in some cases, transcend human intuition.

There is growing interest in the applications of RL to various fields of economics, especially in game theory, finance, and behavioral economics. Atashbar and Shi (2021) provides an up-to-date survey of the applications and categorizes the applications into two areas: using the solution method to solve for an optimal policy or policy response function, or modeling bounded rationality and transition dynamics. We primarily use RL with the former in mind, as our goal is to endogenize several quantities assuming that all agents act according to optimal policies. Our work is, to the best of our knowledge, the first to implement and solve models of growth via creative destruction using RL.

2.2.1 Markov Decision Processes

The models provided in this paper are formulated as Markov Decision Processes (MDPs). An MDP is a tuple (S, A, P, R) where S is a set of states, A is a set of actions, $P(s'|s, a)$ is the probability of transitioning to state s' given that action a is taken in state s, and $R(s, a, s')$ is the immediate reward for transitioning from state s to state s' under action a. Some definitions also include the discount factor $\gamma \in [0, 1]$ as part of the definition of an MDP. γ determines the importance of future rewards relative to immediate rewards.

In general, S and A can be continuous spaces. Consider, for example, applications in robotics where the state is a location and the action is movement in a certain direction. However, continuous MDPs often require additional levels of complexity and computation in order to obtain optimal policies. For our models, we discretize S and A to obtain a finite MDP.

2.2.2 Dynamic Programming Solutions

Given a finite MDP, we can define the value function $V : S \to \mathbb{R}$ and action-value function $Q : S \times A \to \mathbb{R}$. The value function $V(s)$ represents the expected long-term reward that an agent can achieve from a particular state s onwards, by following the optimal policy. The action-value function $Q(s, a)$ represents the expected long-term reward that an agent can achieve from a particular state s onwards, by taking a specific action a and then following the optimal policy thereafter. Mathematically, these are related via the Bellman Equations:

$$Q(s, a) = \mathbb{E}[R_{t+1} + \gamma V(s_{t+1})|s_t = s, a_t = a] \tag{1}$$

$$V(s) = \max_{a \in A} Q(s, a) \tag{2}$$

where R is the immediate reward, γ is the discount factor, s_t and s_{t+1} are the current and next states, and a_t is the action taken in the current state. Value iteration is an algorithm which iteratively computes $V(s)$ using the above equations via dynamic programming. The algorithm is guaranteed to converge to the optimal values, and the optimal policy can then be derived from the optimal value function. The pseudocode is shown below.

Algorithm 1 Value iteration

```
Input: MDP with state space $S$, action space $A$, transition function $P$, reward function $R$, and discount factor $\gamma$
Output: Optimal value function $V^*(s)$ and optimal policy $\pi^*(s)$
Initialize $V_0(s)$ arbitrarily for all $s \in S$; for $k = 0, 1, 2, \ldots$ do
    for each $s \in S$ do
        $V_{k+1}(s) \leftarrow \max_{a \in A} \sum_{s' \in S} P(s'|s,a)\,(R(s,a,s') + \gamma V_k(s'))$;
    end
    if $V_{k+1}(s) \approx V_k(s)$ for all $s \in S$ then
        Set $V^*(s) = V_{k+1}(s)$ and $\pi^*(s) = \arg\max_{a \in A} \sum_{s' \in S} P(s'|s,a)\,(R(s,a,s') + \gamma V_k(s'))$ for all $s \in S$
        break
    end
end
```

2.3 Multi-agent Reinforcement Learning

Multi-agent reinforcement learning (MARL) is a field within RL that deals with multiple agents interacting, possible simultaneously, in the same environment. Each agent may have a different reward function, resulting in a MARL problem that lies somewhere on the spectrum between pure cooperation and pure competition. The formulation of a Schumpeterian Growth model as a MARL problem is motivated by the duopolistic competition in Aghion et al. (2014), which can be adapted to a MARL problem with two agents attempting to maximize their own profits. Because we only use two agents, we describe the algorithm for which we can solve our formulation of the problem.

2.3.1 Minimax Q-learning

The minimax Q-learning algorithm[2], introduced by Littman (1994), is an RL algorithm designed to deal with environments in which an agent faces an adversarial opponent who seeks to minimize the agent's rewards. Minimax Q-learning is based on the traditional Q-learning algorithm, a single-agent RL algorithm that uses action-values [3] to learn of an optimal policy while interacting with the environment. Unlike value iteration, both algorithms assume no knowledge of the environment except for access to all possible actions at every given state and internal documentation of every visited state. In particular, transition probabilities are unknown, and the agent(s) simply move around the environment according to the MDP and keep track of the state-action-new states they encounter.

In settings of two agents, traditional Q-learning can fail because the opponent can actively work against the agent. The minimax Q-learning algorithm works by assuming that the opponent is trying to minimize the agent's reward, and thus it computes the worst-case expected reward for each action in a given state. The algorithm then selects the action that maximizes the minimum expected reward over all possible opponent strategies. By doing so, the agent is able to learn a policy that is robust against the opponent's actions. While the convergence of the algorithm under general settings is an open research problem, the minimax Q-learning algorithm has been shown to perform well in several adversarial games and problems, including chess, checkers, and robot navigation in adversarial environments.

[2] We do not include the full pseudocode in this paper (see Littman (1994)).

[3] Hence the name *Q*-learning

3 Annual Business Survey Data

In this section, we describe 2017-2019 innovation data from the Annual Business Survey (ABS) (National Science Foundation (2021)). The questions on the ABS follow guidelines in the OSLO Manual (OECD/Eurostat (2018)), which is provided by the OECD and Eurostat (the Statistical Office of the European Union). We argue that the data demonstrates a distribution of innovation probabilities that is not adequately explained by existing Schumpeterian models of innovation. Moreover, existing models do not incorporate the idea of business process innovations, a core part of the innovation survey results.

For the purpose of this paper, we utilize a subset of the data containing the following fields: Industry, Number of Companies, Percentage of Industry with Product or Business Innovation, Percentage of Industry with Product Innovation, and Percentage of Industry with Business Innovation. There are a total of 4,857,473 U.S. Companies represented in the survey, and the total number of industries we use is 36. To obtain this number, we chose to refrain from using more granular data that divided certain industries into sub-industries (for instance, we would treat "Chemicals" as one industry as opposed to treating "Pesticide, fertilizer, and other agricultural chemicals", "Pharmaceuticals and medicines", etc. as their own industries).

Innovation, as defined by the OSLO manual, is:

> The introduction of new or improved products (goods or services) or business processes that differed significantly[4] from the business's previous products or processes.

As for business innovation, Eurostat further categorizes business functions into either core business functions or support business functions. Core business functions involve primary, income-generating activities, such as the production of goods and services. Support business functions are secondary activities that assist core business functions, including logistics, marketing, sales, information technology, administrative functions, and engineering and related technical services.

In our upcoming models, we assume that only product quality matters for profit flow. Our model in Section 4 treats business process innovation as improvements in the efficiency of product innovation. Under this framework, a company only engages in business process innovation insofar as it speeds up product innovation.

3.1 Distribution of Innovation Rates

We are interested in the distribution of innovation probabilities across sectors for both product and business process innovations. To obtain these distributions from the data, we make a few simplifying assumptions. First, we assume that the percentage of companies in an industry with at least one innovation in the three years is equivalent to the innovation rate in that sector. This assumption would not be valid if there was a significant number of firms with more than one innovation in the three years. However, given the relatively low percentage of firms with at least one innovation, we think this is unlikely. Thus, assuming the length from 2015-2017 is a single unit of time, we can use the phrases "innovation rate" and "innovation probability" interchangeably.

Moreover, we assume that all product innovation is an increase in the quality level of an existing sector's frontier quality. This assumes that all of the innovations are improvements in quality as opposed to completely new categories of products, which require much more novelty. This assumption is supported by the existence of other data in the

[4] This definition of innovation supports the discretized quality ladder approach in Barro and Sala-i Martin (2004) and Aghion and Howitt (1992), since there must be a "significant" improvement from one quality level to the next.

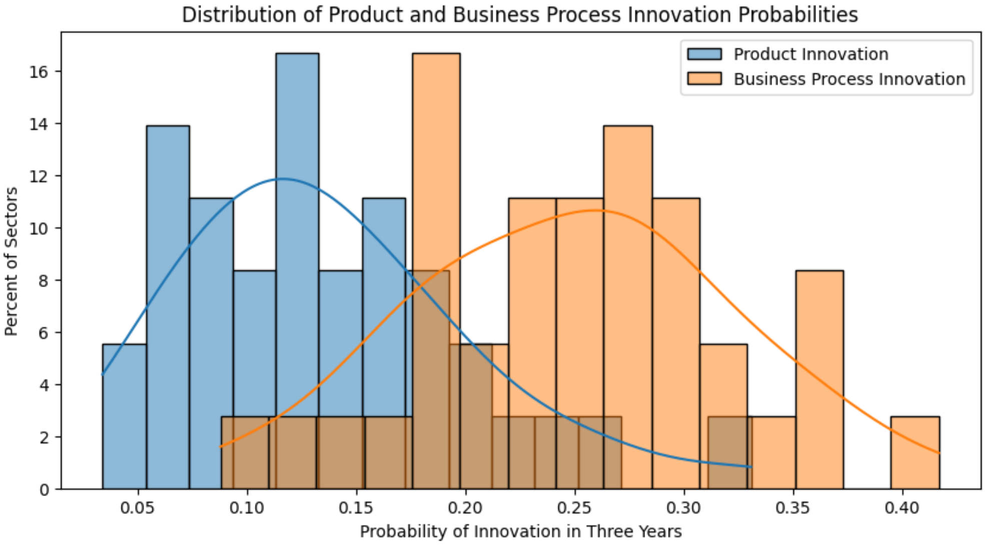

Figure 1: Distribution of Innovation Probabilities

Product: Mean 0.136, Median 0.125, St.dev 0.065

Business Process: Mean 0.249, Median 0.254, St.dev 0.072

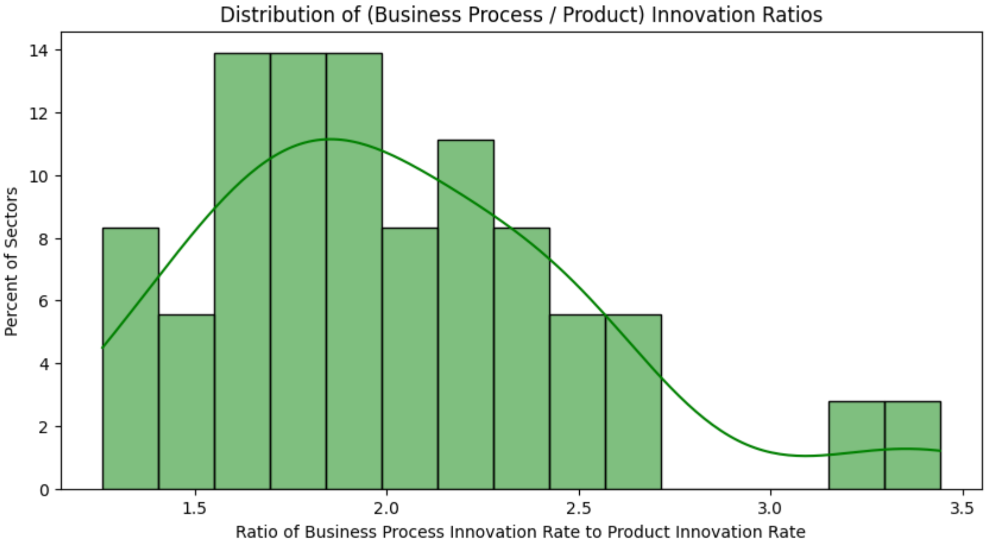

Figure 2: Distribution of Ratio of Innovation Probabilities

Mean 2.02, Median 1.92, St.dev 0.50

ABS survey that demonstrates that innovations consisting of products new to the market are generally more sparse than products only new to the business, indicating that larger degrees of novelty are harder to come by.

Under these assumptions, we can plot the distribution of product and business process innovation rates across sectors in Figure 1. Moreover, we plot the distribution of the ratio of the probability of business process innovation to the probability of product innovation across all industries in Figure 2. We see that the probability of product innovation is not constant across sectors, as specified in Barro and Sala-i Martin (2004). Moreover, business process innovation probability is almost always higher than product innovation probability. Our goal in the following sections is to come up with Schumpeterian models that have solutions explaining these phenomena.

4 Single-Agent Model for Business and Product Innovation

In our first model, we assume that, for any given sector, there are a lot of laggard firms and one leading firm. In particular, given the enormous number of firms in the ABS data, we can essentially assume that all of the firms in the data are laggard firms, so the distributions of product and business process innovation from the previous section is solely due to laggard firms. Thus, our single-agent RL model captures the scenario of a laggard firm catching up to the product quality frontier. We assume that business process innovation only matters insofar as it helps the firm with product innovation. Our state space can thus be represented as a grid of quality points, with product quality and business process quality on the y and x axes. The firm wishes to reach the top rung of product quality, as it then begins to receive great rewards from the monopoly profit flow, and its actions are various levels of expenditure on either product or business R&D innovation.

4.1 Model specification

Let $[N] = \{0, 1, \ldots, N-1\}$ for positive integer N. We fix P to be the number of product qualities, and B to be the number of business process qualities. Denote the product and business process quality levels as κ and ι, respectively, where $\kappa \in [P]$, $\iota \in [B]$. Thus, the highest quality level of product is $P-1$, and the highest quality level of business process is $B-1$. We define our set of states as:

$$S = \{(\kappa, \iota) : \kappa \in [P], \iota \in [B]\}. \tag{3}$$

Additionally, we specify any state where $\kappa = P - 1$ as a goal state, so that any firms that land on such a state terminate their gameplay. In terms of the Schumpeterian model, we can think of the firm as having caught up to the quality frontier.

Let $L \geq 2$ be an integer specifying the degree of discretization of our actions, so higher levels of L better approximate a continuous action space. We require that our firm spends a positive amount on R&D in at least one sector, but the sum of the total R&D expenditure must not exceed 1, which we treat as the “budget" per step. The requirement that the expenditure is less than a certain numerical constant is not as constraining as it seems. Given enough computational power and discretization, we could arbitrarily set the upper bound of R&D expenditure in the most general form of the model. As for the positive requirement, we note that positive R&D expenditure is required (see the transition function below) to attain a goal state, and we are only concerned with firms actively seeking to attain the quality frontier.

Thus, our set of actions is:

$$A = \left\{ \left(\frac{i_p}{L-1}, \frac{i_b}{L-1} \right) : 0 < i_p + i_b \leq L-1,\ i_p \in [L], i_b \in [L] \right\}, \tag{4}$$

where $\frac{i_p}{L-1}$ and $\frac{i_b}{L-1}$ are the fraction of the budget that we spend on product and business process research innovation, respectively. Note that $|A| = \frac{L(L+1)}{2} - 1$, so the number of actions increases quadratically with the degree of discretization.

The transition function $P(s'|s,a)$ is described as follows. Given state $s = (\kappa, \iota)$ and action $a = \left(\frac{i_p}{L-1}, \frac{i_b}{L-1}\right)$, then $p_p = f_p\left(\frac{i_p}{L-1}\right)$ and $p_b = f_b\left(\frac{i_b}{L-1}\right)$ return the independent probabilities of transitioning $\kappa \to \kappa+1$ and $\iota \to \iota+1$, assuming we are in an interior state. Thus, there is probability $1 - p_p p_b$ of staying in the same state, $p_p p_b$ probability of transitioning to $(1+\kappa, 1+\iota)$, $(1-p_p)p_b$ probability of transitioning to $(\kappa, 1+\iota)$, and $p_p(1-p_b)$ probability of transitioning $(\kappa+1, \iota)$. If we are not in an interior state, this implies that we either have $\kappa = P-1$, in which case there are no more transitions, or we have $\iota = B-1$, in which case there are no more business process inventions to be made.

We posit the following formulas for the probabilities of transitions in product quality or business process quality:

$$f_p(x) = \min\left(\frac{1}{\zeta_p} x^\alpha (1+\iota), 1\right) \tag{5}$$

$$f_b(x) = \frac{1}{\zeta_b} x^\alpha \tag{6}$$

where $1 \leq \zeta_p, \zeta_b$ represent costs to research and $0 < \alpha < 1$ results in diminishing returns to more research expenditure. In particular, the $(1+\iota)$ multiplicative factor in f_p highlights the benefits of business process quality. The benefit of higher business process quality to an firm is through raising the probability of product quality innovation dramatically for the same amount of product R&D expenditure.

Finally, the firm receives negative reward $R(s,a,s') = -\frac{i_p}{L-1} - \frac{i_b}{L-1} - c$, where again $a = \left(\frac{i_p}{L-1}, \frac{i_b}{L-1}\right)$, and c represents some small fixed cost. In the special case in which the firm reaches a goal state $s' = (P-1, \iota)$ for some $\iota \in [B]$, the firm also receives a large positive reward $R >> 0$. The interpretation behind this is that the firm receives no profit flows if it is not on the frontier of knowledge (consistent with the monopoly rights assumption from Barro and Sala-i Martin (2004)). However, it begins to receive profit flows when it reaches the frontier, and we can take R to be the net present value of all the profit flows after reaching that state. Meanwhile, we represent all R&D costs as negative rewards. In addition, we include a small fixed cost to represent costs of outside operations not involving R&D and provide more incentive for the firm to reach a goal state faster.

4.2 Solving the Model with Dynamic Programming

For our simulation, we let $P = 5$ and $B = 15$. We chose this value of B in relation to P because the optimal policy no longer attempts to perform any business process innovations beyond $B = 15$. The firm's goal states are where $\kappa = 4$. We let $L = 11$, so the available fractions of R&D expenditure are in tenths. For f_p and f_b, we set $\zeta_p = 50$, $\zeta_b = 5$, and $\alpha = 0.2$. We set the fixed cost $c = 1$, and the final reward for reaching the goal state as $R = 100$.

We use value iteration to solve our MDP model. We set $\gamma = 0.9$, and our stopping condition for value iteration as when the maximum absolute difference between $V_k(s)$ and $V_{k+1}(s)$ is less than $\frac{\epsilon(1-\gamma)}{\gamma}$, $\epsilon = 10^{-4}$. The optimal policy in each state is shown in Figure 3. We see that the optimal policy almost forms "contours" that extend diagonally from the start state. At the beginning, the optimal policy is to put a small amount of research into business process innovation. As business ladder position increases, it becomes increasingly important to put more funding into product R&D innovation because of the increase in innovation probability due to the business process factor. As product ladder position increases, more expenditure occurs in order because the reward from the goal state is discounted less.

Intuitively, these results can be interpreted as follows. For a firm that is currently far from the frontier, its attempts to reach the frontier are unlikely to immediately result in product research investment. Instead, it will first invest

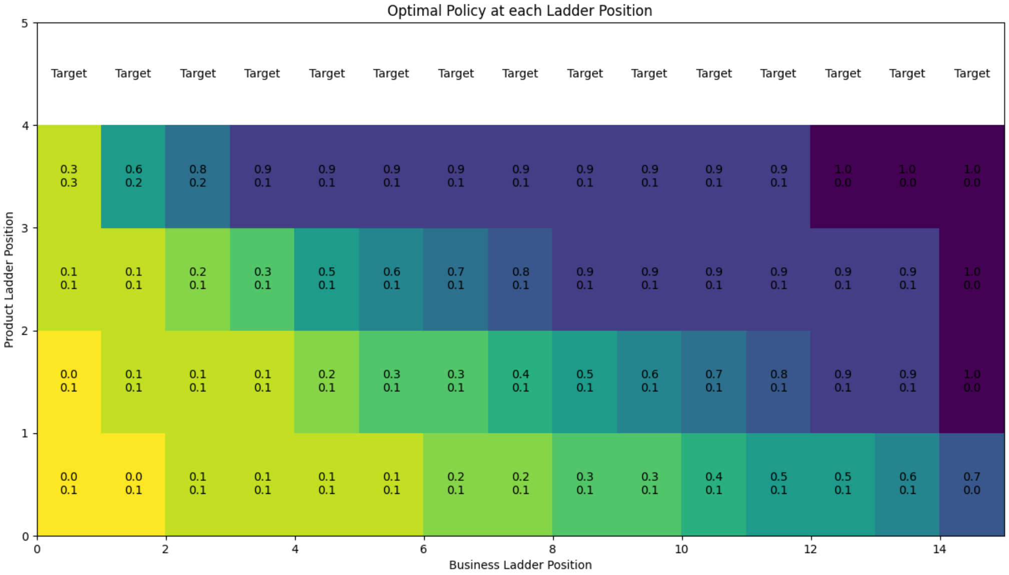

Figure 3: Optimal Policy found via Dynamic Programming. In each state, product expenditure is the top number, and business process expenditure is the bottom number. The squares are colored based on the difference in expenditure between product and business process expenditure. The darker the color, the higher product innovation expenditure is relative to business process expenditure.

in improving the quality of its business processes. As its business processes improve, it will slowly allocate more investment to product investment research because the business process quality allows the firm to advance quicker along the product quality ladder. Moreover, as it gets closer and closer to the frontier, it will invest more into product research in an attempt to rapidly attain the quality frontier.

4.3 Simulations

After finding the optimal policy, we run an 1000 episodes of the game from start ($(\kappa, \iota) = (0, 0)$) to finish ($\iota = P - 1$). For each episode, we calculate the number of product and business process innovations and divide by the number of iterations in the episode to obtain the aggregate product/business process innovation rate in that episode, and the distributions are shown in Figure 4. For each episode, we also calculate the ratio of the business process to product innovation rates, and this distribution is shown in Figure 5.

Our simulation results display, roughly, a similar flavor as that of Figure 1 and Figure 2. We see that our simulation successfully results in a business process innovation rate that is generally larger than the product innovation rate, and the distributions also display significant overlap. While the x-axis displayed is not the same (i.e. the ABS distribution has higher probability of innovations), we note that the time steps in the simulation do not specify units; doing so should be able to shift the simulated distributions to match the ABS distributions. The simulated and ABS survey results also demonstrate slightly right-skewed distributions for both types of innovation. Just as how a small proportion of industries have a high relative innovation probability, a small proportion of firms end up with a higher average aggregated innovation rate before reaching the frontier.

However, there are a few salient differences between the ABS survey and simulated results. First, the simulated results

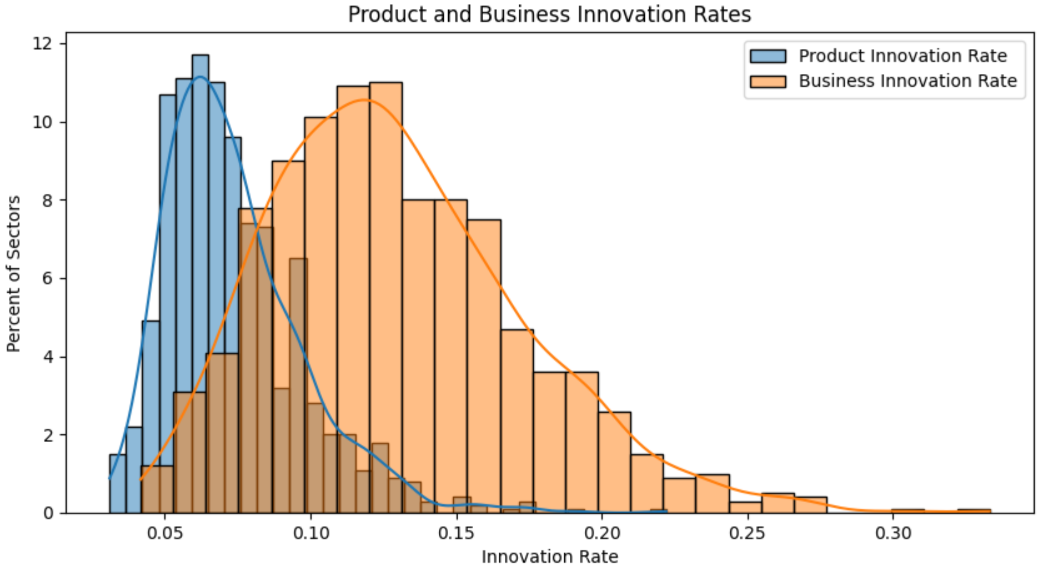

Figure 4: Distribution of Innovation Probabilities

Product: Mean 0.0735, Median 0.069, St.dev 0.024

Business Process: Mean 0.1295, Median 0.125, St.dev 0.044

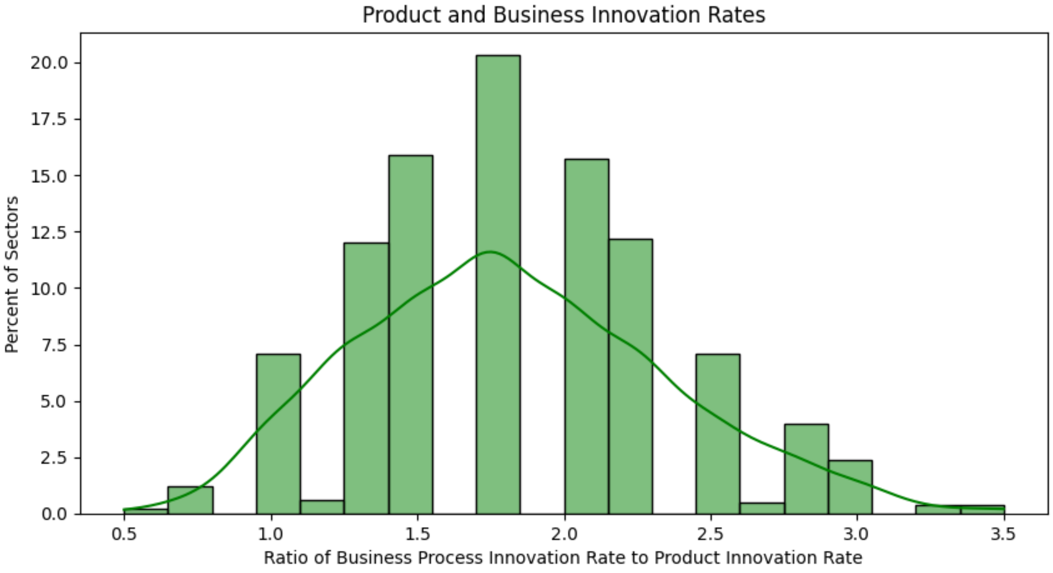

Figure 5: Distribution of Ratio of Innovation Probabilities

Mean: 1.821, Median 1.75, St.dev 0.525

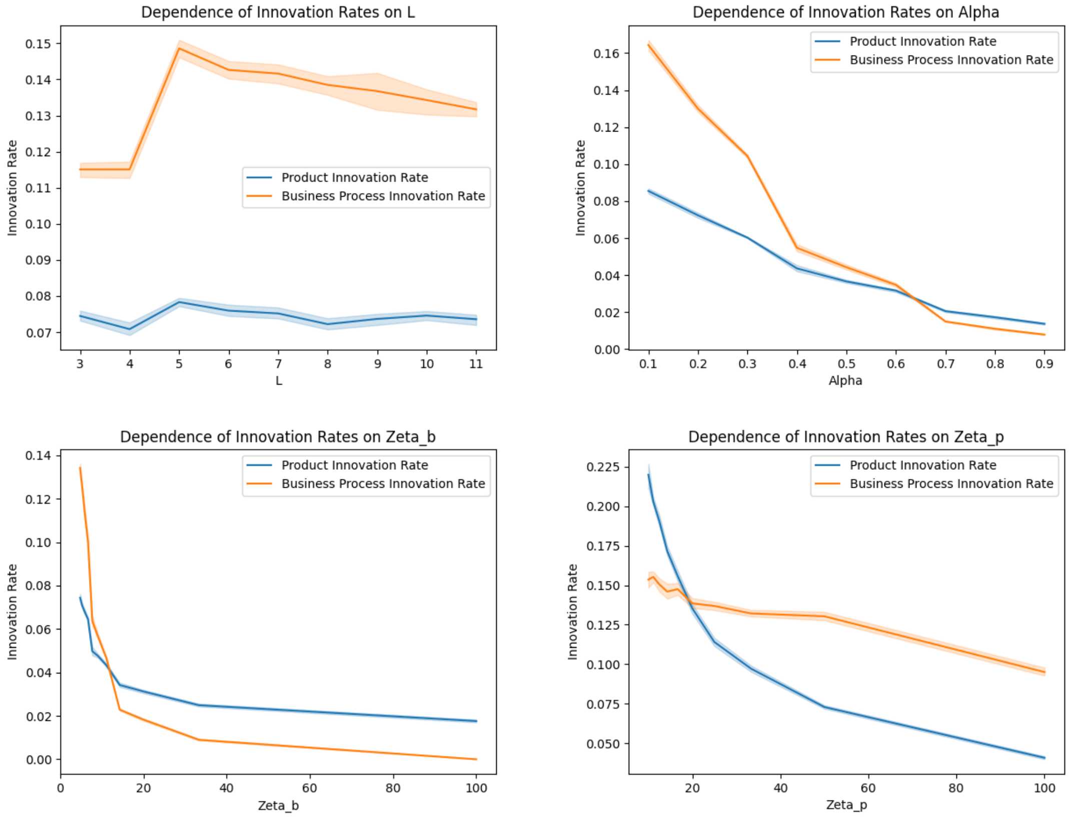

Figure 6: Dependence of Innovation Rate on Parameter Values. For each experiment, we hold all other parameters as the constants specified in Section 4.2 and vary the only the one parameter. We do 10 bootstraps trials to obtain the error bars, each time taking the mean of the innovation rates from 100 episodes.

in Figure 4 display a much heavier tail than Figure 1. In fact, one could potentially argue that the latter would not be skewed if not for a few outlier industries, whereas the simulation demonstrates considerable skewness. This could be an issue with the data, and more granularity within sectors may demonstrate the existence in the data of a heavier skewness or, potentially, no skewness at all. Another issue could be with the model, which allows for lucky firms to reach the goal state quickly and thus display a high average innovation rate. Currently, there is no penalty for moving up the product quality ladder. However, as in Barro and Sala-i Martin (2004), we may construct a model in which the probability of moving up the product quality ladder is proportional to a decreasing function of the current product quality level. This may reduce the "lucky firms" phenomena and decrease the heaviness in the tail.

Comparing the figures for ratio of business process to product innovation rates, it appears that the opposite phenomena occurs. Though there is similarity in the statistics of the simulated and ABS distributions, the simulated distribution in Figure 5 appears to be more symmetric than the results in Figure 2. In the ABS data, an industry with a high business product innovation ratio is more likely to have a average or lower product innovation rate than an industry with a low business product is to have a higher product innovation rate. However, the simulated results demonstrate that, roughly speaking, equal numbers of firms get very lucky/unlucky with business process/product innovations as unlucky/lucky.

Though our simulated results mostly match well with the overall shape of the ABS distributions, one concern is how the choice of parameter values $(P, B, L, \alpha, \zeta_b, \zeta_p)$ in our model affects the results displayed. We believe that reasonable values of P, B do not affect our results because, for any fixed value of P, we know that there exists a threshold value of B above which the results are not affected, because the firm no longer spends any amount on business process innovation above the threshold; consider that the optimal business expenditure is 0 for $\iota = 14$ in Figure 3. We also see from the L figure in Figure 6 that our model is relatively robust to different values of discretization.

However, the same cannot be said for the other parameter values, which can be a limitation to this model. Both product and innovation rates depend heavily on α, ζ_b, and ζ_p, and for each parameter it is possible to find a value such that the mean business innovation rate actually drops below the product innovation rate. We note that business process innovation rates are more sensitive to ζ_b than product innovation rates, but that product innovation rates still respond to changes in business process research costs. The other way around is true for ζ_p. α is also critical for determining the relative means; larger α means less diminishing returns to research, leading to more direct product innovation and higher overall product innovation rates relative to business process innovation.

Despite the sensitivity of the simulated results due to changes in parameter values, we note that this model is capable of obtaining the optimal policy for any valid transition function. In particular, the modifications that would result in formidable mathematical optimization problems in existing Schumpeterian models pose no issue here. For instance, we can find an optimal policy for any function f_p and f_b, allowing full flexibility in specifying the form of innovation probability.

5 Multi-Agent Model for Business and Product Innovation

We now provide a two-firm model that builds off of Section 4 and Aghion et al. (2014). We assume that we have a duopoly attempting to maximize profit flow (formulated as rewards in the MDP). The state space is a 4-tuple describing the current quality for both players, for both product and business process innovation. The actions available for both players remains as the amount of expenditure devoted to product or business innovation. The transition probability function is identical for each firm. The biggest modification is in the reward structure: positive rewards are given out every turn as opposed to only at the end, but the rewards are only given to the firm at the product quality frontier, or split in some fashion if tied. This mimics the monopoly flow of profits for the firm at the product quality frontier from Barro and Sala-i Martin (2004).

5.1 Model specification and Proposed Solution

We adapt almost all notation from Section 4. Let P and B be the number of product and business process qualities again. Technically, we might think of this MDP as infinite, since the quality improvements could theoretically go on forever, but we assume finiteness for the sake of computational tractability. The set of states is then:

$$S = \{(\kappa_1, \iota_1, \kappa_2, \iota_2) : \kappa_1, \kappa_2 \in [P], \iota_1, \iota_2 \in [B]\}. \tag{7}$$

where the subscript denotes the qualities associated with each firm. We can specify any state in which $\kappa_1 = P - 1$ or $\kappa_2 = P - 1$ as a terminal state, perhaps with the philosophical interpretation that at some point the entire industry becomes obsolete. The set of actions available to each firm is identical to our first model, described by Equation (4). Similarly, the transition function for each firm is identical to the transition function for the first model.

The reward structure is the biggest difference between this model and the first model. We continue to give negative reward to each firm based on their actions, but we now add on a positive reward for being on the frontier. We can write our reward function for firm i (denoting the opposing firm as firm j) as:

$$R_i(s,a,s') = -\frac{i_p}{L-1} - \frac{i_b}{L-1} - c + g(\kappa_i', \kappa_j') \tag{8}$$

$$g(\kappa_i, \kappa_j) = \begin{cases} M & \kappa_i > \kappa_j \\ M(1-\Delta) & \kappa_i = \kappa_j \\ 0 & \kappa_i < \kappa_j \end{cases} \tag{9}$$

where $a = \left(\frac{i_p}{L-1}, \frac{i_b}{L-1}\right)$ is firm i's expenditure, (κ_i', κ_j') are the product quality levels of firms i and j in state s', c is a fixed constant, M represents monopoly reward, and $0 \le \Delta \le 1$ represents the competitiveness in an industry (idea based on Aghion et al. (2014)). We also remove the additional reward given in Section 4 for reaching the terminal state, since there monopoly profits are already incorporated in every step.

We propose solving this model using the minimax Q-learning algorithm because, barring edge cases for Δ, maximizing one's own reward is equivalent to minimizing the opposing firm's reward. Under ideal circumstances, this algorithm would provide optimal policies for each firm that should be symmetric, which we could plot similarly to Figure 3. We could also determine endogenously the product and business innovation distributions across sectors, treating different episodes of the game as different sectors.

As in the previous model, we expect minimax Q-learning to return a solution regardless of the functional specifications for the transition probabilities and the rewards. However, unlike value iteration, we do not know the guarantees of convergence, so it is possible (though unlikely, given the widespread application of this algorithm) that we run into tractability concerns. Further implementation and testing work is required to determine if this is the case.

6 Conclusions

In conclusion, we present two Markov Decision Process (MDP) models that follow ideas of creative destruction and Schumpeterian growth theory that demonstrate a distribution for product and business process innovations across different sectors. We solve the first model, a single agent model, using dynamic programming, and visualize the optimal policy. For the second model, we discuss a possible multi-agent reinforcement learning solution. Finally, we compare the simulation results from the first model with real product and business process innovation data from 2015-2017 Annual Business Survey data and discuss its benefits and drawbacks.

We acknowledge the following limitations and avenues for future work.

- Business processes serve many functions, and the models currently specify its interaction with product innovation in a very particular manner. In particular, it could be the case that modeling product innovation with reduced costs at higher levels of business process interaction would result in a more accurate model. Future work may involve specifying and solving this model and comparing the results with the ones in Section 4.
- The ABS data is very coarse-grain, and the assumptions made in order to obtain a distribution of product and business process innovation rates across sectors may not be particularly accurate. Data at the level of smaller sectors, and data that distinguishes what types of product or business process innovation occurs, may be helpful for future analysis.
- Solving Section 5 and identifying the characteristics of the optimal solution (assuming that minimax-Q finds the optimal solution) is the next direct line of work. This may also shed light on the sensitivity issues identified in Section 4.

References

Robert J. Barro and Xavier Sala-i Martin. *Economic Growth*. MIT Press, Cambridge, MA, 2004. ISBN 9780262025539.

Joseph A Schumpeter. *Capitalism, socialism and democracy*. Routledge, 1942.

Philippe Aghion and Peter Howitt. A model of growth through creative destruction. *Econometrica*, 60(2):323–351, 1992. doi: 10.2307/2951599.

Gene M. Grossman and Elhanan Helpman. *Innovation and Growth in the Global Economy*. MIT Press, Cambridge, MA, 1991. ISBN 9780262071277.

Philippe Aghion, Ufuk Akcigit, and Peter Howitt. What do we learn from schumpeterian growth theory? In *Handbook of economic growth*, volume 2, pages 515–563. Elsevier, 2014.

Daron Acemoglu and Ufuk Akcigit. Intellectual property rights policy, competition and innovation. *Journal of the European Economic Association*, 10(1):1–42, 2012.

Philippe Aghion, Christopher Harris, Peter Howitt, and John Vickers. Competition, imitation and growth with step-by-step innovation. *Review of Economic Studies*, 68(3):467–492, 2001.

Hamed Atashbar and Tianyuan Shi. Deep reinforcement learning: Emerging trends in macroeconomics and future prospects. Technical Report WP/21/116, International Monetary Fund, 2021. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4313684#.

Michael L Littman. Markov games as a framework for multi-agent reinforcement learning. In *Machine learning proceedings 1994*, pages 157–163. Elsevier, 1994.

National Science Foundation. National center for science and engineering statistics (ncses) - publications and data products. https://ncses.nsf.gov/pubs/nsf23310, 2021. Accessed: May 7, 2023.

OECD/Eurostat. *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation*. OECD Publishing, Paris, 2018. ISBN 9789264304604. URL https://doi.org/10.1787/9789264304604-en.