Destructive Innovation and Productivity Dispersion

Yang Ming

August 2020

Abstract

The dramatic rise of productivity dispersion in the U.S. since the 90s can be explained by the emergence of destructive innovations. Empirically, I find these innovations persistently destroy the rival firms' sales, and leads to hump-shaped impulse response of productivity dispersion. I propose a technology-ladder model to invesigate the channel through which destructive innovation affects productivity dispersion. The model predicts that the firm with low R&D cost exerts higher R&D effort when there is no destructive innovation shock; upon the shock, the low-cost firm responds by raising its R&D effort, and the high-cost firm by lowering. Therefore, the greater difference in R&D transforms into a hump-shaped impulse response of the technology gap. Empirical tests using U.S. publicly listed firms as sample confirms the heterogeneous responses of R&D as implied by my theory.

1 Introduction

Since around 1990, there has been a dramatic and persistent rise in productivity dispersion across firms in the U.S., Canada and other OECD countries. This is documented by a series of recent researches[1] using different empirical strategies and datasets and has become a stylized fact. As it is now clear that this upward trends in productivity dispersion in these developed economies cannot be accidental or classified as some one-time events, economists have started the inquiry of its cause, this paper is one of these efforts.

Having observed the coincidence of the time series of productivity dispersion and the market value of major innovations[2] in the U.S. economy, I propose a theory in which the arrival of major innovations contributes to the dispersion of firm productivity. Whenever a firm achieves high-valued, major innovations, it brings negative impact on the optimized profit of its competitors, which has its renowed name as creative destruction since Schumpeter (1942). In response to such shock, the affected firms reoptimize their R&D strategies. If the direction of the adjustment of R&D is conditional on pre-shock R&D, in a way that firms with high pre-shock R&D responds to the destructive innovation shock by increasing R&D effort, and the low pre-shock R&D firms responds in the opposite fashion, then cross-firm R&D efforts will be more dispersed as a result of the destructive innovation. This will transform into a larger productivity dispersion. In this paper, I provide a theoretical framework to reproduce the above mechanism, and present empirical evidence in support of it.

Overview

I start by an empirical investigation of the impact the high-valued innovations to the sales of competing firms. I detrend the firm-year market value of patents provided by Kogan et al. (2017), and obtain the 95% percentile for each industry. Any firm whose yearly new patents' value exceeds its industry-specific 95% percentile is said to be a major innovator in that year[3]. An major innovation shock is defined as the event that there is at least one major innovator in an industry-year cell. Using the local projection method developed by Jordà (2005), I identify the impulse response function of firm-level and industry-level sales to the major innovation

[1] For the empirical evidence on the increasing productivity dispersion in the U.S., see Kehrig (2015), Barth et al. (2016) and Decker et al. (2018). For that of Canada, see Gouin-Bonenfant (2020). Evidence for other OECD countries include Andrews, Criscuolo and Gal (2016) and Berlingieri, Blanchenay and Criscuolo (2017).

[2] The market value of an innovation refers to the value of the corresponding patent estimated by Kogan et al. (2017), who measure such values by stock market's responses to the news of granting of patents.

[3] A recent work, Celik and Tian (2020), also construct the "tail innovation" index in similar way, but using the distribution of number of citations rather than patent values. The idea of "tail innovations" goes further back to Acemoglu, Akcigit and Celik (2014)

shock, and find that there is a persistent drop of around 50% in sales at both levels. Placebo tests indicate this plunge in sale is likely due to the shock, which is thus called the destructive innovation shock interchangeably in the rest of this paper. Applying the same methodology to the impulse response of productivity dispersion to the destructive innovation shock, I find hump-shaped response function of industry-level productivity dispersion under various definitions (the inter-quartile range, 90-10 percentile, and standard deviation). Again, placebo tests are in favor of the causal interpretation between the destructive innovation shock and higher productivity dispersion.

Having shown the rise in productivity dispersion as a response to the destructive innovation shock, I proceed to seek for a theoretical explanation of the mechanism. The class of models of technology ladders, or quality ladders, is by its nature suitable for this question. In this type of models, the positions of firms (usually two) along a "ladder" of technology is stochastically determined by their R&D efforts. From that, the technology gap between the two firms can be calculated, and serves as an analogy of the productivity dispersion observed from the data. To the best of my knowledge, I am (at least among) the first to study the response of technology gap to the destructive innovation shock using a technology-ladder model. I start by characterizing the equilibrium without shock in a baseline model, where the firms' R&D efforts and technology gap are continuous-time Markov processes, whose corresponding Markov chains are of finite states. I prove that when two firms are exactly identical, then the expected technology gap is zero under the limiting distribution. Intuitively, the two firms have equal chance to be the leader or laggard at any degree. Otherwise, if one firm's marginal R&D cost is strictly lower than the other's, then the expected technology gap of the low-cost firm is strictly positive under the limiting distribution. In other words, the low-cost firm is expected to be the leader in technology. For this claim I prove the special case where the maximum technology gap is one, due to the complexity of the dynamical system. However, numerical experiments support the extention of the proposition to any finite upper bound of gap. My work with the baseline model benefits from Aghion et al. (2001), Aghion et al. (2005) and Ludkovski and Sircar (2016), though the technology gap is the focus of none of them.

I extend the baseline model to study the impact of the destructive innovation shock. When such shock happens, the optimized profit earned by each firm for the same technology gap is reduced. This is to reflect the aforementioned fact that major innovations destroy rival firms' sales. I introduce the concept of distances to the technology frontier to formally model this idea. When there is no shocks, as in the baseline model, the leading firm (or both if they are neck-to-neck) is on the technology frontier, thus its distance to the frontier is zero. Upon the shock, both firms are pushed backwards from the technology frontier, with their relative position to each other – or technology gap – unchanged. The degree of destruction of the shock is increasing in how far the leading firm is pushed away from the technology frontier. Firms hit by the shock can recover the loss in profit by advancing towards the technology gap, and this is done through innovations whose arrival rates are functions of the R&D efforts. The concern is what are firms' instantaneous responses to the destructive innovation shock in their R&D decisions. I prove that when the degree of shock is sufficiently large, both firms will lower their R&D efforts; if the degree of shock is extremely small, however, the leading firm will increase its R&D effort. Again, due to the complexity of this stochastic dynamical system, to the question how firms' respond to shocks with intermediate degrees I have no analytical answer, and I shall attempt to provide a numerical one.

I pin down the values of model parameters by calibration and estimation, and solve the value and policy functions numerically using the value function iteration method. The policy function of each firm is a mapping from the two firms' distances to the technology frontier to the firm's R&D effort. With that I simulate the system where the arrival of innovation is a Poisson process, with state-dependent arrival rate given by the policy functions. I hit the system with a destructive innovation shock with a degree consistent with the empirical impulse response function, and observe the responses of R&D efforts taking average over all repetitions. It turns out that the low-cost firm responds by raising R&D effort, and the high-cost firm by lowering. This instantly increases the difference in the two firms' R&D efforts, which gradually converges to the pre-shock level. The heterogeneous responses in R&D results in a hump-shaped curve of the technology gap after the shock, resembling those obtained empirically using the local projection method.

To bring my proposed mechanism to data, I empirically test for heterogeneity of firms' R&D responses to the destructive innovation shock. The results suggest that firms with high R&D intensity (RDI) in the past incline to raise their RDI in the year after a destructive innovation shock, while firms with low past RDI are found to lower their RDI in response to the same shock. This evidence coincides with the patterns from the simulation, and thus verifies the channel through which the destructive innovations contributes to productivity dispersion.

Contribution to Literature

This paper contributes to the understanding of the source of productivity dispersion, and claims that the rapidly rising productivity dispersion observed in U.S. since the 90's can be explained by the emergence of destructive innovations, in response to which within-industry R&D efforts becomes more dispersed. This provides a new, data-backed perspective to the extant literature explaining the productivity dispersion by other factors, such as the surge of entry (Foster et al. (2019)), or the life-cycle of businesses (Haltiwanger, Jarmin and Miranda (2013); Decker et al. (2016)).

My paper is also related to the question that how we should see productivity dispersion in a normative way. The prominent work of Hsieh and Klenow (2009) has made productivity dispersion the synonym of resource misallocation. A recent work of Gouin-Bonenfant (2020) shows that productivity dispersion leads to lower labor share. Kehrig (2015) and Caballero and Hammour (1994) together imply the negative cyclical productivity dispersion may be associated with the cleansing effect during recessions. My theory holds that given the heterogeneous responses in R&D described above, productivity dispersion is inevitable by-product of creative destruction. And although I do not study firm entry or exit due to the nature of my data[4], it is easy to see that how these heterogeneous responses incentivize those firms with the lowest productivity to exit. Thus productivity dispersion, instead of a sign of misallocation, can be a process of reallocation, in the sense of Acemoglu et al. (2018).

My theory contributes to the literature of the technology ladder models tracing back to Segerstrom, Anant and Dinopoulos (1990), Grossman and Helpman (1991) and Aghion and Howitt (1992), by studying the behavior of the gap along the technology ladder between two firms under the limiting distribution. The contributions of my empirical work is to show the dynamic destruction of major innovations to rival firms' sales. The similar question is answered by Klette and Kortum (2004), Broda and Weinstein (2010) and Garcia-Macia, Hsieh and Klenow (2019), all in different lines[5].

Layout

The remainder of this paper is organized as follows. Section 2 establish an empirical strategy to identify major innovations, and employs the local projection method to obtain the impulse response functions of sale as well as productivity dispersion to the arrival of major innovation. Section 3 sets up the baseline model to study the behavior of R&D efforts and technology gap under the limiting distribution. Section 4 extends the baseline model to analyze the impact of the destructive innovation shock. Section 5 and 6 assign parameter values through calibration and estimation, solve the baseline and extended models numerically,

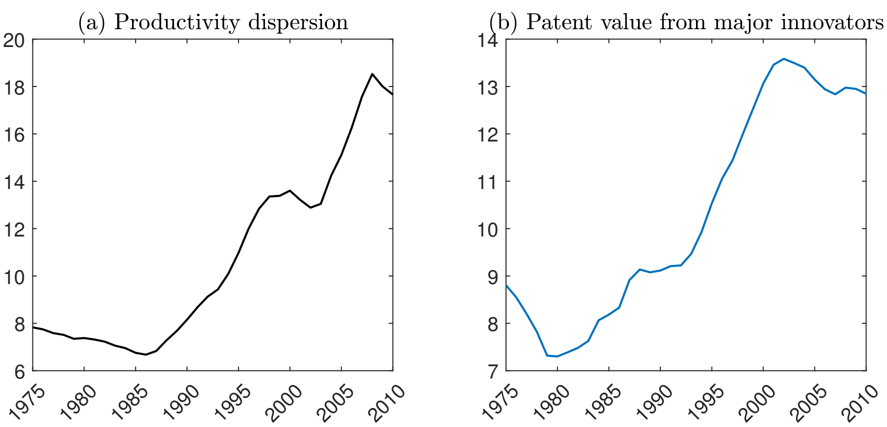
[4]Entering into or disappearing from the Compustat database is not good proxies for firms' entry or exit.

[5]Innovations are found to be destructive to firms (who exit) in Klette and Kortum (2004), to products in Broda and Weinstein (2010), and to jobs in Garcia-Macia, Hsieh and Klenow (2019).

and portray the responses of variables of interest to the destructive innovation in simulations. Section 7 provides empirical evidence in support of the proposed mechanism. Section 8 concludes.

2 Impulse Responses to Destructive Innovations

This section documents empirical findings on the destructiveness of high-valued innovations to rival firms, and the enlargement of within-industry productivity dispersion as a result. These facts shed light on the cause of the upward trend of productivity dispersion over U.S. manufacturing firms in recent decades (see Figure 1)[6], and motivate a theoretical treatment of the underlying mechanism to be presented in Sections 3 and 4.



Panel (a) shows the annual inter-quartile range of TFP over all sampled firms; panel (b) shows the natural log of all firm-year patent values from major innovators. Both time series are in 5-year moving average.

Figure 1: Productivity dispersion and patent value from major innovators

2.1 Identifying Destructive Innovators

Destructive innovators, interchangeably referred to as major innovators in this paper, are those firms who have made high-valued innovations in a year relative to other firms in the same industry. The rest of this subsection formalizes this conceptual definition, constructs the major innovation index, whose impact on rival firms' sales as well as on productivity dispersion is later studied.

[6]External evidence of the rising productividy dispersion in the U.S. is listed in Section 1; for that of the ascending innovation measures during the same period, see Figure IV in Kogan et al. (2017).

Data

Information on the firm-year patent values is from the database constructed by Kogan et al. (2017), where the economic value of each single patent from U.S. listed firms is measured by the response from the stock market in a short time window upon the news of patent granting, and the patent values are normalized to 1982 million dollars and aggregated to firm-year cells. This dataset is merged with CRSP/Compustat Merged Database to incorporate firm fundamentals (sales, R&D expenditures, etc.). The NBER-CES Manufacturing Industry Database is also used for its documentation of industry-year level labor share, which helps to obtain cost-share based revenue total factor productivity (TFPR, or TFP).

Merging the above three databases yields an unbalanced panel dataset ranging in years from 1970 to 2010, covering 4,074 U.S. manufacturing firms from 135 industries defined by 4-digit Standard Industrial Classification (SIC) codes. The details on data description and pre-processing are left to the appendix.

Major Innovators

Firstly, to make patent values from different years comparable, remove the year fixed effect by running the regression

$$Tsm_{i,j,t} = \alpha + \delta_t + u_{i,j,t}, \tag{1}$$

where Tsm is the firm-year patent value with the same notation as in Kogan et al. (2017). Subscripts i, j, t are the firm, industry and year indicators, respectively. δ_t is the year fixed effect and u the error term. The residual from the above regression, denoted by Tsm^{res}, is orthogonal to the year fixed effects and thus used as the detrended firm-year patent values.

Industries may have different standards in the degrees above which a firm-year patent value should be considered high. To account for this, for each industry j', I pool all the detrended patent values $Tsm^{\text{res}}_{i,j,t}$, and take the industry-specific 95th percentile over $\left\{Tsm^{\text{res}}_{i,j,t} | j = j'\right\}$, denote it as $c_{j'}$. A firm i from industry j is thus said to be a *major innovator* at year t if and only if $Tsm^{\text{res}}_{i,j,t} \geq c_j$. The indicator variable MI^{firm} identifies major innovators according to the above criterion:

$$MI^{\text{firm}}_{i,j,t} = \begin{cases} 1 & \text{if } Tsm^{\text{res}}_{i,j,t} \geq c_j; \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The industry level major innovation indicator $MI_{j,t}$ obtains value 1 if and only if there is at least one major innovator in industry j at year t:

$$MI_{j,t} = \begin{cases} 1 & \text{if } \max_i \left\{ MI^{\text{firm}}_{i,j,t} \right\} \geq 1; \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

2.2 Destructiveness of Major Innovations

I employ the *local projection method* (Jordà (2005)) to study the dynamic impact of major innovations on rival firms' sales[7], which is meausred at both firm and industry levels. To make sure the identification strategy captures the impact of major innovations on other firms, I exclude firms who have been major innovators in past ten years, with ten being the maximum period examined in the *impulse response function* (IRF). Formally, the firm- and industry-level sales are thus defined as follows:

$$S_{i,j,t} = \text{sales}_{i,j,t} \times \mathbb{1} \left\{ \max_{s=t-10,\cdots,t} MI^{\text{firm}}_{i,j,s} = 0 \right\}, \tag{4}$$

and

$$S_{j,t} = \sum_i S_{i,j,t}. \tag{5}$$

And their corresponding regression specifications are

$$\log\left(S_{i,j,t+h}\right) = \alpha_i^h + \delta_t^h + \beta^h MI_{j,t} + \boldsymbol{\gamma}^h \boldsymbol{X}_{i,j,t} + \boldsymbol{\lambda}^h \boldsymbol{X}_{j,t} + u^h_{i,j,t}, \tag{6}$$

and

$$\log\left(S_{j,t+h}\right) = \alpha_j^h + \delta_t^h + \beta^h MI_{j,t} + \boldsymbol{\lambda}^h \boldsymbol{X}_{j,t} + u^h_{j,t}. \tag{7}$$

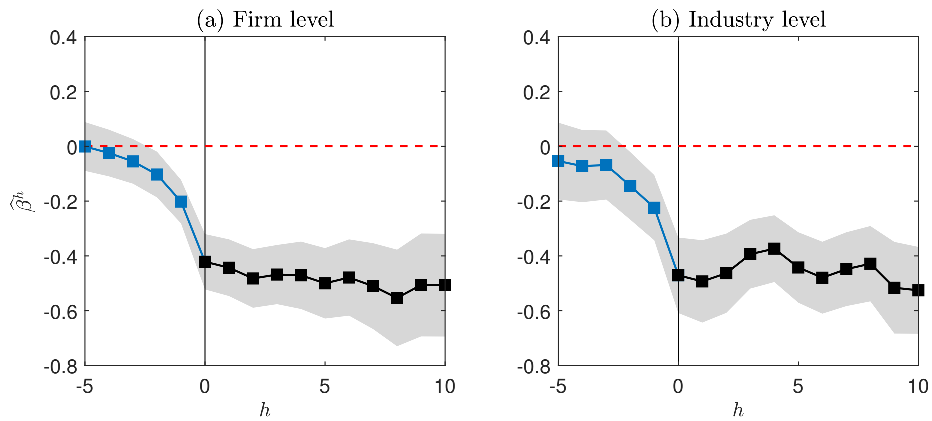
In the above expressions, α_i and α_j are firm and industry fixed effects; δ_t is the year fixed effect; $MI_{j,t}$ the major innovation indicator defined in (3); $\boldsymbol{X}_{i,j,t}$ and $\boldsymbol{X}_{j,t}$ the firm- and industry-level set of control variables. The industry-level controls are firm number in the industry-year cell, and the *Herfindahl-Hirschman Index* (HHI)[8], both at one-year lag. The firm-level controls

[7] Jordà (2005) identifies shocks first by regressing variable of interest y at time $t+h$ on lagged values of y. Since I identify the shocks directly, the specifications here are in line with those as in Ramey (2016).

[8] The HHI is a standard index to measure the degree of within-industry concentration, defined as $HHI_{j,t} = \sum_i \left(ms_{i,j,t}\right)^2$, where ms is the market share in percentage. For details, see Herfindahl (1950) and Hirschman (1945).

include TFP[9], market share, capital stock, employment – all at one-year lag, and firm age.

The coefficients $\{\beta^h\}_{h=0}^{10}$ captures the dynamic impact of major innovations on the market sizes of other competing firms from the current year to ten years later. The figure below plots the estimated $\widehat{\beta}_h$ with $h = -5, \cdots, 0, \cdots, 10$, where negative h's are for placebo test to see if the impulse responses are truly caused by the shock.



Panel (a) plots the IRF estimated using specification (6); panel (b) using specification (7). Shaded areas are 90% confidence bands, determined by robustness standard errors clustered at the industry level.

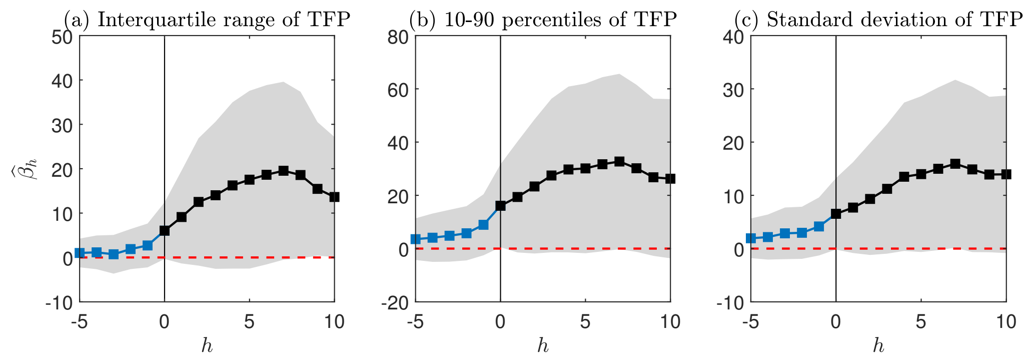
Figure 2: Impulse responses of firm- and industry-level market size to major innovations

From Figure 2, major innovations are persistently destructive to rival firms' sales. The downward pretrend implies that such impact is in place two years prior to the occurance of major innovation. To this, one explanation is that these major innovators don't wait until their filed patents to be granted to utilize their new technologies, in which case the treatment effect is before the documentation of the treatment (at $h = 0$).

2.3 Macro Impact of Destructive Innovations on Technological Changes

The destructive innovations may have implications on multiple aspects of the macro economy, among which that on technological changes is of concern. As Figure 3 shows, the destructive innovations broadens the within-industry productivity dispersion under three different meausres.

[9] Firm-year TFP is estimated using cost-share based method, to which a recent detailed treatment is Foster et al. (2017). Details on the methodology are left to the appendix.



Industry-year obeservations with firm number less than 2 are dropped from local projection regressions. Shaded areas are 90% confidence bands, determined by robustness standard errors clustered at the industry level.

Figure 3: Impulse responses of productivity dispersion to major innovations

The above impulse response functions are obatined through local projection regressions, whose specifications is the same as that of (7). The only difference is the explained variables, which are three canonical measures of productivity dispersion: interquartile range (IQR), the 10-90 percentile range, and the standard deviation of TFP. All of them are calculated at the industry-year level.

Figure 3 suggests the major innovations are responsible for a persistent expansion of productivity dispersion in the industry where they occur. This is consistent with the resemblance between the time paths of patent values from major innovators and the measure of productivity dispersion in Figure 1. To explore the mechanism by which destructive innovations contributes to productivity dispersion, I propose a theory in the following two sections where firms have heterogeneous responses in R&D to destructive innovations.

3 Technology Ladder and Heterogeneous Climbers

In this section I introduce a dynamic model of R&D racing under duopolistic market where firms' R&D strategies endogenously generate technological progress and technology gap in the stationary equilibrium. In the next section, I will extend the model to study firms' responses in R&D to destructive innovations, and their implications on the evolvement of productivity dispersion.

My model is closely related to Aghion et al. (2001) in the sense that firms compete by climbing the technology ladder[10]. The major contribution of my more generalized model is

[10]Also known as the quality ladder as in Grossman and Helpman (1991).

twofold: firstly, I assume heterogeneous R&D capacities among firms, and thus study the behavior of technology gap in the stationary equilibrium; secondly, I propose a way to incorporate destructive innovations into this class of models.

3.1 Model Setup

Technology and Profit

Two firms, indexed by $f \in \{A, B\}$, consist a duopolistic industry. Time is continuous: $t \in [0, \infty)$. At every instant of time t, firm f's profit flow $\pi_f(t)$ depends on the locations of the two firms on a *technology ladder* with countably infinite steps: $\boldsymbol{n}(t) = (n_A(t), n_B(t)) \in \mathbb{N}_+^2$. The higher $n_f(t)$, the more advanced in technology firm f is. The relationship between firms' profit and their positions on the ladder are assumed to be as follows:

Assumption 1 *The profit function $\pi_f(\boldsymbol{n}(t))$, $f \in \{A, B\}$ has the following properties:*
1. Time-invariancy: *for any $\widetilde{\boldsymbol{n}} \in \mathbb{N}_+^2$, and any $s, t \in [0, \infty)$, $\pi_f(\boldsymbol{n}(s) = \widetilde{\boldsymbol{n}}) = \pi_f(\boldsymbol{n}(t) = \widetilde{\boldsymbol{n}})$.*
2. Monotonicity: *$\pi_f(\boldsymbol{n})$ is strictly increasing in the* technology gap $\Delta n_f \coloneqq n_f - n_{-f}$.
3. Symmetry: for any $n_1, n_2 \in \mathbb{N}_+$, $\pi_A(n_1, n_2) = \pi_B(n_2, n_1)$.
4. Zero-sum: for any $\widetilde{\boldsymbol{n}} \in \mathbb{N}_+^2$, $\pi_A(\widetilde{\boldsymbol{n}}) + \pi_B(\widetilde{\boldsymbol{n}}) = \overline{\pi} > 0$.

Under these assumptions, the profit flow of either firm is uniquely determined by the current technology gap, regardless of time, the absolute number of steps, or firm label; whoever holding certain relative leading position on the technology ladder earns the same profit. This feature allows me to reduce the strategy space from uncountably infinite, as dependent on the continuous time, to finite as on the technology gap. How this is done will be shown in details in Lemma 1.

Among the literature regarding technology ladder, Aghion et al. (2001) assumes demands faced by the two firms are induced by constant elastisity of substitution (CES) production function; Ludkovski and Sircar (2016) studies a similar problem in Cournot duopolistic market; in Aghion et al. (2005) there is Betrand competition. The first three properties of Assumption 1 are satisfied in all these models, while Assumption 1.4 holds in the last reference.

Innovation

Firms exert R&D efforts to climb upwards along the technology ladder, to leave the *laggard* (or *follower*) farther behind or to catch up with the *leader*. Either way, an arrival of innovation increases the firm's technology gap relative to its rival, and grants it higher profit flows.

When an innovation arrives, the innovator moves one step ahead along the technology ladder. The occurance of innovations is stochastic: for each firm f, the location it is on the technology ladder is a random variable governed by a Poisson counting process $\{N_f(t); t > 0\}$. Steps $N_f(t)$ and $N_{-f}(t)$ are independent at any $t > 0$, unless in the special case where the absolute value of the technology gap, $|N_f(t) - N_{-f}(t)|$, equals its upper bound. What's different there will soon be elaborated.

At any time t, the Poisson hazard rate (or more intuitively, the arrival rate of innovation) for firm f is $\lambda_f(t) = \lambda a_f(t) + h \cdot \mathbb{1}\{\Delta n_f < 0\}$ where $a_f(t) \geq 0$ is the R&D effort chosen by firm f. The parameter h controls the rate of imitation[11] of the laggard: the laggard has a fixed positive rate to move one step ahead by imitating the leader, and that rate is independent of its own R&D effort. Therefore, at time t, the conditional probability mass function (PMF) of firm f's count of coming successful innovaton, or its advancement along the technology ladder by some future instant $s > t$, is[12]

$$\Pr\left\{N_f(s) - N_f(t) = k | n_f(t)\right\} = \frac{\left[\int_t^s \lambda_f(\tau) d\tau\right]^k \exp\left[-\int_t^s \lambda_f(\tau) d\tau\right]}{k!}, \quad k = 0, 1, 2, \cdots . \tag{8}$$

Can the technology gap between two competing firms be infinitely large? It would be unimaginable in reality. For example, the horse-drawn vehicles and automobiles were once close competitors as means of transportation, today they are no longer deemed to be on the same market. In this spirit I assume that there is an upper bound $\overline{m} \in \mathbb{N}_+$ of technology gap Δn_f for $f \in \{A, B\}$. If the leader is at this maximum gap when an innovation arrives, the follower will also jumps a step upwards automatically and simultaneously, so that the gap remains to be $\overline{m}$, and shall never exceed it[13]. Therefore, the leader has no incentive to do any R&D when the maximum gap is reached. The automatic catch-up at the maximum gap can be understood with two scenarios. First, there is a spillover from expired patents where the learning of old-fashioned technologies by the laggard imposes no concerns to the far-reaching leader. In the second instance, the laggard firm is so dropped behind that it is replaced by an outsider who is just one-step ahead of it when the leader succeeds in innovating. As

[11]Some authors use the name "spillover" instead of "imitation" to refer the same phenomenon in which the laggard automatically advances in technology.

[12]See Gallager (1995), Chapter 2, page 43, Theorem 2.

[13]In Aghion et al. (2001), the maximum technology gap (called the lead size in their paper) is implicitly determined by restriction on stationary distribution. In Aghion et al. (2005), it is set to be 1. In Ludkovski and Sircar (2016), the largest gap possible is pinned down by the fact firms stop doing R&D at some point due to marginal cost bounded away from zero.

a result, the space for technology gap $\Delta n_A(t) := n_A(t) - n_B(t)$[14] is finite, and denoted by $\mathcal{M} := [-\overline{m}, -\overline{m}+1, \cdots, -1, 0, 1, \cdots, \overline{m}-1, \overline{m}]$.

Doing R&D incurs cost, the cost function $\psi_f(t) = \psi_f\left(a_f(t)\right)$ is firm-specific and satisfies the following properties:

Assumption 2 *For any $f \in \{A, B\}$, and $a_f \geq 0$, the R&D cost function $\psi_f(a_f)$ satisfies:*
1. $\psi_f(0) = 0$*;*
2. ψ_f *is second-order differentiable;*
3. ψ_f *is strictly increasing;*
4. ψ_f *is strictly convex.*

These assumptions are standard as regards cost functions. Assumption 2.1 means there is no fixed cost to do R&D; Assumption 2.4 ensures that the marginal cost of R&D is strictly increasing.

Strategy and Payoff

A strategy profile in this R&D racing game is $(\boldsymbol{a}(t))_{t\geq 0} = (a_A(t), a_B(t))_{t\geq 0} \in \mathbb{R}^2_{++} \times [0, \infty)$ that fully specifies each firm's R&D effort at any instant. For an arbitrary strategy profile, firm f's expected discounted sum of net profit flow, also called *performance measure* for ease, is

$$\mathcal{J}_f = \mathbb{E}_{\boldsymbol{N}}\left\{\int_0^\infty e^{-\rho t}\left[\pi_f\left(\boldsymbol{N}(t)\right) - \psi_f\left(a_f(t)\right)\right] dt \middle| \boldsymbol{n}(0), (\boldsymbol{a}(t))_{t\geq 0}\right\}, \tag{9}$$

where ρ is the discount factor. The expectation is over the stochastic process $(\boldsymbol{N}(t))_{t\geq 0} = (N_A(t), N_B(t))_{t\geq 0}$, whose realization $(\boldsymbol{n}(t))_{t\geq 0} = (n_A(t), n_B(t))_{t\geq 0}$ pins down the path of the steps of the two firms on the technology ladder. The evolvement of $\boldsymbol{N}(t)$ is governed by PMF (8) and the R&D strategy profile $(\boldsymbol{a}(t))_{t\geq 0}$ that determines the arrival rates of innovations.

Given the strategy played by its competitor, firm f chooses its own strategy to maximize its performance measure in (9), whose optimized value defines the value function:

$$v_f\left(\boldsymbol{n}(0); \left(a_{-f}(t)\right)_{t\geq 0}\right) = \sup_{\left(a_f(t)\right)_{t\geq 0}} \mathbb{E}_{\boldsymbol{N}}\left\{\int_0^\infty e^{-\rho t}\left[\pi_f\left(\boldsymbol{N}(t)\right) - \psi_f\left(a_f(t)\right)\right] dt \middle| \boldsymbol{n}(0), (\boldsymbol{a}(t))_{t\geq 0}\right\}. \tag{10}$$

The performance measure and the value function, expressed as the discounted sum of net profit flows, have clear economic intuition. However, the game in this form is highly

[14]Similarly, the technology gap from the standpoint of firm B is $\Delta n_B(t) := n_B(t) - n_A(t) = -\Delta n_A(t)$. For the ease and clarity of discussion, in the rest of this paper, I refer to "technology gap" as $\Delta n_A(t)$ from firm A's view, unless otherwise specified.

untractable. In the following equilibrium analysis, I will simplify this stochastic game through a transformation into a finite-state equivalent.

Equilibrium Concepts

In this dynamic stochastic game, the *Nash equilibrium* (NE) is a strategy profile from which no player has the incentive to deviate:

Definition 1 *A Nash equilibrium is a pair of strategies $\left(a_A^*(t), a_B^*(t)\right)_{t\geq 0}$, where given $\left(a_{-f}^*(t)\right)_{t\geq 0}$, the strategy $\left(a_f^*(t)\right)_{t\geq 0}$ is the solution to firm f's optimization problem (10).*

Specifically, if $\left(a_{-f}^*(t)\right)_{t\geq 0}$ is a *Markov strategy*, in which action $a_{-f}^{(}t)$ depends solely on the state of technology gap $\Delta n_{-f}(t) = n_{-f}(t) - n_f(t)$, then firm f's strategy in the NE must also be Markovian:

Lemma 1 *In any Nash equilibrium, for any $f \in \{A, B\}$, suppose $s \geq 0$, and $t \geq 0$ satisfy $s \neq t$ and $\Delta n_{-f}(s) = \Delta n_{-f}(t)$. If $a_{-f}^*(s) = a_{-f}^*(t)$ and $a_f^*(t)$ is right-continuous, then $a_f^*(s) = a_f^*(t)$.*

Proof. Fix an arbitrary time $s \geq 0$, and let t be such that $\Delta n_f(s) = \Delta n_f(t)$ and $s \neq t$. Since firm $-f$ plays a Δn_{-f}-dependent strategy, by Assumption 1 and performance measure 9,then firm f faces the same optimization problem. Thus by the strict convexity of R&D cost function, it implies that the set $\left\{t | s \neq t, \Delta n_f(s) = \Delta n_f(t), a_f^*(s) \geq a_f^*(t)\right\}$ has zero measure, and it must be empty set if $a_f^*(t)$ is right-continuous.

□

The restriction on right-continuous strategies[15] is to exclude the case where there are discontinuity points of the first kind in the time paths of R&D efforts, which is mathematically possible but of no economic interest.

Lemma 1 says if in the NE, one firm plays strategy contingent on its technology gap, the other must do the same[16]. Using this property, I reduce the strategy space from an uncountably infinite set (of continous time) to a finite set (of technology gap). In the rest of this paper, most discussion will be based on the latter approach, except for a few contexts in which the time-path interpretation is more convenient for analytical purposes.

The above reduction in dimension of the strategy space leads to the second equilibrium concept, the *Markov sub-game perfect equilibrium* (MPE). The following definition is in line with the equilibrium concept developed by Tirole (1988) and Maskin and Tirole (2001)[17].

[15] In Lemma 1, the condition $a_{-f}^*(t) = a_{-f}^*\left(\Delta n_{-f}(t)\right)$ implies that $\left(a_{-f}^*(t)\right)_{t\geq 0}$ is right-continuous.

[16] Since $\Delta n_f = -\Delta n_{-f}$, to say firm f's strategy is Δn_{-f}-contingent is equivalent to saying that it is Δn_f-contingent.

[17] For a textbook treatment and more examples, see Miao (2014).

Definition 2 *A Markov sub-game perfect equilibrium is a pair of state-contingent strategies* $\left\{a_A^*(\Delta n_A), a_B^*(\Delta n_B)\right\}_{\Delta n_A=-\Delta n_B\in\mathcal{M}}$*, where given* a_{-f}^**, the strategy* a_f^* *is the solution to firm* f*'s optimization problem:*

$$\sup_{\left\{a_f(\Delta n_f)\right\}_{\Delta n_f\in\mathcal{M}}} \mathbb{E}_{\Delta N_f}\left\{\int_0^\infty e^{-\rho t}\left[\pi_f\left(\Delta N_f(t)\right)-\psi_f\left(a_f\left(\Delta N_f(t)\right)\right)\right]dt\Big|\Delta n_f(0),\left\{\boldsymbol{a}(\Delta n_f)\right\}\right\}. \tag{11}$$

In the rest of this paper, for analytical tractability as well as numerical computability, I shall restrict my discussion to the class of MPE.

3.2 Existence of the Markov Perfect Equilibrium

I have shown how the infinite-horizon game can be transformed to a finite-state equivalent, in this subsection I shall discuss the existence and characterization of the MPE. Denote by random variable Z the time interval between the current instance and the next arrival of innovation from either firm A or B; and by Z_f the firm-specific arriving time. Due to the memoryless property based on Poisson PMF (8), the value function in MPE can be re-written as

$$\begin{aligned} v_f\left(\Delta n_f\right) =& \mathbb{E}_Z\Bigg\{\int_0^Z e^{-\rho t}\left[\pi_f(\Delta n_f)-\psi_f\left(a_f^*(\Delta n_f)\right)\right]dt \\ &+ e^{-\rho Z}\mathbb{1}\{Z=Z_f\}v_f\left(\Delta n_f+1\right)+e^{-\rho Z}\mathbb{1}\{Z=Z_{-f}\}v_f\left(\Delta n_f-1\right)\Bigg\}, \end{aligned} \tag{12}$$

where event $\{Z=Z_f\}$ means the next innovation is achieved by firm f. This stochastic Bellman equation can be further simplified through removing the expectation operator, which is then easy to compute for a numerical solution of the equilibrium. The existence of equilibrium is guaranteed by the application of Kakutani's Fixed-Point Theorem. The technical detals are left to the appendix.

Proposition 1 *A Markov sub-game perfect equilibrium* $a_f^*(\Delta n_f)$ *for* $f\in\{A,B\}$ *and* $\Delta n_f=-\Delta n_{-f}\in\mathcal{M}$ *exists, and is the solution to the following optimization problem:*

$$\begin{aligned} v_f(\Delta n_f) = \sup_{a_f\geq 0}& \frac{1}{\lambda(a_f+a_{-f}^*)+h\cdot\mathbb{1}\{\Delta n_f\neq 0\}+\rho}\Big\{\pi_f(\Delta n_f)-\psi(a_f) \\ &+(\lambda a_f+h\cdot\mathbb{1}\{\Delta n_f<0\})v_f(\Delta n_f+1)+(\lambda a_{-f}^*+h\cdot\mathbb{1}\{\Delta n_f>0\})v_f(\Delta n_f-1)\Big\} \end{aligned} \tag{13}$$

for all $\Delta n_f < \overline{m}$, with boundary conditions

$$a^*_f(\overline{m}) = 0; \tag{14}$$

$$v_f(\overline{m}) = \frac{1}{\lambda a^*_{-f} + h + \rho}\Big\{\pi_f(\overline{m}) + (\lambda a^*_{-f} + h) v_f(\overline{m} - 1)\Big\}. \tag{15}$$

Proof. See Appendix.

In the above, a^*_f and v_f are the policy function and value function in the MPE. Equation (13) shows the state-contingent value function transformed from (11); and the boundary conditions are from the fact that the leading firm, at the maximal technology gap, has no incentive to do R&D. This is because it gains nothing from an innovation due to the automatic advancing of its competitor.

The first order condition of (13) with respect to R&D effort implicitly determines the policy function a^*_f. Substitute it back to the value function, and the game is characterized by the following system of non-linear equations. This is later used to find the numerical solution of the model.

Corollary 1 *The value funtions $v_f(\Delta n_f)$ and policy functions $a_f * (\Delta n_f)$ for $f \in \{A, B\}$ are solution to the system of equations:*

$$\begin{aligned}
v_f(\Delta n_f) =& \frac{1}{\lambda\left[a^*_f(\Delta n_f) + a^*_{-f}(-\Delta n_f)\right] + h \cdot \mathbb{1}\{\Delta n_f \neq 0\} + \rho}\Big\{\pi_f(\Delta n_f) - \psi_f\left(a^*_f(\Delta n_f)\right) \\
&+ \left[\lambda a^*_f(\Delta n_f) + h \cdot \mathbb{1}\{\Delta n_f < 0\}\right] v_f(\Delta n_f + 1) \\
&+ \left[\lambda a^*_{-f}(-\Delta n_f) + h \cdot \mathbb{1}\{\Delta n_f > 0\}\right] v_f(\Delta n_f - 1)\Big\}, \quad \textit{if } \Delta n_f < \overline{m};
\end{aligned} \tag{16}$$

$$\frac{d\psi_f\left(a^*_f(\Delta n_f)\right)}{da_f} = \lambda\left[v_f(\Delta n_f + 1) - v_f(\Delta n_f)\right], \qquad \textit{if } \Delta n_f < \overline{m}; \tag{17}$$

$$a^*_f(\overline{m}) = 0; \tag{18}$$

$$v_f(\overline{m}) = \frac{1}{\lambda a^*_{-f} + h + \rho}\Big\{\pi_f(\overline{m}) + (\lambda a^*_{-f} + h) v_f(\overline{m} - 1)\Big\}. \tag{19}$$

Proof. This is directly from Proposition 1, with first order conditions with respect to R&D efforts.

□

Equation (17) conveys clearly the principle which pins down the equilibrium R&D efforts:

the marginal cost of R&D should be equal to the increment of firm value from innovation, scaled by the marginal arrival rate of innovation λ. Since the R&D cost function is strictly convex, the intuition is that if the firm value increases more from innovation, or if the next innovation is expected to arrive sooner, then the firm exerts higher R&D efforts in the equilibrium.

In the system in Corollary 1, there are $(8\overline{m} - 4)$ unknowns[18], and the same number of equations. By Proposition 1, the MPE exists, thus so does the solution to this system.

3.3 Stationary Distribution of the Technology Gap

This subsection intensively uses concepts and resultsregarding Markov processes to study the limiting behavior (when $t \to \infty$) of the technology gap in the MPE. Notice that the state space of gap Δn_A, $\mathcal{M} = \{-\overline{m}, -\overline{m}+1, \cdots, -1, 0, 1, \cdots, \overline{m}-1, \overline{m}\}$, has $2\overline{m}+1$ components. Let $\mathcal{M}_i$ be the ith component in $\mathcal{M}$, $i = 1, \cdots, 2\overline{m}+1$. The *transition rate* of the process $(\Delta N_A(t))_{t\geq 0}$ from state $\mathcal{M}_i$ to an adjacent state $\mathcal{M}_j$, where $j = i-1$ or $j = i+1$, is denoted by $q_{i,j}$, whose value is determined according to the following rule:

$$q_{i,j} = \begin{cases} \lambda a_A^* \left(\mathcal{M}_i\right) + h \cdot \mathbb{1}\{\mathcal{M}_i < 0\}, & \text{if} \quad j = i+1; \\ \lambda a_B^* \left(-\mathcal{M}_i\right) + h \cdot \mathbb{1}\{\mathcal{M}_i > 0\}, & \text{if} \quad j = i-1. \end{cases} \tag{20}$$

When $|i-j| = 1$, the arrving time Z_f of innovation by firm f, as mentioned in value function (12), follows an exponential distribution with rate $q_{i,j}$. It means that between two successive jumps of process $(\Delta N_A(t))_{t\geq 0}$, to $\mathcal{M}_i$ and $\mathcal{M}_j$ respectively, the expected time it takes in the MPE is $1/q_{i,j}$. Similarly, the rate with which $(\Delta N_A(t))_{t\geq 0}$ leaves its current state $\mathcal{M}_i$ is $q_i \coloneqq q_{i,i+1} + q_{i,i-1}$, a property of the sum of Poisson processes.

Now define the transition rate matrix Q to be a square matrix of dimension $(2\overline{m}+1) \times (2\overline{m}+1)$, whose components $q_{i,j}$ satisfy (20) if $|i-j| = 1$; $q_{i,j} = -q_i$ if $|i-j| = 0$; and $q_{i,j} = 0$ if $|i-j| > 1$. The last case is because it is impossible for the technology gap to jumps with size greater than one. The figure below is a graphical representation in accordance with the Q-matrix.

[18] There are $(2\overline{m} - 1)$ states, under each four unknowns: a_A^*, a_B^*, v_A and v_B.

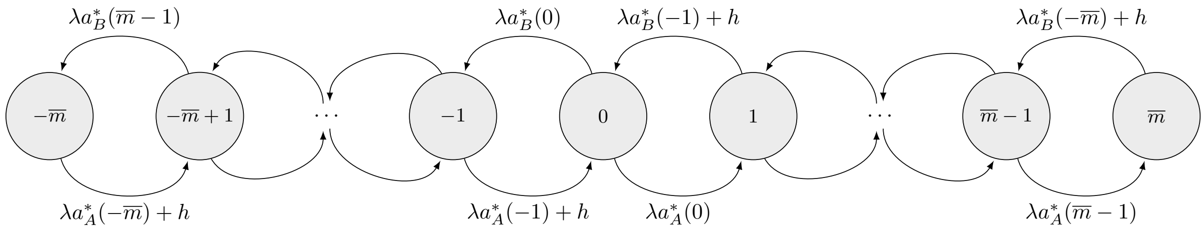


Figure 4: Transition rates across states of technology gap

In this way, the Q-matrix governs the *jump process* of the technology gap, from which distribution of the time between two jumps onto certain states can be inferred. However, the Q-matrix is unclear on the probability distribution over the new states when there is a jump. To study the distribution of the technology gap when t is sufficiently large, another matrix describing such probability is in need, and it can be derived from the Q-matrix.

When the technology gap jumps, the probability over where it heads is given by the *jump matrix P*. One way to construct it is through the entries of the Q-matrix:

Definition 3 *The jump matrix P is defined as follows:*

$$P = \left(p_{i,j} : i, j \in \{1, \cdots, 2\overline{m}+1\}\right), \tag{21}$$

where

$$p_{i,j} = \begin{cases} -q_{i,j}/q_{i,i}, & \text{if} \quad i \neq j, q_{ii} \neq 0; \\ 0, & \text{if} \quad i \neq j, q_{ii} = 0. \end{cases} \tag{22}$$

$$p_{i,i} = \begin{cases} 0, & \text{if} \quad q_{ii} \neq 0; \\ 1, & \text{otherwise.} \end{cases} \tag{23}$$

Just like with discrete-time Markov chains, the *stationary distribution* $\mu(m), m \in \mathcal{M}$ is defined by the invariance under multiplication by the P-matrix:

Definition 4 *A $1\times(2\overline{m}+1)$ vector μ is a stationary distribution over state space $\mathcal{M}$ if and only if:*
(1) $\mu P = \mu$;
(2) $\sum_{m\in\mathcal{M}} \mu(m) = 1$.

The stationary distribution μ has its name from the invariance property. It is also called the *limiting distribution* out of the fact[19]:

$$\mu(m) = \lim_{t \to \infty} \Pr(\Delta N_A(t) = m), \quad \forall m \in \mathcal{M}. \tag{24}$$

Thus when time t is sufficiently large, the probability that the technology gap $\Delta N_A(t)$ assumes certain value m doesn't depend on the initial gap $\Delta n_A(0)$. And so the limiting distribution μ provides the probablistic behavior of the technology gap in the long-term. In the rest of the paper, when calculating moments of variables that are state-dependent on the technology gap, it is this limiting distribution I shall use. And I will use the terms "stationary" and "limiting" interchangeably according to which fits the context better.

For a jump matrix P, if the associated Markov chain is irreducible and recurrent, then the stationary distribution μ in Definition 4 exists and is unique. From Figure 4 it is obvious that these conditions are met. Moreover, the stationary distribution can be explicitly expressed using the transition rate matrix Q:

Proposition 2 *In the MPE, the stationary distribution μ of technology gap $(\Delta N_A(t))_{t \geq 0}$ exists and is unique. Specifically, $\mu_(m) = -\xi_i q_{i,i}$, where $m = \mathcal{M}_\rangle$ and ξ_i is the i-th component of the solution to $\xi Q = 0$.*

Proof. See Theorems 3.5.1 and 3.5.2 in Norris (1998)[20].

□

Therefore, when there is a solution to equilibrium R&D strategies $\left\{a_A^*(\Delta n_A), a_B^*(\Delta n_B)\right\}_{\Delta n_f \in \mathcal{M}}$, the stationary distribution can be calculated through either Definition 4 or Proposition 2.

3.4 Expected Technology Gap under Limiting Distribution

This subsection explores the expected technology gap under the limiting distribution. I contrast two cases: in one the R&D costs of the two firms are identical; in the other, firm A's marginal R&D cost is pointwise lower than that of firm B. I start with the first case, the following result it obvious, but nonetheless necessary as a baseline.

Proposition 3 *If $\psi_A(a) = \psi_B(a)$ for all $a \geq 0$, then $\lim_{t \to \infty} \mathbb{E}\left[\Delta N_A(t)\right] = 0$.*

Proof. This is from the symmetry of the game, details are left to the Appendix.

[19]See Theorem 5.11 in Cinlar (1975).

[20]These theorems are about the existence and uniqueness of the *invariant measure*, between whicn and the limiting distribution there is a one-to-one mapping.

If two firms face the same cost on any level of R&D, the expected technology gap under the limiting distribution is zero. This is not surprising since the firms are identical, so are their strategies, and hence there is no reason for asymmetry in the long term.

The more interesting case is where firms have heterogeneous R&D costs. Without loss of generality, let firm A be more efficient in R&D than firm B. Particularly, assume the marginal R&D cost of firm A is strictly lower than that of firm B at all R&D effort levels. With its supremacy in R&D capacity, firm A's value function exceeds its rival's over all possible states:

Lemma 2 *If* $\dfrac{d\psi_A(a)}{da_A} < \dfrac{d\psi_B(a)}{da_B}$ *for all* $a > 0$*, then for any* $m \in \mathcal{M}$*,* $v_A(m) > v_B(m)$*.*

Proof. See Appendix.

The low-cost firm is not necessarily more R&D intensive at every state. However, in the simple case where the maximum technology gap is one, it is true that firm A's R&D effort dominates firm B's at all states.

Lemma 3 *In the case* $\overline{m} = 1$*, if the following two conditions hold, then* $a_A^*(m) > a_B^*(m)$ *for* $m = -1, 0$*.*

(1) $\dfrac{d\psi_A(a)}{da_A} < \dfrac{d\psi_B(a)}{da_B}$ *for all* $a > 0$*; and*

(2) *The discount factor* ρ *and the arrival rate multiplier* λ *are sufficiently small.*

Proof. See Appendix.

Naturally, the difference in R&D efforts leads to asymmetric probability masses on state space $\mathcal{M}$ under the limiting distribution.

Proposition 4 *In the case* $\overline{m} = 1$*, under the conditions in Lemma 3, in the limiting distribution, the probability mass function of technology gap satisfies* $\mu(1) > \mu(-1)$*.*

Proof. By the definition of Q-matrix and Lemma 3, the transition rate matrix Q is

$$Q = \begin{bmatrix} -\left(\lambda a_A^*(-1) + h\right) & \lambda a_A^*(-1) + h & 0 \\ \lambda a_B^*(0) & -\left(a_A^*(0) + a_B^*(0)\right) & \lambda a_A^*(0) \\ 0 & \lambda a_B^*(-1) + h & -\left(\lambda a_B^*(-1) + h\right). \end{bmatrix} \tag{25}$$

The *invariant measure* ξ is the solution to $\xi Q = 0$, whose components satisfy

$$\frac{\xi(1)}{\xi(-1)} = \frac{\lambda a_A^*(0)}{\lambda a_B^*(0)} \cdot \frac{\lambda a_A^*(-1) + h}{\lambda a_B^*(-1) + h} > 1. \tag{26}$$

By Definition 3, 4 and Proposition 2, $\mu(1) = \xi(1)\left(\lambda a_A^*(-1) + h\right) > \xi(-1)\left(\lambda a_B^*(-1) + h\right) = \mu(-1)$.

$\square$

Intuitively, if the marginal R&D cost of one firm is pointwisely lower than the other's, then under the limiting distribution, the low-cost firm's has a larger chance of being the leader ($\mu(1)$) than does its rival ($\mu(-1)$). It's implication on the technology gap is direct.

Corollary 2 *In the case* $\overline{m} = 1$*, under the conditions in Lemma 3,* $\lim_{t\to\infty} \mathbb{E}\left[\Delta n_A(t)\right] > 0$*.*

Proof. This follows from Proposition 4.

With heterogeneous R&D costs, the expected technology gap is no longer zero, but strictly positive. Due to the complexity of the non-linear system, this statement is proved under the simple case where the maximum gap is one. Meanwhile, parameter restriction in Lemma 3 results from the lack of analytical expressions of equilibrium R&D efforts and firm values. When it comes to numerical experiments in later sections, Proposition 4 and Corollary 2 hold with arbitrary finite $\overline{m} > 1$ and all randomized sets of parameters used for trial.

4 Extended Model with Destructive Innovation Shock

In the baseline model, the only state variable is the technology gap, whose expected value under the limiting distribution depends on the R&D cost functions of the two firms. In this section, I extend the baseline model to incorporate the destructive innovation shock, and examine its impact on R&D efforts and the technology gap.

4.1 Modelling Destructive Innovation

The destructive innovation shock is exogenous and unanticipated. When hit by it, both firms lag behind the technology frontier by a certain degree reflecting the magnitude of the shock, and are left with a smaller aggregate profit. Such conceptualization is justified by the impulse response functions of total sales of firms hurt by the major innovation shock (Figure 2).

Denote firm A and B's distances to the technology frontier on the technology ladder by d_A and d_B. For either firm $f \in \{A, B\}$, $d_f \in \{0, 1, 2, \cdots\}$, where $d_f = 0$ means that firm f is on the technology frontier. Therefore, $\min\{d_A, d_B\} = 0$ if and only if there is currently no destructive innovation shock. It is equivalent to saying that $\min\{d_A, d_B\} > 0$ if and only if there has been a destructive innovation shock from which the two firms have yet to recover, as the most advanced between them is still behind the frontier.

Profit Function

To capture the idea that the destructive innovation shock destroys the two firms' profits simultaneously, assume when such shock occurs, the distances of the two firms to the technology frontier increase by the same degree. The larger this degree, the smaller total profit is for the two firms to share. This is summarized in the following assumption:

Assumption 3 *Given distances to technology frontier d_f and d_{-f}, the profit of firm f is*

$$\pi_f(d_f, d_{-f}) = (1 - \delta \min\{d_f, d_{-f}\})\widetilde{\pi}(d_{-f} - d_f), \tag{27}$$

where $\delta \in (0, 1)$, function $\widetilde{\pi}$ satisfies Assumption 1.

When there is no shock, i.e. $\min\{d_A, d_B\} = 0$, the profit function and the whole game are reduced to those analyzed in the previous section. To see that the baseline model is a special case, note that the technology gap Δn_f is the difference between the distances, $d_{-f} - d_f$. If there is shock, the distance of the leading firm to the technology frontier is the degree of destruction to the total profit. Since $\widetilde{\pi}$ in profit function (27) satisfies Assumption 1, after the shock, the reduced total profit is divided between the two firms according to the technology gap between them, just like in the baseline model.

The upper bound of the degree of shock is the largest integer such that the profit remains non-negative: $\overline{D} = \max\{n \in \mathbb{N} : \delta n \leq 1\}$. The distance of one firm from the technology frontier can thus never exceeds $\overline{D} + \overline{m}$, the sum of the upper bounds of shock and of the technology gap.

Destructive Innovation Shock

The arrival of the destructive innovation shock is unanticipated by the firms, neither is it taken into consideration when firms form their strategies. Formally, by occurance of the shock, I mean the following:

Definition 5 *A destructive innovation shock with magnitude $D \in \{1, \cdots, \overline{D}\}$ occurs at time t if and only if:*

(1) There exists some $\tau > 0$ such that for any $s \in (t - \tau, t)$, $\min\{d_A(s), d_B(s)\} = 0$; and

(2) For $f = A, B$, $d_f(t) - \lim_{s \to t^-} d_f(s) = D$.

Transition of States

Since the state space is now two-dimensional, the evolvement of state (d_A, d_B) is more complex than in the baseline model. The exhaustive discussion of all possible transisions after an arrival of innovation is as follows:

Case 1: $d_f(t) = d_{-f}(t) = 0$. There is no destructive innovation shock and the two firms are neck-to-neck on the technology frontier. If the next innovation is made by firm f at time s, then $d_f(s) = 0$ and $d_{-f}(s) = 1$, meaning that it becomes the leader and still on the frontier.

Case 2: $d_f(t) = 0$ *and* $0 < d_{-f}(t) < \overline{m}$. There is no destructive innovation shock and the firm f is the leader. If the next innovation is done by the leader at time s, then $d_f(s) = 0$ and $d_{-f}(s) = d_{-f}(t) + 1$; if it is the laggard who innovates, $d_f(s) = 0$ and $d_{-f}(s) = d_{-f}(t) - 1$.

Case 3: $d_f(t) = 0$ *and* $d_{-f}(t) = \overline{m} > 0$. The only difference from case 2 is that now the leader f has no incentive to do R&D, just like in the baseline model. Therefore, the next innovation must be from the laggard $-f$.

Case 4: $d_f(t) > 0$ *and* $d_{-f}(t) > 0$. There has been a destructive innovation shock at least as late as time t. If the next innovation is from firm f at time s, then $d_f(s) = d_f(t) - 1$. For the other firm $-f$, if $d_{-f}(t) = d_f(t) + \overline{m}$, then $d_{-f}(s) = d_{-f}(t) - 1$ out of the automatic catching up at the maximum technology gap; otherwise it remains where it was: $d_{-f}(s) = d_{-f}(t)$.

To formalize the above rules of transition, for fixed $f \in \{A, B\}$, let $\left(d_f(t), d_{-f}(t)\right)$ be the state at time t. Suppose the next innovation arrives at time $s > t$. If the next innovation is done by firm f, the new state $\left(d_f(s), d_{-f}(s)\right)$ is given by the transition function T_f

$$T_f\left(d_f(t), d_{-f}(t)\right) = \begin{cases} \left(d_f(t), d_{-f}(t) + 1\right), & \text{if } d_f(t) = 0 \wedge 0 \leq d_{-f}(t) < \overline{m}; \\ \left(d_f(t) - 1, d_{-f}(t)\right), & \text{if } d_f(t) > 0 \wedge d_f(t) - d_{-f}(t) > -\overline{m}; \\ \left(d_f(t) - 1, d_{-f}(t) - 1\right), & \text{if } d_f(t) > 0 \wedge d_f(t) - d_{-f}(t) = -\overline{m}. \end{cases} \tag{28}$$

Similarly, if the next innovation is from firm $-f$[21], $\left(d_f(s), d_{-f}(s)\right)$ will be

$$T_{-f}\left(d_f(t), d_{-f}(t)\right) = \begin{cases} \left(d_f(t) + 1, d_{-f}(t)\right), & \text{if } d_{-f}(t) = 0 \wedge 0 \leq d_f(t) < \overline{m}; \\ \left(d_f(t), d_{-f}(t) - 1\right), & \text{if } d_{-f}(t) > 0 \wedge d_{-f}(t) - d_f(t) > -\overline{m}; \\ \left(d_f(t) - 1, d_{-f}(t) - 1\right), & \text{if } d_{-f}(t) > 0 \wedge d_{-f}(t) - d_f(t) = -\overline{m}. \end{cases} \tag{29}$$

In this way, for either firm f, function T_f returns the updated state upon the arrival of innovation depending on who the innovator is.

[21] I shall not discuss the case where both firms innovate simultaneously, as such an event has zero probability.

As in the baseline model, I allow for the imitation effect, with which the gap between the two firms shrinks by one step at a fixed rate $h \geq 0$ regardless of firms' R&D efforts.

4.2 Analysis of Equilibrium

In the extended model, given the initial distances to the technology frontier $\left(d_f(0), d_{-f}(0)\right)$ and any arbitrary strategy profile $\left(a_f(t), a_{-f}(t)\right)_{t\geq 0}$ specifying the time paths of R&D efforts of both firms from the beginning to infinity, the performance measure of firm f is, similar to its counterpart in the baseline model (9), the expected discounted sum of its net profit flow:

$$\mathcal{J}_f\left(d_f(0), d_{-f}(0) \middle| \left(a_f(t), a_{-f}(t)\right)_{t\geq 0}\right) = \mathbb{E}\left\{\int_0^{\infty} e^{-\rho t}\left[\pi_f\left(d_f(t), d_{-f}(t)\right) - \psi_f\left(a_f(t)\right)\right] dt\right\}. \tag{30}$$

Again, an equilibrium is a state-dependent strategy profile $\left(a_A^*(t), a_B^*(t)|d_A(0), d_B(0)\right)_{t\geq 0}$, such that given its rival's strategy $\left(a_{-f}^*(t)\right)_{t\geq 0}$, firm f's strategy $\left(a_f^*(t)|d_A(0), d_B(0)\right)_{t\geq 0}$ maximizes its performance measure:

$$\left(a_f^*(t)|d_A(0), d_B(0)\right)_{t\geq 0} \in \underset{\left(a_f(t)\right)_{t\geq 0}}{\operatorname{argmax}}\ \mathcal{J}_f\left(d_f(0), d_{-f}(0) \middle| \left(a_f(t), a_{-f}^*(t)\right)_{t\geq 0}\right). \tag{31}$$

As in the baseline model, the state space of the extended model can be reduced from uncountably infinity in time to a finite subset of the two firms' distances to the technology frontier (d_A, d_B). Again, I will restrict the discussion on equilibrium to the Markov Perfect Equilibrium, in which the state-dependent equilibrium value of performance measure of firm f is characterized by its value function $v_f(d_f, d_{-f})$, and its equilibrium strategy the policy function $a_f^*(d_f, d_{-f})$.

The argument for the existence of a Perfect Markov Equilbrium in the extended model is similar to that for the baseline model given by Proposition 1, only with an additional state variable. The proposition below lays out the system of equations to which the value functions v_f and policy functions a_f^* consist a solution:

Proposition 5 *There exists a Perfect Markov Equilibrium $a_f^*(d_f, d_{-f})$ for $f \in \{A, B\}$, where $(d_f, d_{-f}) \in \{0, \cdots, \overline{D}+\overline{m}\}^2, \min\{d_f, d_{-f}\} \leq \overline{D}$ and $|d_f - d_{-f}| \leq \overline{m}$. Moreover, the equilibrium strategy profile satisfies the following system of equations:*

$$v_f(d_f, d_{-f}) = \frac{1}{\lambda\left[a_f^*(d_f, d_{-f}) + a_{-f}^*(d_{-f}, d_f)\right] + h \cdot \mathbb{1}\{d_f \neq d_{-f}\} + \rho}\Big\{\pi_f(d_f, d_{-f})$$

$$
\begin{aligned}
& -\psi_f\left(a^*_f(d_f,d_{-f})\right)+\left[\lambda a^*_f(d_f,d_{-f})+h\cdot\mathbb{1}\{d_f>d_{-f}\}\right]v_f\left(T_f(d_f,d_{-f})\right)\\
& +\left[\lambda a^*_{-f}(d_{-f},d_f)+h\cdot\mathbb{1}\{d_f<d_{-f}\}\right]v_f\left(T_{-f}(d_f,d_{-f})\right)\Big\} \qquad (32)
\end{aligned}
$$

$$
\frac{d\psi_f\left(a^*_f(d_f,d_{-f})\right)}{da_f}=\lambda\left[v_f\left(T_f(d_f,d_{-f})\right)-v_f(d_f,d_{-f})\right] \qquad (33)
$$

$$
a^*_f(0,\overline{m})=0 \qquad (34)
$$

The proof of the above proposition is skipped, as it acquires nothing more than a trivial modification of that of Proposition 1.

Due to the lack of an explicit expression of firm value function, it is difficult to see the impact of destructive innovation shock in a qualitative way. However, it can be shown that with fixed technology gap between the two firms, a step farther from the technology frontier always means lower value for either firm.

Lemma 4 *For all $f\in\{A,B\}$ and $(d_f,d_{-f})\in\mathbb{N}^2_+$, $v_f(d_f,d_{-f})>v_f(d_f+1,d_{-f}+1)$ if and only if v_f is defined at (d_f,d_{-f}) and $(d_f+1,d_{-f}+1)$.*

Proof. See Appendix.

Corollary 3 *For all $f\in\{A,B\}$ and $(d_f,d_{-f})\in\mathbb{N}^2_+$, $v_f(d_f+1,d_{-f})<v_f(d_f,d_{-f})<v_f(d_f,d_{-f}+1)$ if and only if v_f is defined at all of these states.*

The proof of Corollary 3 is similar to that of Lemma 4.

Directly from Lemma 4, the destructive innovation shock lowers the values of both firms, and the degree of such destruction is increasing in the magnitude D as in Definition 5:

Proposition 6 *If there is a destructive innovation shock at time $t>0$, for all $f\in\{A,B\}$, $v_f\left(d_f(t),d_{-f}(t)\right)<\lim\limits_{s\to t^-}v_f\left(d_f(s),d_{-f}(s)\right)$, and $\lim\limits_{s\to t^-}v_f\left(d_f(s),d_{-f}(s)\right)-v_f\left(d_f(t),d_{-f}(t)\right)$ is strictly increasing in the magnitude of shock, $D\in\{1,\cdots,\overline{D}\}$.*

Proposition 6 corresponds to the empirical observation in section 2.2, where high-valued innovations are found to be associated with the loss in sales of rival firms. The model here not only accounts for this fact, but also claims that the degree of loss is in positive proportion to the magnitude of the destructive innovation shock. As a caveat, in empirical work, it would be difficult to distinguish between the magnitude of shock, D, from the stepwise destructiveness δ as in profit function (27).

4.3 Responses of R&D to Destructive Innovations

In this subsection I employ the extended model to study firms' responses in R&D efforts to the destructive innovation shock, especially to see if there is heterogeneity directions of such responses. This is interesting because the heterogeneous responses in R&D may help to explain the comovement between the patent value from major innovations and productivity dispersion (Figure 1).

Without explicit expressions of firms' R&D efforts as functions of state variables, the model is silent on the responses in R&D to the destructive innovation shock of any arbitrary degree, and I shall later let numerical experiments speak on that question. In the following I analyze the R&D responses to shocks with extreme degrees.

When the magnitude of shock, D, is sufficiently large, so that the aggregate gross profit of the two firms is vanishingly small, both firms reduce their R&D efforts in response to the shock, regardless of the functional form or parameters of the R&D cost functions.

Proposition 7 *If $\delta > 0$ is small enough, there exists $\widehat{D} \in \{1, \cdots, \overline{D}\}$, such that if there is a destructive innovation shock with magnitude $D \geq \widehat{D}$ at time $t > 0$, $a_A^*(t) < \lim_{s \to t^-} a_A^*(s)$ and $a_B^*(t) < \lim_{s \to t^-} a_B^*(s)$.*

Proof. See Appendix.

Proposition 7 claims that when the magnitude of the shock is large enough, both firms will respond by reducing their R&D efforts. When the destruction to profit is so overwhelming and there is a long way to go before the loss can be recovered, both firms will find the marginal value of an innovation lower than before, thus adjust their R&D efforts downward.

At the other extreme, when the magnitude of shock is minimal, the firm on the technology frontier will always respond by increasing its R&D effort.

Proposition 8 *If a destructive innovation shock with magnitude 1 occurs at time $t > 0$ and $\lim_{s \to t^-} d_f(s) = 0$, then $a_f^*(t) > \lim_{s \to t^-} a_f^*(s)$.*

Proof. See Appendix.

Upon being hit by the shock with degree 1, the current firm value is strictly lower for the leading firm, as can be seen from Lemma 4. However, this loss can be fully recovered by one innovation, whose marginal value is hence higher than that before the shock. This is why at least the leader has incentive to do more R&D after small shocks.

From the above two propositions, it is clear that when δ in Assumption 3 is small enough, for each firm f, there is a minimal threshold of magnitude of destructive innovation shock, above which its response will be to lower R&D effort. Denote that minimum threshold by $\widehat{D}_f(d_f, d_{-f})$ to reflect the fact that it depends on the state just prior to the shock. Specifically, if firm f is on the technology frontier prior to the shock, $\widehat{D}_f(0, d_{-f})$ is strictly greater than 1.

Suppose there is a leader and a laggard just prior to the destructive innovation shock. Without loss of generality, let them be firm A and B, respectively. Following the above analysis, if it is the case that $\widehat{D}_A(0, d_B) > \widehat{D}_B(d_B, 0)$, then there exists magnitude of shock D satisfying $\widehat{D}_B(d_B, 0) \leq D < \widehat{D}_A(0, d_B)$ to which the leader responds by higher R&D effort and the laggard lower. consequently, in a period after the shock, the expected technology gap is strictly higher than that prior to the shock. This is the channel through which the innovation shock contributes to enlarged technology gap, and it explains the empirical observation of greater productivity dispersion that is likely to be caused by high-valued innovations from competing firms (Figure 3). The above pattern is found in simulation of the extended model in Section 6, and confirmed empirically in Section 7.

5 Model Parameterization

To ready the extended model for numerial experiments and see how firms in my model react to the destructive innovation shock in the simulation, in this section I assign data-based parameter values. I will use the baseline model in Section 3 for calibration and *generalized method of moments* (GMM) estimation[22]. This is valid because the sole extra parameter in the extended model is the destructiveness of the shock, δ, in the profit function (27). The choice of its value doesn't affect the other parameters in the stationary equilibrium without shock.

To begin with, use the conventional values of the discount factor $\rho = 0.1$ and the multiplier on the R&D effort $\lambda = 1$, from Ludkovski and Sircar (2016) and Aghion et al. (2005), respectively[23].

So far the analytical results do not rely on specifications of profit functions $\pi_f(\Delta n_f)$ and R&D cost functions $\psi_f(a_f)$ for $f \in \{A, B\}$. However, I need to adopt explicit functional

[22] An alternative approach, *simulated method of moments* (SMM), is also feasible. However, the GMM is preferable in this case as stationary distribution of the states is available. It is computationally easier and free of additional assumptions required by SMM. For the latter, see Duffie and Singleton (1993). Strebulaev and Whited (2012) provides a practical guide for both approaches.

[23] In Aghion et al. (2005), the arrival rate of innovation equals the R&D effort n (a_f in my notation) plus the rate of spillover h. This is equilvalent to setting $\lambda = 1$ in my model.

forms of them to pin down paramters, for which my choices are:

$$\pi_f(\Delta n_f) = \frac{\overline{\pi}}{2}\left(1 + \frac{\Delta n_f}{\overline{m}}\right), \tag{35}$$

$$\psi_f(a_f) = \frac{\kappa_f}{2} a_f^2. \tag{36}$$

The profit function π_f is linearly increasing in the technology gap Δn_f of firm f with slope $\frac{\overline{\pi}}{2\overline{m}}$. If the firm is at the largest possible gap, $\Delta n_f = \overline{m}$, it takes all the profit $\overline{\pi}$ on the market; if it lags behind to the extreme degree, i.e. $\Delta n_f = -\overline{m}$, it has zero profit. It is easy to verify that profit function (35) satisfies Assumption 1. For the numerical experiments in the next section I use the value $\overline{m} = 5$ and it doesn't have qualitative influence on the outcome[24]. The quadratic R&D cost function (36) is standard in the literature and satisfies Assumption 2. Without loss of generality, impose the constraint that the marginal R&D cost (thus also the R&D cost) of firm A is strictly less than that of firm B: $0 < \kappa_A < \kappa_B$.

Combining the R&D cost function (36) and the first-order condition (17) of the R&D effort, the parameter κ_f for $f = A, B$ can be calibrated as follows:

$$\kappa_f = \frac{v_f(\Delta n_f + 1) - v_f(\Delta n_f)}{a_f^*(\Delta n_f)} = \frac{\left[v_f(\Delta n_f + 1) - v_f(\Delta n_f)\right]^2 / 2}{\psi_f\left(a_f^*(\Delta n_f)\right)} \qquad \forall\ \Delta n_f, \tag{37}$$

where the second equality is from equation (36). In the data, the equilibrium R&D cost $\psi_f(a^*)$ can be approximated using firm-year R&D intensity. And I use the firm-year value of patents, scaled by sale, as the proxy of the increment in firm value from innovaton on the numerator. This is because the stock market's reponse to patent grants in Kogan et al. (2017) itself reflects the value added to the firm by innovations.

I use firms with RDI between the second and third deciles of the industry-year cell to calibrate κ_B, and use those between the seventh and eighth deciles for that of κ_A. The range between the two groups of firms is consistent with the fact that the productivity dispersion is often measured by the inter-quartile range.

The details regarding the moment conditions are left to the appendix. The calibrated values are $\kappa_A = 0.0868$ and $\kappa_B = 0.1057$. In the data, the firms with high R&D intensity on

[24] I refrain from calibrating or estimating the maximum gap $\overline{m}$ because it is difficult to say what one unit of gap corresponds to in reality. And the choice of its value in the literature is arbitrary. Therefore in this paper I am not aiming to pin down this parameter, but would like to emphasize that if one is to add $\overline{m}$ into the parameterization as an extra degree of freedom, the fitting of the model will be no worse than with the arbitrary choice of $\overline{m} = 5$.

average are found to have lower R&D cost, as would be predicted by Lemma 3.

The two parameters remains to be determined are $\overline{\pi}$ and h. The former is the total gross profit of the two firms, and the later the rate of imitation with which the laggard automatically advances one step along the technology ladder independent of its R&D effort. For these two parameters there are no analytical moment conditions similar to (37). However, with stationary distribution available from Definition 3 and Proposition 2, it is feasible to estimate them by GMM. There needs to be two target moments: the first one I pick is the expected ratio of the RDI of low-cost firms to that of the high-cost firms; the second is the expected ratio of the value of innovation between the two types of firms. Again, the technical details on the procedure of the estimation are left to the appendix. The estimated parameter values are $\pi = 0.0924$ and $h = 0.0772$.

Table 1 summarizes the list of the model parameters determined as above.

Table 1: Model Parameters

Parameter	Description	Source	Value
ρ	discount factor	Ludkovski and Sircar (2016)	0.1
λ	multiplier on R&D effort	Aghion et al. (2005)	1
$\overline{m}$	maximum technology gap	assigned	5
κ_A	R&D cost parameter of firm A	calibrated	0.0868
κ_B	R&D cost parameter of firm B	calibrated	0.1057
$\overline{\pi}$	total profit	GMM	0.0924
h	rate of spillover	GMM	0.0772

The table below reports the performance of calibration and estimation in fitting moments. The perfect fitting of the first two moments is due to the explicit expression of κ_f in (37), where κ_f is consistent over all states Δn_f thus invariant to the stationary distribution.

Table 2: Fitting of Moments

Description	Data	Model
Mean of ratio (37) of low-cost firms	0.0868	0.0868
Mean of ratio (37) of high-cost firms	0.1057	0.1057
Average ratio of R&D intensities, low-cost to high-cost firms	3.9053	3.9040
Average ratio of innovation values, low-cost to high-cost firms	3.4108	3.4094

In Table 2, the first two moments informs κ_A and κ_B, respectively. The latter two moments jointly informs $\overline{\pi}$ and h in the GMM estimation. The baseline model does a satisfactory job in fitting the selected moments from the data, thus justifies the usage of its extended version to study the impact of destructive innovation shocks.

6 Quantitative Analysis

With the model parameters obtained in Section 5, I proceed by showing the performance of the model in simulation. I will firstly show the numerical solutions of the value functions and policy functions of the two firms in the baseline model from Section 3, where the sole state variable is the technology gap[25]. Then I turn to the numerical study of the extended model from Section 4, where the state is two-dimensional in firms' distances to the technology frontier. The simulation of this extended model helps in understanding the pattern of firms's responses in R&D to the destructive innovation shock, and its implication on the dynamics of technology gap.

6.1 Numerical Solution of the Baseline Model

The baseline model is numerically solved using value function iteration[26]. Firstly set a initial value function $\boldsymbol{v}_0 = \{v_{0,A}(m), v_{0,B}(m)\}_{m \in \mathcal{M}}$, for example $v_f(m) = 0$ for all $m \in \mathcal{M}$ and $f \in \{A, B\}$. Then calculate the policy function $\boldsymbol{a}_0^* = \{a_{0,A}^*(m), a_{0,B}^*(m)\}_{m \in \mathcal{M}}$ according to equations (17) and (18). For the next step, update value function $v_{1,f}(m)$ using $\boldsymbol{v}_0$ and $\boldsymbol{a}_0^*$ on the right-hand side in (16) and (19) to get $\boldsymbol{v}_1$. Keep iterating until the sequence of value functions $\{\boldsymbol{v}_k\}_{k=0}^{\infty}$ converges. The outcome of the iteration is plotted in the figure below:

The left panel of Figure 5 conveys two messages. Firstly, for either the low-cost firm (firm A in this example) or the high-cost firm (firm B), the value function is strictly increasing in its technology gap. This is intuitive as a firm benefits from a large technology gap not only from a high profit flow when it stays at the current state, but also from the chance of jumping to an even larger gap where profit is higher. Secondly, for any state $m \in \mathcal{M} = \{-\overline{m}, -\overline{m}+1, \cdots, \overline{m}-1, \overline{m}\}$, it is the case that $v_A(m) > v_B(m)$, meaning the low-cost firm's

[25]Strictly speaking, there are two state variables in the baseline model: the gaps of firm A and B. However, recall that there is a bijection between these two, as by definition one is the opposite to the other: $\Delta n_A = n_A - n_B = -(n_B - n_A) = -\Delta n_B$.

[26]The function forms of the profit and R&D cost functions are the same as (35) and (36) which I use for model parameterization.

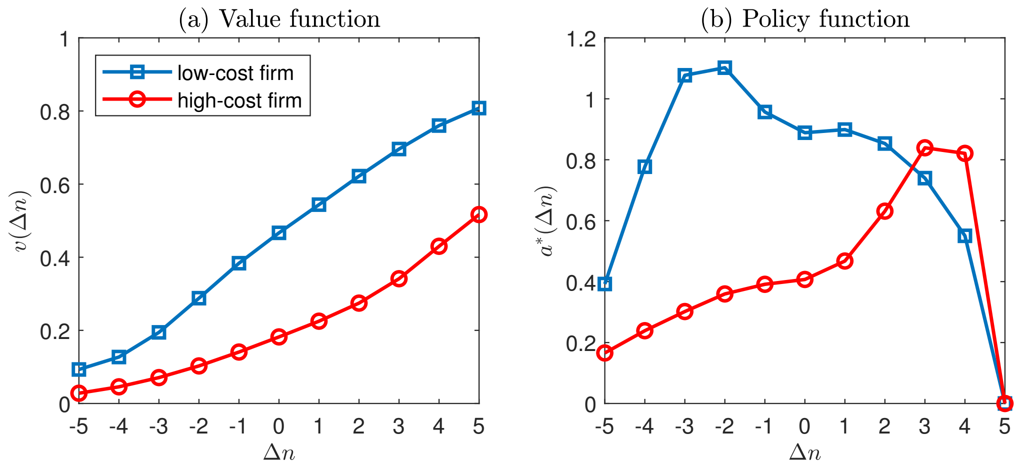


Figure 5: Numerical solution of the baseline model

value strictly dominates that of the high-cost firm's. This is consistent with Lemma 2[27], whose explanation is that in the equilbrium, the low-cost firm can achieve a performance measure at least the same as the high-cost firm, and from that there's still room for improvement by optimizing.

It is not the case, however, that the low-cost firm will always choose higher R&D effort than the high-cost firm, as is suggested by the right panel. The figure of policy functions implies that unconditionally, the low-cost firm is more likely to be the leader (since $a_A^*(0) > a_B^*(0)$); but conditional on being the leader with gap $\Delta n_f \geq 3$, the high-cost firm would make greater effort to maintain its advantage. One possible incentive behind this is the high-cost firm treasures the chance of being the leader more because it is more rare for it.

From above it is obvious that Lemma 3 doesn't extend to cases where the maximum technology gap $\overline{m} > 1$. Nonetheless, arguments similar to Proposition 4 and Corollary 2 still holds in the example here with $\overline{m} = 5$. That is, for any $m \in \{1, 2, \cdots, 5\}$, the stationary probability that the low-cost firm is at state m is strictly higher than that at $-m$: $\mu_A(m) > \mu_A(-m)$, $\forall m > 0$. Consequently, the expected technology gap of the low-cost firm under the stationary distribution is strictly positive: $\mathbb{E}[\Delta n_A(t)] > 0$. This can be seen from the following figure depicting the stationary distribution of firm A's technology gap.

[27]In the special case of quadratic R&D costs $\psi_f(a_f) = \frac{\kappa_f}{2} a_f^2$ with $\kappa_A < \kappa_B$, the condition in Lemma 2 that $\frac{d\psi_A(a)}{da_A} < \frac{d\psi_B(a)}{da_B}$ for all $a > 0$ is satisfied.

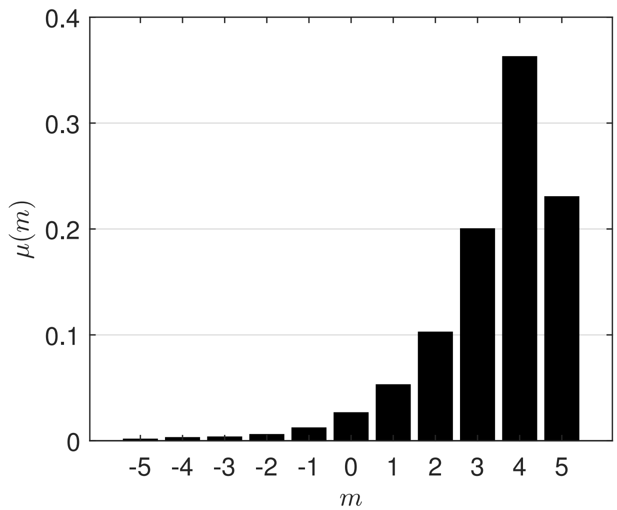


Figure 6: Stationary distribution of firm A's technology gap

From Figure 6, for any $m > 0$, the stationary probability for firm A's technology gap to be $\Delta n_A = m$ is greater than that of $\Delta n_A = -m$. And its expected technology gap is calculated to be 3.4073, thus the low-cost firm is expected to be the leader in the long run. So far, the properties of the the numerical solution are in accordance with the theoretical arguments for the baseline model.

6.2 Numerical Solution of the Extended Model

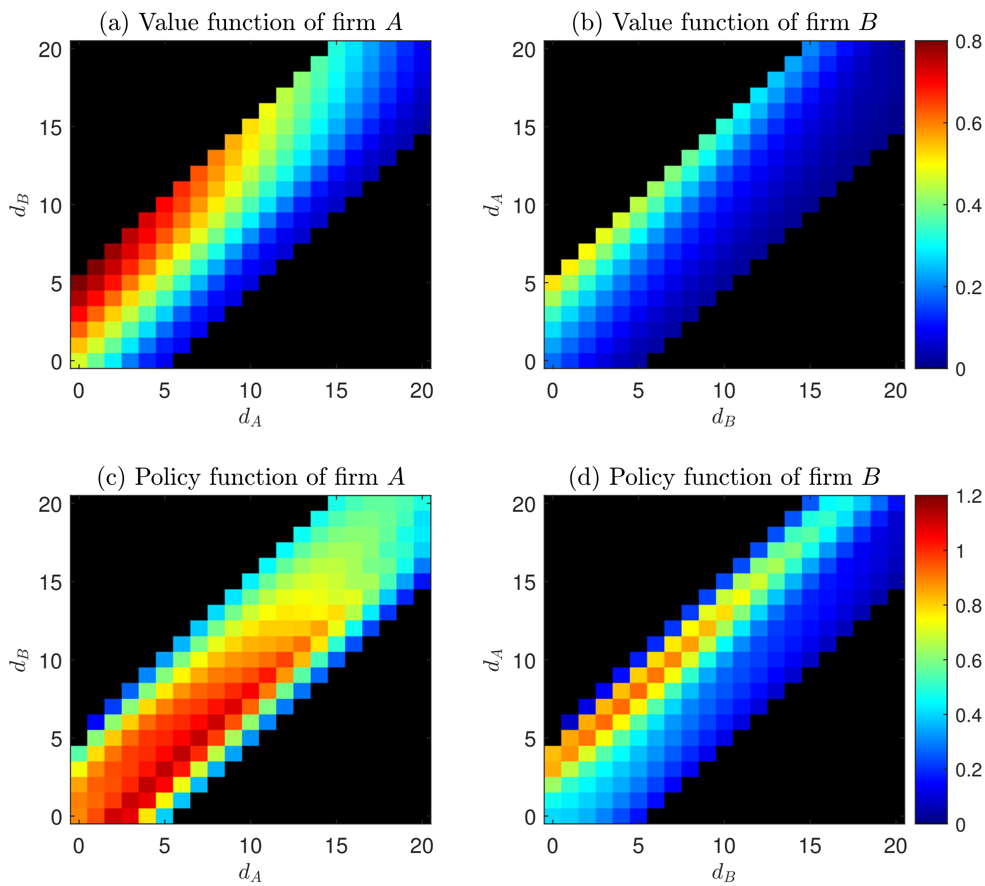
With the destructive innovation shock absent, the baseline and the extended models are equivalent. From now on, I use the extended model to see how the two firm with heterogeneous R&D costs respond to the destructive innovation shock, and how would that impact the dynamics of the technology gap.

Since the profit function is different from that in the baseline model, there needs to be a new specification in the form of equation (27), where I arbitrarily assign $\delta = 0.05$, and

$$\widetilde{\pi}_f(d_f, d_{-f}) = \frac{\overline{\pi}}{2}\left(1 + \frac{d_{-f} - d_f}{\overline{m}}\right). \tag{38}$$

Thus one degree of the destructive innovation shock reduce the total profit of the two firms by 5 percent. And just like in the baseline model, the share of the total profit by firm $f \in \{A, B\}$ is linearly increasing in its technology gap, which is equivalent to the difference in the distances to the technology frontier: $\Delta n_f = d_{-f} - d_f$. The R&D cost function admits the same form as in the baseline model (36). This is how the whole dynamical system reduces to the baseline model without the shock.

The value function $v_f(d_f, d_{-f})$ and policy function $a^*_f(d_f, d_{-f})$ are again solved numerically using value function iteration. Since they are now two-dimensional, the outcomes are represented with heatmaps as follows, where black blocks correspond to zero value. The symmetric off-diagonal black areas reflect the restriction that the absolute difference between d_A and d_B cannot exceed the maximum technology gap $\overline{m} = 5$.



For firm f the horizontal and vertical axes are the distances to the technology frontier of its own and of its rival's, d_f and d_{-f}, respectively. The two panels in the same row share a common color bar.

Figure 7: Numerical solution of the extended model

Compare panels (a) and (b) in Figure 7, for either firm, its value is higher when it's closer to the technology frontier, or when its leading position is more prominent. More precisely, for any fixed d_f, the value function $v_f(d_f, d_{-f})$ is increasing in its technology gap $d_{-f} - d_f$. Also, for any integers $d_1, d_2 \in \{0, 1, \cdots, 20\}$, it always holds that $v_A(d_1, d_2) > v_B(d_1, d_2)$. Both of these two patterns extend those shown by panel (a) of Figure 5 for the baseline model.

From panels (c) and (d) it is clear that the equilibrium R&D effort of either firm is not monotone in technology gap. Firm A (the low-cost firm) tends to exert higher R&D effort when it is lagged behind, as the high values in panel (c) is below the diagonal; for firm B (the high-cost firm) it is the opposite: its high R&D efforts are seen when it's at the leading position. This coincides with the patterns of R&D efforts in the baseline model, as is shown by panel (b) of Figure 5.

An interesting contrast between the two rows of Figure 7 is that, for the value functions, fix any colored grid $(d_f, d_{-f}) < (20, 20)$, it is always the case that $v_f(d_f, d_{-f}) > v_f(d_f + k, d_{-f} + k)$ for any integer $k > 0$ with which the value function is defined. It means a firm's value function is strictly decreasing in the degree of the destructive innovation shock, as stated by Lemma 4. This is not true for the policy functions: it is easy to see in panel (c) that the equilibrium R&D effort of firm A increases in the degree of shock (darker color towards red along the diagonal) before it finally decreases. The same can be found in panel (d) but is less obvious due to the small scale of change in color.

Furthermore, a closer inspection will show that the areas where the shock causes firm A to raise its R&D effort are those would lead firm B to cut down its R&D. This verifies the existence of certain magnitudes of the destructive innovation shock, to which the firms respond by adjusting R&D in opposite directions, as conjectured in the end of Section 4.

6.3 Simulation

As the first step of simulating the continuous-time dynamical system characterized in Section 4, I discretize the time horizon to $T = 1,000$ periods. The simulation is repeated by $B = 50,000$ times. For each repetition b, I create two vectors of dimension $1 \times T$, $\widehat{d}_{A,b}$, $\widehat{d}_{B,b}$ to restore the realization of the distances to the technology frontier. Set $\widehat{d}_{A,b}(1) = \widehat{d}_{B,b}(1) = 0$ for all $b = 1, 2, \cdots, B$, such that for all repetitions the system begins with the state that the two firms are neck-to-neck on the technology frontier.

By the Markov property of the equilibrium, at each period $t = 1, 2, \cdots, T$, the R&D effort chosen by firm f depends on the current state $\widehat{d}_{f,b}(t)$ only. Store the time-paths of R&D efforts of firms A and B in vectors $\widehat{a}_{A,b}$ and $\widehat{a}_{B,b}$. Obviously, for any $t = 1, 2, \cdots, T$, $\widehat{a}_{f.b}(t) = a^*_f\left(\widehat{d}_{f,b}(t), \widehat{d}_{-f,b}(t)\right)$, where $a^*_f(d_f, d_{-f})$ is firm f's policy function. For each repetition b, from $t = 2$ on, the probability distribution of the state $\left(\widehat{d}_{A,b}(t), \widehat{d}_{B,b}(t)\right)$ is determined by the last period's state and R&D efforts $\left(\widehat{d}_{A,b}(t-1), \widehat{d}_{B,b}(t-1)\right)$. The details on discretization and updating rule of state are left in the appendix.

For each repetition, let there be a destructive innovation shock at $t = 500$[28] with degree $D = 10$. That means after the realization of the state $\left(\widehat{d}_{A,b}(500), \widehat{d}_{B,b}(500)\right)$, I change it to $\left(\widehat{d}_{A,b}(500) + 10, \widehat{d}_{B,b}(500) + 10\right)$[29]. The choice of degree $D = 10$ is to be consistent with the empirical finding in Section 2, that the major innovation destroys about 50% of its rivals' sales on average over a ten-year horizon. Thus in the extended model with the destructiveness of one degree of shock $\delta = 0.05$ as in (27), it requires a shock with degree 10 to be a proper analogy. The values of δ and D are not essential for the simulation qualitatively.

With the above design of shock, from $t = 1, \cdots, 499$, the leading firm is on the technology frontier, thus the dynamical system is equivalent to that of the baseline model in the sense the technology gap and distances satisfy $m_f = d_{-f} - d_f$ and $\min\{d_f, d_{-f}\} = 0$. At $t = 500$, the two firms are pushed backward from the technology frontier by ten steps simultaneously, while their relative positions remain unchanged. From this period on, until one firm returns to the frontier, the system can no longer be described by the baseline model, but only by the extended model.

The impulse response of expected R&D efforts to the destructive innovation shock is approximated by the mean of the simulated paths:

$$\widehat{a}_f(t) = \frac{1}{B} \sum_{1 \leq b \leq B} \widehat{a}_{f,b}(t), \tag{39}$$

where $\widehat{a}_f(t)$ without repetition subscript b is the mean of the simulated R&D effort of firm f at period t. For $t \geq 500$, it serves as the simulated impulse response function of firm f's R&D to the shock. Similarly, the simulated technology gap of firm f, $\Delta\widehat{n}_f$, is calculated as

$$\Delta\widehat{n}_f(t) = \frac{1}{B} \sum_{1 \leq b \leq B} \left[\widehat{d}_{-f,b}(t) - \widehat{d}_{f,b}(t)\right]. \tag{40}$$

The following figure reports the time paths of $\Delta\widehat{n}_A(t)$, $\widehat{a}_A(t)$ and $\widehat{a}_B(t)$:

[28] I leave some periods for the stochastic processes to evolve and approach to the limiting distribution.

[29] The R&D efforts at $t = 500$ are determined after the shock happens.

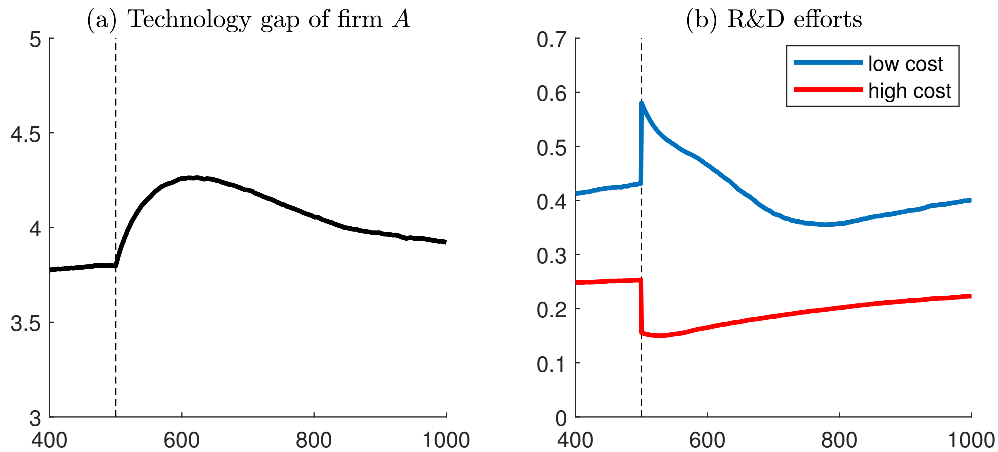


Figure 8: Simulation of the impact of destructive innovation shock

In panel (b), upon being hit by the shock, the low-cost firm instantly increases its R&D effort, and the high-cost firm responds by reducing its R&D effort. For a period afterward, the difference in R&D efforts is higher than the pre-shock level. This causes a hump-shaped increase in the technology gap, as shown in panel (a). Such change in technology gap is gradual rather than a jump because the path of gap must travel through all the intermediate states before reaching a higher level. In longer horizon, as the the difference in R&D converges back to the pre-shock level, the technology gap also closes.

Compare the simulaton results with the empirical findings. Panel (a) in Figure 8 resembles the impulse responses in Figure 1, Section 2, in that the response of productivity dispersion to the shock is gradual and persistent. Such consistency between the model's products and empirical observations suggests that the heterogeneous responses of firm-level R&D is a reasonable mechanism explaining the higher productivity dispersion as the effect of destructive innovation shock. The next section present empirical evidence confirming the heterogeneous responses of R&D in panel (b). Regression models in that section identify firms' responses of current year's R&D intensity to last year's destructive innovation shock, which exhibit the heterogeneity in line with that in panel (b).

7 Empirical Evidence on Heterogeneous Responses in R&D to Destructive Innovation

In this section I provide supportive evidence in favor of the claim from the extended model and its simulation, that firms with different R&D capacities will respond to the destructive innovations in opposite directions. Particularly, as the following empirical investigation reveals, it is the case that the high-capacity (low R&D cost) firm responds positively with its R&D intensity, while the low-capacity (high R&D cost) firm negatively. The approaches in this section extends Chen and Ming (2020).

The ability to conduct R&D activities, or the R&D cost function of a firm, is unobservable from the data. However, Lemma 3 predicts that low-cost firms invest more in R&D than high-cost firms. Thus the implication to be tested is formulated as follows: firms with high R&D intensity (RDI) will respond to destructive innovation shocks by increasing its RDI; while firms with low RDI will do the opposite.

The R&D intensities of firms in different industries are not directly comparable, because each industry may have its own standard about which levels of RDI should be considered high or low. Thus the value of industry-year *empirical cumulative distribution function* (ECDF) is used for comparison of RDI across firms. The ECDF value in RDI of firm i in industry j at year t is defined as follows:

$$ecdfRDI_{i,j,t} = \frac{\sum_{i'} \mathbb{1}\{\log(RDI_{i',j,t}) < \log(RDI_{i,j,t})\}}{\sum_{i'} \mathbb{1}\{\log(RDI_{i',j,t}) \in \mathbb{R}\}}. \tag{41}$$

In this way, $ecdfRDI \in [0, 1)$ gives the percentage of firms in an industry-year cell with RDI below a certain firm, thus it serves as a measure comparable across industries and years. The regression specification is as follows:

$$\log(RDI_{i,j,t}) = \alpha_i + \delta_t + \beta MI_{j,t-1} + \gamma MI_{j,t-1} \times ecdfRDI_{i,j,t-1} + \eta ecdfRDI_{i,j,t-1} + \boldsymbol{\lambda} \boldsymbol{X}_{i,j,t} + u_{i,j,t}, \tag{42}$$

where α_i and δ_t are firm and year fixed effects, respectively; MI is the major innovation index defined in (3). The set of control variables is denoted by $\boldsymbol{X}$, which includes the ECDF of revenue TFP obtained in the same way as in (41), market share, number of firms in the industry-year cell, and the Herfindahl-Hirschman index. The alternative setting $\boldsymbol{X} = \emptyset$ will also be used for robustness check.

With this specification, the coefficient γ of the cross-term $MI \times ecdfRDI$ is of interest.

It captures the heterogeneity in the reponses of current RDI from firms sorted by past RDI, to the destructive innovation shocks. For example, if the estimated $\widehat{\gamma}$ is significantly positive, the interpretation should be that firms with high RDI in the past reponds to destructive innovations by raising RDI, as compared to firms with low past RDI.

One concern about the above approach is that there may be high year-to-year fluctuation in one firm's ranking of its RDI relative to others from the same industry, thus the ECDF value in the last year could fail to reflect a firm's long-term performance which is more informative on its capacity to do R&D. I thus perform an additional robustness check by replacing the ECDF of annual RDI ($ecdfRDI$) in regression (42) by the ECDF of a three-year moving average of RDI, denoted by $ecdfRDI^{\text{MA3}}$:

$$ecdfRDI^{\text{MA3}}_{i,j,t} = \frac{\sum_{i'} \mathbb{1}\{\sum_{s=t-2}^{t} \log(RDI_{i',j,s}) < \sum_{s=t-2}^{t} \log(RDI_{i,j,s})\}}{\sum_{i'} \mathbb{1}\{\sum_{s=t-2}^{t} \log(RDI_{i',j,s}) \in \mathbb{R}\}}. \tag{43}$$

The outcome of regression (42) is reported in the following table:

Table 3: Responses of R&D Intensity to Destructive Innovation Shocks

$\log(RDI_t)$	(1)	(2)	(3)	(4)
MI_{t-1}	-0.181	-0.234^{**}	-0.143	-0.201^{*}
	(0.110)	(0.116)	(0.109)	(0.106)
$MI_{t-1} \times ecdfRDI_{t-1}$	0.343^{*}	0.369^{**}		
	(0.174)	(0.180)		
$ecdfRDI_{t-1}$	1.376^{***}	1.444^{***}		
	(0.128)	(0.136)		
$MI_{t-1} \times ecdfRDI^{\text{MA3}}_{t-1}$			0.242	0.281^{*}
			(0.161)	(0.161)
$ecdfRDI^{\text{MA3}}_{t-1}$			0.894^{***}	0.949^{***}
			(0.090)	(0.097)
Control variables	NO	YES	NO	YES
R^2	0.229	0.396	0.188	0.372
N	26,555	26,555	18,233	18,233

Notes: Standard errors clustered at the industry level are in parenthesis. ***, ** and * indicate significance at the 1%, 5% and 10% levels. For columns 1 and 2, observations of firms who were major innovators in the last year are excluded; for columns 3 and 4, those of firms who have been major innovators in the past three years are excluded.

In Table 3, columns 1 and 2 use the ECDF values based on the one-year lag measure defined by (41); columns 3 and 4 based on the alternative three-year moving average counterpart. For all specifications other than that of column 3, the coefficient of interest, which is on the cross-term between the major innovation indicator and the ECDF value of past RDI (in the second and fourth rows), are significantly positive. Again, this is in support of the existence of heterogeneity in firms' reponses in their RDI to destructive innovation shocks. Specifically, it is shown that firms with higher RDI in the past invest more in R&D in response to such shocks.

The first row in Table 3 captures the negative impact of the destructive innovations over all firms' RDI. Combining these estimated values with those reported in the second or fourth rows, it appears that the destructive innovations cause the least R&D intensive firms to drop their R&D intensity by about 20 percent; and they cause the most R&D intensive firms to increase their RDI by a degree ranging from 8.0 to 16.2 percent. For some firms with past RDI close to the industry median, the destructive innovation shock should have little or no impact on their current choice of R&D input. The above empirical evidence on the directions of heterogeneous R&D adjustment is consistent with the simulation in the previous section.

Since the treatment – the destructive innovation – is assigned at the industry level, the standard errors in Table 3 are clustered at the industry level using the method as in Liang and Zeger (1986), which according to Abadie et al. (2017) is conservative. Actually, under alternatives like robust standard errors or firm-level standard errors, the esimated coefficients reported in that table will all be significant at 1% level.

In regression (42), the effect of destructive innovation on the conditional expectation of a firm's R&D intensity is linearly increasing in the value of ECDF of its past RDI. In the rest of this section I will loosen this linearity constraint with non-parametric regressions, and study the responses of firms sorted into different deciles by their past RDI.

To do so, firstly divide the observations of R&D intensity from each industry-year cell into ten intervals by the industry-year specific deciles, $\overline{c}^s_{j,t}$, $s = 1, \cdots, 10$, where

$$\overline{c}^s_{j,t} = \underset{i'}{\operatorname{argmax}} \left\{ \log(RDI_{i',j,t}) : ecdfRDI_{i',j,t} \leq s/10 \right\} \tag{44}$$

in which the variable $ecdfRDI$ is defined by (41). Thus $\overline{c}^s_{j,t}$, $s = 1, \cdots, 10$ are the ten deciles of $\log(RDI)$ in industry j at year t. And secondly, generate ten dummy variables $c^s_{i,j,t}$ from $s = 1$ to $s = 10$, defined as follows:

$$c^1_{i,j,t} = \begin{cases} 1 & \text{if} \quad \log(RDI_{i,j,t}) \leq \overline{c}^1_{j,t} \\ 0 & \text{otherwise} \end{cases} \tag{45}$$

$$c_{i,j,t}^{s} = \begin{cases} 1 & \text{if} \quad \overline{c}_{j,t}^{s-1} < \log(RDI_{i,j,t}) \leq \overline{c}_{j,t}^{s} \text{ and } 2 \leq s \leq 9 \\ 0 & \text{otherwise} \end{cases} \tag{46}$$

$$c_{i,j,t}^{10} = \begin{cases} 1 & \text{if} \quad \log(RDI_{i,j,t}) > \overline{c}_{j,t}^{9} \\ 0 & \text{otherwise} \end{cases} \tag{47}$$

For example, $c_{i,j,t}^{3} = 1$ means that firm i's log-RDI at year t falls between the second and third deciles among the log-RDIs from all firms in industry j at the same year. Finally, use the following specification to identify the responses in RDI to destructive innovations of firms in different groups partitioned by (45) to (47):

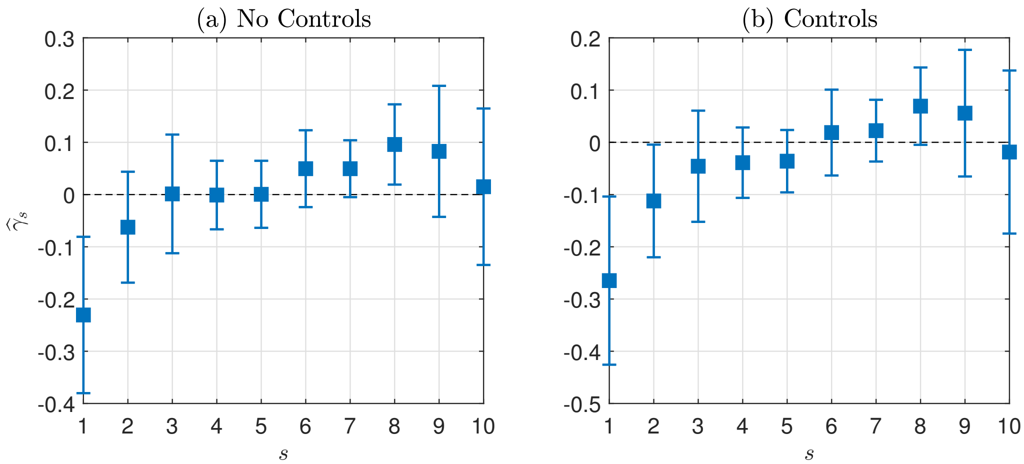
$$\log(RDI_{i,j,t}) = \alpha_i + \delta_t + \sum_{s=1}^{10} \gamma_s MI_{j,t-1} \times c_{i,j,t-1}^{s} + \sum_{s=1}^{10} \eta_s c_{i,j,t-1}^{s} + \boldsymbol{\lambda X}_{i,j,t} + u_{i,j,t}. \tag{48}$$

Similar to specification (42), α_i and δ_t are firm and year fixed effects. MI stands for the major innovation indicator; $\boldsymbol{X}$ the same set of control variables. The single term MI is omitted because the group of cross-dummies $\left\{MI_{j,t} \times c_{i,j,t}^{s}\right\}_{s=1}^{10}$ is fully saturated, thus including $MI_{j,t}$ incurs perfect multicolinearity.

The estimates $\{\widehat{\gamma}_s\}_{s=1}^{10}$ captures the decile-grouped, potentially heterogeneous responses of RDI to destructive innovations. Their estimated values and 90% confidence intervals are plotted in the figure below.

The interpretation of $\widehat{\gamma}_s$ in specification (48) is similar to $\widehat{\beta}+\widehat{\gamma}$ in (42), which is the impact of destructive innovations on RDI of firms in decile group s. The message from Figure 9 is consistent with that from Table 3: such impact will be the reduction in RDI for low-RDI firms, as well as the increase in RDI for high-RDI firms. For firms with past RDI close the the median (from the second to the fifth deciles in the industry-year cell, to be precise), there is no significant impact from destructive innovation on current RDI. This is also true for the most R&D intensive firms in the 10-th decile group, about its reason I do not have an tested explanation. All these patterns are robust whether control variables are included or not.

To summarize, my theoretical model in Section 4 predicts opposite responses of firms' R&D to destructive innovation shock. However, as the parameter values and functional forms are undetermined, I don't give sufficient conditions on the range of the magnitude of shock which triggers this type of heterogeneous responses. Assigned with data-based parameter values and reasonable specification of functions, the simulation produces the predicted op-



Panel (a) plots the decile-grouped responses of firm-level RDI to destructive innovations ($\{\widehat{\gamma}^s\}_{s=1}^{10}$) and the 90% confidence intervals using no control variables; panel (b) plots those regressed with control variables. The confidence intervals are based on robust standard errors clustered on the industry level.

Figure 9: Decile-grouped responses of firm RDI to major innovations

posite directions of instantaneous responses in R&D efforts, and persistently larger gap in both R&D and technology (8). This prediction is further confirmed by empirical tests documented in this section, suggesting the model captures the channels through which destructive innovations impact the dynamics of productivity dispersion.

8 Conclusion

This paper invesigates the role of destructive innovations as a source of the rapidly rising productivity dispersion in the U.S. since 1990. A firm's achievement with high-valued innovations in a certain period destroys its rival firms' sales, to which the industry-level productivity dispersion responses by a hump-shaped increase. To explore the mechanism, I build a model where firms' positions along a technology ladder, as well as their technology gap, are stochastically determined by their R&D efforts. I simulate the impact of the destructive innovation shock, to which the firm with low R&D cost responds by raising its R&D effort, and the high-cost firm by lowering. Thus the cross-firm R&D difference jumps instantaneously, and converge gradually to its pre-shock level. During this process, the technology gap, reflecting the dynamic changes in R&D efforts, exhibits a hump-shaped impulse response. This heterogeneity in impact on R&D provides an theoretical explanation on the relationship between destructive innovation and productivity dispersion, and is confirmed by empirical tests showing heterogeneous responses of R&D by firms with high and low past R&D intensities, as implied by the theory.

Restricted by the nature of data, I do not discuss the pattern of firm exit and entry in response to destructive innovations. It would be reasonable to postulate though, if allowed, firm(s) close to the edge of zero value will exit upon the shock, thus the aggregate productivity would improve either by selection or reallocation. This might be a future work worthwhile if data permits.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey Wooldridge. 2017. "When Should You Adjust Standard Errors for Clustering?" National Bureau of Economic Research Working Paper 24003.

Acemoglu, Daron, Ufuk Akcigit, and Murat Alp Celik. 2014. "Young, Restless and Creative: Openness to Disruption and Creative Innovations." National Bureau of Economic Research Working Paper 19894.

Acemoglu, Daron, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William Kerr. 2018. "Innovation, Reallocation, and Growth." *American Economic Review*, 108(11): 3450–91.

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica*, 60(2): 323–351.

Aghion, Philippe, Christopher Harris, Peter Howitt, and John Vickers. 2001. "Competition, Imitation and Growth with Step-by-Step Innovation." *The Review of Economic Studies*, 68(3): 467–492.

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt. 2005. "Competition and Innovation: an Inverted-U Relationship." *The Quarterly Journal of Economics*, 120(2): 701–728.

Andrews, Dan, Chiara Criscuolo, and Peter N. Gal. 2016. "The Best versus the Rest: The Global Productivity Slowdown, Divergence across Firms and the Role of Public Policy." OECD Publishing OECD Productivity Working Papers 5.

Barth, Erling, Alex Bryson, James C. Davis, and Richard Freeman. 2016. "It's Where You Work: Increases in the Dispersion of Earnings across Establishments and Individuals in the United States." *Journal of Labor Economics*, 34(S2): S67–S97.

Berlingieri, Giuseppe, Patrick Blanchenay, and Chiara Criscuolo. 2017. "The Great Divergence(s)." OECD Publishing OECD Science, Technology and Industry Policy Papers 39.

Broda, Christian, and David E. Weinstein. 2010. "Product Creation and Destruction: Evidence and Price Implications." *American Economic Review*, 100(3): 691–723.

Caballero, Ricardo J., and Mohamad L. Hammour. 1994. "The Cleansing Effect of Recessions." *The American Economic Review*, 84(5): 1350–1368.

Celik, Murat Alp, and Xu Tian. 2020. "Agency Frictions, Managerial Compensation, and Disruptive Innovations." Working Paper.

Chen, Shuowen, and Yang Ming. 2020. "Heterogeneous Responses of R&D to Destructive Innovations." Working Paper.

Cinlar, Erhan. 1975. *Introduction to Stochastic Processes.* Englewood Cliffs, NJ:Prentice-Hall, Inc.

Corbae, Dean, Maxwell B. Stinchcombe, and Juraj Zeman. 2009. *An Introduction to Mathematical Analysis for Economic Theory and Econometrics.* Princeton, NJ:Princeton University Press.

Decker, Ryan A, John C Haltiwanger, Ron S Jarmin, and Javier Miranda. 2018. "Changing Business Dynamism and Productivity: Shocks vs. Responsiveness." National Bureau of Economic Research Working Paper 24236.

Decker, Ryan A., John Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2016. "Declining Business Dynamism: What We Know and the Way Forward." *The American Economic Review*, 106(5): 203–207.

Duffie, Darrell, and Kenneth J. Singleton. 1993. "Simulated Moments Estimation of Markov Models of Asset Prices." *Econometrica*, 61(4): 929–952.

Foster, Lucia, Cheryl Grim, John Haltiwanger, and Zoltan Wolf. 2019. "Innovation, Productivity Dispersion, and Productivity Growth." In *Measuring and Accounting for Innovation in the Twenty-First Century. NBER Chapters.* National Bureau of Economic Research, Inc.

Foster, Lucia S., Cheryl A. Grim, John Haltiwanger, and Zoltan Wolf. 2017. "Macro and Micro Dynamics of Productivity: From Devilish Details to Insights." National Bureau of Economic Research Working Paper 23666.

Gallager, Robert G. 1995. *Discrete Stochastic Processes. The Kluwer International Series in Engineering and Computer Science*, Norwell, MA:Kluwer Academic Publishers.

Garcia-Macia, Daniel, Chang-Tai Hsieh, and Peter J. Klenow. 2019. "How Destructive Is Innovation?" *Econometrica*, 87(5): 1507–1541.

Gouin-Bonenfant, Emilien. 2020. "Productivity Dispersion, Between-Firm Competition, and the Labor Share." Working Papers.

Grossman, Gene M., and Elhanan Helpman. 1991. "Quality Ladders in the Theory of Growth." *The Review of Economic Studies*, 58(1): 43–61.

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda. 2013. "Who Creates Jobs? Small versus Large versus Young." *The Review of Economics and Statistics*, 95(2): 347–361.

Herfindahl, Orris Clemens. 1950. "Concentration in the steel industry." PhD diss. Columbia University New York.

Hirschman, Albert O. 1945. *National Power and The Structure of Foreign Trade.* University of California Press.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *The Quarterly Journal of Economics*, 124(4): 1403–1448.

Jordà, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review*, 95(1): 161–182.

Kehrig, Matthias. 2015. "The Cyclical Nature of the Productivity Distribution." Working Papers.

Klette, Tor Jakob, and Samuel Kortum. 2004. "Innovating Firms and Aggregate Innovation." *Journal of Political Economy*, 112(5): 986–1018.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. 2017. "Technological Innovation, Resource Allocation, and Growth." *The Quarterly Journal of Economics*, 132(2): 665–712.

Liang, Kung-Yee, and Scott L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika*, 73(1): 13–22.

Ludkovski, Michael, and Ronnie Sircar. 2016. "Technology Ladders and R&D in Dynamic Cournot Markets." *Journal of Economic Dynamics and Control*, 69: 127–151.

Maskin, Eric, and Jean Tirole. 2001. "Markov Perfect Equilibrium: I. Observable Actions." *Journal of Economic Theory*, 100(2): 191–219.

Miao, Jianjun. 2014. *Economic Dynamics in Discrete Time.* Cambridge, MA:The MIT Press.

Norris, J.R. 1998. *Markov Chains. Cambridge Series in Statistical and Probabilistic Mathematics*, New York, NY:Cambridge University Press.

Ramey, Valerie A. 2016. "Macroeconomic Shocks and Their Propagation." In . Vol. 2 of *Handbook of Macroeconomics*, , ed. John B. Taylor and Harald Uhlig, Chapter 2, 71–162. Elsevier.

Schumpeter, Joseph A. 1942. *Capitalism, Socialism and Democracy.* New York, NY:Harper and Brothers.

Segerstrom, Paul S., T. C. A. Anant, and Elias Dinopoulos. 1990. "A Schumpeterian Model of the Product Life Cycle." *The American Economic Review*, 80(5): 1077–1091.

Strebulaev, Ilya A., and Toni M. Whited. 2012. *Dynamic Models and Structural Estimation in Corporate Finance. Foundations and Trends in Finance*, Norwell, MA,:Now Publishers Inc.

Tirole, Jean. 1988. *The Theory of Industrial Organization.* Cambridge, MA:The MIT Press.

Appendices

A Data

A.1 Data Source

Firm Fundamentals

I use CRSP/Compustat Merged (CCM) Database for firm fundamentals reported at the annual frequency. Individuals in this database is a subset of U.S. publicly listed firms. They can be identified in two ways: either by CRSP's permanent company and security identifiers (`permco` or `permno`), or by Compustat's permanent company identifier (`gvkey`). Variables to be used include R&D expense (`xrd`), sale (`sale`), number of employees (`emp`), book value of capital (`ppent`), value added output (`ouput`) and etc.

Industry-level Cost Shares and Prices Deflators

The industry-level cost shares and prices deflators are obtained from the NBER-CES Manufacturing Database, which can be merged to the CCM database mentioned above by the four-digit SIC codes. The NBER-CES database informs on industry-level payroll (`pay`), value added (`vadd`), shipment price deflator (`piship`), investment price deflator (`piinv`) and etc. These variables combined with those firm fundamentals are used to estimate revenue TFP, I shall elaborate how in later subsection.

Patent Value

To evaluate the outcome of firms' R&D activities, I employ the patent value dataset published by Kogan et al. (2017). They estimate the private value (or market value) and scientific value of each U.S. patent issued from 1926 to 2010 that can be matched to a publicly list firm. Their approach is to measure the patent value by the response of the stock market to the news of the issuance in a short (two days) time window. They then aggregate patent-level value to firm-year level, and the outcome can be matched to the CCM database through permanent company and security identifier.

The difference between the firm-year private patent value (`tsm`) and the firm-year scientific patent value (`tcw`) is that for the latter patents are weighted by the number of forward citations. In this paper I use the variable of private value following the argument that it is the private – instead of scientific – value, that the managers care about when it comes to R&D decisions.

The authors also provided scaled private value (`v7`) and scientific value (`v6`) of patents, which are firm-year patent value divided by book assets.

GDP Deflator

In Kogan et al. (2017), patent values are normalized to 1982 million dollars. Thus when I match them to R&D expenditure (RDE)[30], I would like to normalize the latter to the same year as well. For this purpose I use the annual GDP Implicit Price Deflator in United States, available at the website of Federal Reserve Economic Data (FRED).

A.2 Data Pre-processing

For the CRSP/Compustat merged data, I restrict the sample to U.S–based firms that provide final versions of statements. We omit regulated utilities (SIC codes 4900 to 5000) and financial firms (SIC codes 6000 to 7000), get rid of firm-year observations with values of acquisitions greater than 5% of assets, and keep only if the firm exists in the data for at least two years. I also drop observations with negative or missing book value of assets, book value of capital, number of employees, capital investment or revenue. Because Compustat records end-of-year captal values, we shift the reported book value forward one year.

For each industry definded by a four-digit SIC code and year in the NBER-CES database, I compute the following two variables: the labor share in value added (payroll cost divided by value added, with variable name `labshare`), the ratio of value added to gross output (`vaddfrac`). I then replace these two variables by their respective 10-year moving average, and generate the capital share (`capshare`) as the residual of the labor share, where I make the underlying assumption that the production function is homogeneous of degree one in labor and capital.

Finally, merge the above two datasets by industry and year indicator (`gvkey` and `year`, respectively), and then merge with it the Kogan et al. (2017) firm innovation value dataset by the permanent company and security identifier (`permno`) and `year`. This yields an unbalanced panel dataset, whose time spans annualy from 1970 to 2010, and covers 4,074 firms (identified by Compustat's permanent company identifier, `gvkey`) out of 135 four-digit SIC industries. There are 43,800 observations in total.

A.3 Construction of Key Variables

Revenue Total Factor Productivity (TFPR)

[30] Data on RDE is used for calibration in Section 5,

The revenue TFP in this paper is estimated using the cost-share based approach. Define variable `capital` to be the book value of capital (`ppent` in CCM) deflated by the investment deflator (`piinv` from the NBER-CES database). And define variable `output` as `sale` multiplied by the value added to gross output ratio (`vaddfrac`), and then deflated by the shipments deflator (`piship`). The level of TFPR is calculated as follows:

$$tfp_{i,j,t} = \exp\left[\log(output_{i,j,t}) - capshare_{j,t}\log(capital_{i,j,t}) - labshare_{j,t}\log(emp_{i,j,t})\right]. \tag{A.1}$$

The unit of the variable `tfp` generated as above is million dollars, the same as that of `output` and `capital`. In this way, `tfp` is the residual of revenue that is not explained by the factors capital and labor in a production function homogeneous of degree one, whose factor shares are invariant across firms in each industry.

R&D Expenditure (RDE) and Intensity (RDI)

The CCM database provides firm-year observations of R&D expenses (`xrd`), I scale it by firm size, approximated by sale, to get the RDI. When the level of RDE is need, the original variable `xrd` is normalized to 1982 million dollars using the GDP deflator, the outcome is named `rde`.

In Section 7 where R&D intensity is concerned, I take the natural log of RDI (`lnrdi`) as the explained variable instead of using the level of it, because the latter is highly right-skewed, with mean 448.56 and the maximum as high as 2568440. This may be the result of the fact some firms may have sales close to zero at times when their RDE is far from zero. For the same reason, in Section 2, the natural log of sale (`lnsale`) is used as the explained variable in the local projection models.

Control Variables

Throughout the empirical studies in this paper (Sections 2 and 7), the sets of firm- and industry-level control variables are consistent, and are elaborated as follows. The firm-level controls are:

1. Firm age (`age`): current year minus the year of first appearance of the firm.
2. One-year lag of revenue TFP (`tfp`).
3. One-year lag of capital (`capital`).
4. One-year lag of number of employees (`emp`).

5. One-year lag of market share (`mkshare`). The market share is the ratio of the firm-year sale to the sum of sales from all firms in its industry-year cell.

There are two industry-level control variables:

1. One-year lag of number of competitors (`compt`). The number of competitors is the count of firms in the industry-year cells.

2. One-year lag of the Herfindahl-Hirschman Index (HHI) (`hhi`). The description of HHI is in footnote 3 on page 4.

A.4 Descriptive Statistics

The standard descriptive statistics of variables mentioned in the previous subsection reported in Table 4 below. The statistics of RDI is from the sample after dropping outliers, whose reason was introduced previously in subsection A.3. The last two variables are calculated at the industry-year instead of firm-year level because of the way they are constructed.

Table 4: Descriptive Statistics of Key Variables

Variable name	Mean	Std. Dev.	Min	Max
tsm	612.38	4101.60	0.00	154092.1
v7	0.16	0.39	0.00	12.67
rde	40.84	228.51	0.00	7250.38
lnrdi	1.80	1.96	−5.64	14.76
lnsale	4.32	2.32	−6.91	12.96
age	12.07	10.73	1	60
tfp	25.46	124.59	0.00	5475.91
capital	384.92	2627.27	0.00	95284.93
emp	6.47	27.94	0.00	853
mkshare	0.11	0.21	0.00	1
compt	8.77	12.33	1	133
hhi	5188.16	2772.35	430.05	10000

B Proofs

B.1 Proof of Proposition 1

In stochastic Bellman equation (12), the composite arriving time $Z \in (0, \infty)$ follows exponential distribution:

$$\begin{aligned} F_Z(z) =& 1 - \Pr(Z > z) = 1 - \Pr(Z_A > z, Z_B > z) \\ =& 1 - \exp\left[-\left(\lambda(a_A^* + a_B^*) + h \cdot \mathbb{1}\{\Delta n_i \neq 0\}\right) z\right]. \end{aligned} \tag{B.1}$$

Let $Y = e^{-\rho Z} < 1$, it can be shown that Y follows a Beta distribution: $Y \sim \text{Beta}(\overline{\lambda}/\rho, 1)$, where $\overline{\lambda} := \lambda(a_A^* + a_B^*) + h \cdot \mathbb{1}\{\Delta n_i \neq 0\}$.

$$F_Y(y) = \Pr\left(Z \leq -\frac{\ln y}{\rho}\right) = \int_0^{-\frac{\ln y}{\rho}} \overline{\lambda} e^{-\overline{\lambda} z} dz = 1 - y^{\overline{\lambda}/\rho}; \tag{B.2}$$

$$f_Y(y) = -\frac{\overline{\lambda}}{\rho} y^{(\overline{\lambda}/\rho)-1} = \frac{1}{\mathrm{B}(\overline{\lambda}/\rho, 1)} y^{(\overline{\lambda}/\rho)-1}, \tag{B.3}$$

where $\mathrm{B}(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt$ is the Beta function.

Therefore, in (12),

$$\begin{aligned} \mathbb{E}_Z\left[\int_0^Z e^{-\rho t} dt\right] =& \frac{1}{\rho}\left[1 - \mathbb{E}_Z\left(e^{-\rho Z}\right)\right] = \frac{1}{\rho}\left[1 - \mathbb{E}_Y(Y)\right] \\ =& \frac{1}{\lambda(a_A^* + a_B^*) + h \cdot \mathbb{1}\{\Delta n_i \neq 0\} + \rho}; \end{aligned} \tag{B.4}$$

$$\begin{aligned} \mathbb{E}_Z\left[e^{-\rho Z} \mathbb{1}\{Z = Z_i\}\right] =& \mathbb{E}_Z\left[e^{-\rho Z}\right] \mathbb{E}_Z\left[\mathbb{1}\{Z = Z_i\}\right] = \mathbb{E}_Y(Y) \Pr(Z_i < Z_{-i}) \\ =& \frac{\lambda(a_A^* + a_B^*) + h \cdot \mathbb{1}\{\Delta n_i \neq 0\}}{\lambda(a_A^* + a_B^*) + h \cdot \mathbb{1}\{\Delta n_i \neq 0\} + \rho} \cdot \frac{\lambda a_i^* + h \cdot \mathbb{1}\{\Delta n_i < 0\}}{\lambda(a_A^* + a_B^*) + h \cdot \mathbb{1}\{\Delta n_i \neq 0\}} \\ =& \frac{\lambda a_i^* + h \cdot \mathbb{1}\{\Delta n_i < 0\}}{\lambda(a_A^* + a_B^*) + h \cdot \mathbb{1}\{\Delta n_i \neq 0\} + \rho}. \end{aligned} \tag{B.5}$$

Substitute equations (B.4) and (B.5) into Bellman equation (12), I derive the optimization system as in Proposition 1. The boundary conditions are from the fact that a firm has no incentive to do R&D at the maximal technology gap $\overline{m}$.

The existence of a Markov Perfect Equilbrium is guaranteed by the application of Kakutani's Fixed-Point Theorem. An easy representation is Theorem 5.11.15 in Corbae, Stinchcombe and Zeman (2009).

□

Proof of Proposition 3

By the symmetry of the game, firms A and B have the same value function and policy function[31]. Therefore, $\{\Delta n_A(t)\}$ and $\{\Delta n_B(t)\}$ have the same limiting distribution μ. By definition of the technology gap, $\Delta n_A(t) = -\Delta n_B(t)$, thus for any $m \in \mathcal{M}$, $\mu_i = \lim_{t\to\infty} \Pr(\Delta n_A(t) = m) = \lim_{t\to\infty} \Pr(\Delta n_B(t) = -m) = \lim_{t\to\infty} \Pr(\Delta n_A(t) = -m) = \mu_{2\overline{m}+2-i}$. Since $\mathcal{M}_i = -\mathcal{M}_{2\overline{m}+2-i}$, $\lim_{t\to\infty} \mathbb{E}\left[\Delta n_A(t)\right] = \sum_{i=1}^{2\overline{m}+1} \mu_i \mathcal{M}_i = 0$.

□

B.2 Proof of Lemma 2

Suppose the opposite, i.e. there exists $m_0 \in \mathcal{M}$, such that $v_A(m_0) \le v_B(m_0)$. Obviously, $a_A^*(m) \ne a_B^*(m)$ for some m, otherwise $v_A(m) > v_B(m)$ for all m because firm A has lower marginal cost of R&D.

By Definition 2, denote the strategy profile by $\left\{a_A^*(\Delta n_A), a_B^*(\Delta n_B)\right\}$, that $v_A(m_0) \le v_B(m_0)$ means

$$\begin{aligned}
& v_A\left(m_0 \middle| \left\{a_A^*, a_B^*\right\}\right) \\
=& \mathbb{E}_{\Delta N_A}\left\{\int_0^\infty e^{-\rho t}\left[\pi_A\left(\Delta N_A(t)\right) - \psi_A\left(a_A^*\left(\Delta N_A(t)\right)\right)\right] dt \middle| \Delta N_A(0) = m, \left\{a_A^*, a_B^*\right\}\right\} \\
\le& \mathbb{E}_{\Delta N_B}\left\{\int_0^\infty e^{-\rho t}\left[\pi_B\left(\Delta N_B(t)\right) - \psi_B\left(a_B^*\left(\Delta N_B(t)\right)\right)\right] dt \middle| \Delta N_B(0) = m, \left\{a_A^*, a_B^*\right\}\right\} \\
=& v_B\left(m_0 \middle| \left\{a_A^*, a_B^*\right\}\right).
\end{aligned} \tag{B.6}$$

Now let firm A play B's strategy, and denote firm B's best response by a_B^{br}. Compare firm A's original strategy a_A^* and a_B^{br}, there are three cases:

Case 1: $a_B^{\text{br}}(m) = a_A^*(m)$ for all m. In this case, it is easy to see from inequality (B.6) that $v_A\left(m_0 \middle| \left\{a_B^*, a_B^{\text{br}}\right\}\right) > v_A\left(m_0 \middle| \left\{a_A^*, a_B^*\right\}\right)$ because the positions of the two firms are mirrored, and firm A has strictly lower marginal cost than firm B.

Case 2: $a_B^{\text{br}}(m) > a_A^*(m)$ for some m. This contradicts that equilibrium strategy a_A^* is the

[31] Otherwise it's easy to show there is contradiction by switching firm labels

best response to a_B^*. This is because firm B faces the same problem when A plays a_B^* as A faces in the original MPE. However, the marginal cost of R&D is strictly higher for firm B, thus $a_B^{\mathrm{br}}(m) > a_A^*(m)$ is impossible.

Case 3: $a_B^{\mathrm{br}}(m) \le a_A^*(m)$ for all m, and this inequality holds strictly for some m. This implies that for any $t > 0$ and any $k \in \{1, 2, \cdots, \overline{m}\}$,

$$\begin{aligned}
\Pr\Big(\Delta N_A(t) = k|\Delta N_A(0) = m; \left\{a_B^*, a_B^{\mathrm{br}}\right\}\Big) \ge& \Pr\big(\Delta N_A(t) = k|\Delta N_A(0) = m; \left\{a_B^*, a_A^*\right\}\big) \\
=& \Pr\big(\Delta N_B(t) = k|\Delta N_B(0) = m; \left\{a_A^*, a_B^*\right\}\big) \qquad \text{(B.7)}
\end{aligned}$$

The last equality is derived from the symmetry of this dynamical system: once the initial states and R&D are flipped and strategies of players A and B swapped, the random variable $\Delta N_f(t)$ is governed by the same stochastic process which $\Delta N_{-f}(t)$ initially follows. For the same reason,

$$\Pr\Big(\Delta N_A(t) = k|\Delta N_A(0) = m; \left\{a_B^*, a_B^{\mathrm{br}}\right\}\Big) \le \Pr\big(\Delta N_B(t) = k|\Delta N_B(0) = m; \left\{a_A^*, a_B^*\right\}\big) \qquad \text{(B.8)}$$

for any $t > 0$ and any $k \in \{-\overline{m}, -\overline{m}+1, \cdots, 0\}$.

In either case 1 or case 3, $\mathcal{J}_A\left(m_0 \middle| \left\{a_B^*, a_B^{\mathrm{br}}\right\}\right) > v_A\left(m_0 \middle| \left\{a_A^*, a_B^*\right\}\right)$, i.e. firm A can achieve a strictly higher performance measure by deviating from a_A^* to a_B^*, however it chooses not to, which contradicts the rational agent assumption. Therefore, the major premise, that the existence of $m_0 \in \mathcal{M}$ such that $v_A(m_0) \le v_B(m_0)$, is false.

□

B.3 Proof of Lemma 3

Firstly, if $a_A^*(-1) \le a_B^*(-1)$, it must be that $a_A^*(0) < a_B^*(0)$. To see this, notice that from Corollary 1,

$$\begin{aligned}
v_f(1) - v_f(0) =& \frac{1}{\lambda a_{-f}^*(-1) + \rho + h}\left[\pi_f(1) + \left(\lambda a_{-f}^*(-1) + h\right) v_f(0) - \left(\lambda a_{-f}^*(-1) + h\right) v_f(0) - \rho v_f(0)\right] \\
=& \frac{1}{\lambda a_{-f}^*(-1) + \rho + h}\left[\pi_f(1) - \rho v_f(0)\right]. \qquad \text{(B.9)}
\end{aligned}$$

Suppose $a_A^*(-1) \leq a_B^*(-1)$, from Corollary 1, Lemma 2 and equation (B.9),

$$\begin{aligned}
\frac{d\psi_A\left(a_A^*(0)\right)}{da_A} =& \lambda[v_A(1) - v_A(0)] = \frac{\lambda}{\lambda a_B^*(-1) + \rho + h}\left[\pi_A(1) - \rho v_A(0)\right] \\
<& \frac{\lambda}{\lambda a_A^*(-1) + \rho + h}\left[\pi_B(1) - \rho v_B(0)\right] = \lambda[v_B(1) - v_B(0)] \\
=& \frac{d\psi_B\left(a_B^*(0)\right)}{da_B},
\end{aligned} \tag{B.10}$$

Where the fact $\pi_A(m) = \pi_B(m)$ for any $m \in \mathcal{M}$ comes from Assumption 1. Since $\frac{d\psi_A(a)}{da_A} < \frac{d\psi_B(a)}{da_B}$, inequality (B.10) implies $a_A^*(0) < a_B^*(0)$. Now that $a_A^*(-1) \leq a_B^*(-1)$ and $a_A^*(0) < a_B^*(0)$, for either $m = -1$ or $m = 0$, $\frac{d\psi_A\left(a_A^*(m)\right)}{da_A} < \frac{d\psi_B\left(a_B^*(m)\right)}{da_B}$, thus

$$\begin{aligned}
&[v_A(1) - v_B(1)] - [v_A(-1) - v_B(-1)] = [v_A(1) - v_A(-1)] - [v_B(1) - v_B(-1)] \\
=& [v_A(1) - v_A(0) + v_A(0) - v_A(-1)] - [v_B(1) - v_B(0) + v_B(0) - v_B(-1)] \\
=& \left(\frac{d\psi_A\left(a_A^*(0)\right)}{da_A} + \frac{d\psi_A\left(a_A^*(-1)\right)}{da_A}\right) - \left(\frac{d\psi_B\left(a_B^*(0)\right)}{da_B} + \frac{d\psi_B\left(a_B^*(-1)\right)}{da_B}\right) \\
<& 0.
\end{aligned} \tag{B.11}$$

On the other hand, from Corollary 1, in the case $\overline{m} = 1$, I have

$$\begin{aligned}
v_A(1) - v_B(1) =& \frac{1}{\lambda a_B^*(-1) + \rho + h}\left[\pi(1) + \left(\lambda a_B^*(-1) + h\right) v_A(0)\right] \\
&- \frac{1}{\lambda a_A^*(-1) + \rho + h}\left[\pi(1) + \left(\lambda a_A^*(-1) + h\right) v_B(0)\right];
\end{aligned} \tag{B.12}$$

$$\begin{aligned}
v_A(-1) - v_B(-1) =& \frac{1}{\lambda a_A^*(-1) + \rho + h}\left[\pi(-1) - \psi_A\left(a_A^*(-1)\right) + \left(\lambda a_A^*(-1) + h\right) v_A(0)\right] \\
&- \frac{1}{\lambda a_B^*(-1) + \rho + h}\left[\pi(-1) - \psi_B\left(a_B^*(-1)\right) + \left(\lambda a_B^*(-1) + h\right) v_B(0)\right].
\end{aligned} \tag{B.13}$$

Equations (B.12) and (B.13) implies

$$\begin{aligned}
&[v_A(1) - v_B(1)] - [v_A(-1) - v_B(-1)] \\
=& \frac{1}{\lambda a_B^*(-1) + \rho + h}\left[\pi(1) + \pi(-1) + \psi_B\left(a_B^*(-1)\right) + \left(\lambda a_B^*(-1) + h\right)\left(v_A(0) + v_B(0)\right)\right]
\end{aligned}$$

$$-\frac{1}{\lambda a_A^*(-1)+\rho+h}\left[\pi(1)+\pi(-1)+\psi_A\left(a_A^*(-1)\right)+\left(\lambda a_A^*(-1)+h\right)\left(v_A(0)+v_B(0)\right)\right] \tag{B.14}$$

From $\psi_f(0)=0$ in Assumption 2 and that $\frac{d\psi_A(a)}{da_A}<\frac{d\psi_B(a)}{da_B}$ for all $a\geq 0$, it's easy to show that $\psi_A(a)<\psi_B(a)$ for all $a>0$. Therefore, in equation (B.14), $[v_A(1)-v_B(1)]-[v_A(-1)-v_B(-1)]>0$ when arrival rate multiplier $\lambda>0$ is small enough, which contradicts inequality (B.11). Therefore, $a_A^*(-1)\leq a_B^*(-1)$ is negated.

Now that $a_A^*(-1)>a_B^*(-1)$, similar to inequality (B.10), it's easy to see that $\frac{d\psi_A\left(a_A^*(0)\right)}{da_A}>\frac{d\psi_B\left(a_B^*(0)\right)}{da_B}$ when the discount factor $\rho>0$ is small enough, which implies $a_A^*(0)>a_B^*(0)$. □

B.4 Proof of Lemma 4

Define the lower contour set $C_f\left(v|d_f,d_{-f}\right)$ the set of all strategy profiles $\{a_A(t),a_B(t)\}_{t=0}^{\infty}$ by which the performance measure of firm f is no greater than v:

$$C_f\left(v|d_f,d_{-f}\right):=\left\{\{a_A(t),a_B(t)\}_{t=0}^{\infty}:\mathcal{J}_f\left(d_f,d_{-f}\middle|\left\{a_f(t),a_{-f}(t)\right\}_{t=0}^{\infty}\right)\right\} \tag{B.15}$$

It is important that both contour sets C_A and C_B, whenever non-empty, have their elements in the same order of the strategies of firms A and B, otherwise any operation of these two sets is meaningless.

Fix d_A and d_B, suppose for each $f\in\{A,B\}$, value function v_f is defined at (d_f,d_{-f}) and $(d_f+1,d_{-f}+1)$. By profit function in Assumption 3, for any arbitrary strategy profile $\{\widetilde{a}_A(t),\widetilde{a}_B(t)\}_{t=0}^{\infty}$, $\mathcal{J}_f\left(d_f,d_{-f}\middle|\{\widetilde{a}_A(t),\widetilde{a}_B(t)\}_{t=0}^{\infty}\right)>\mathcal{J}_f\left(d_f+1,d_{-f}+1\middle|\{\widetilde{a}_A(t),\widetilde{a}_B(t)\}_{t=0}^{\infty}\right)$. Therefore, for $f\in\{A,B\}$,

$$C_f\left(v_f(d_f+1,d_{-f}+1)|d_f,d_{-f}\right)\subsetneq C_f\left(v_f(d_f+1,d_{-f}+1)|d_f+1,d_{-f}+1\right) \tag{B.16}$$

and

$$\partial C_f\left(v_f(d_f+1,d_{-f}+1)|d_f,d_{-f}\right)\cap\partial C_f\left(v_f(d_f+1,d_{-f}+1)|d_f+1,d_{-f}+1\right)=\emptyset, \tag{B.17}$$

where ∂C means the boundary of set C.

By definition, the equilibrium strategy profile $\left\{a_A^*(t), a_B^*(t)|d_A+1, d_B+1\right\}_{t=0}^{\infty}$ belongs to $C_f\left(v_f(d_f+1, d_{-f}+1)|d_f+1, d_{-f}+1\right)$, but not to $C_f\left(v_f(d_f+1, d_{-f}+1)|d_f, d_{-f}\right)$, which means

$$C_f\left(v_f(d_f+1, d_{-f}+1)|d_f+1, d_{-f}+1\right) \setminus C_f\left(v_f(d_f+1, d_{-f}+1)|d_f, d_{-f}\right) \neq \emptyset. \tag{B.18}$$

Now discuss whether the equilibrium strategy profile at state $\left\{a_A^*(t), a_B^*(t)|d_A, d_B\right\}_{t=0}^{\infty}$ belongs to $C_f\left(v_f(d_f+1, d_{-f}+1)|d_f, d_{-f}\right)$. If it doesn't, by the definition of lower contour set, it implies $v_f(d_f, d_{-f}) > v_f(d_f+1, d_{-f}+1)$ and the proof thus finishes. If it does, by (B.17) and (B.18), firm f can deviate to any strategy in the non-empty difference set in (B.18), where for any strategy $-f$ can choose, firm f will end up with a strictly higher performance measure than $v_f(d_f, d_{-f})$, which contradicts that $\left\{a_A^*(t), a_B^*(t)|d_A, d_B\right\}_{t=0}^{\infty}$ consists of the optimal strategies for both firms.

□

B.5 Proof of Proposition 7

Suppose the state prior to the shock is (d_A, d_B). For either firm $f \in \{A, B\}$, by Lemma 4, both $v_f(d_f+k-1, d_{-f}+k)$ and $v_f(d_f+k, d_{-f}+k)$ are monotonically decreasing in k and bounded below by zero. Therefore, as $\delta \to 0^+$, $\lim\limits_{k\to\infty} v_f(d_f+k-1, d_{-f}+k)$ and $\lim\limits_{k\to\infty} v_f(d_f+k, d_{-f}+k)$ exists and are equal. This implies $\lim\limits_{k\to\infty}\left[v_f(d_f+k-1, d_{-f}+k) - v_f(d_f+k, d_{-f}+k)\right] = 0$, or that for any $\varepsilon > 0$, there exists $\widetilde{D}(\varepsilon)$ such that for both $f = A$ and $f = B$, $v_f(d_f+k-1, d_{-f}+k) - v_f(d_f+k, d_{-f}+k) < \varepsilon$ if $k \geq \widehat{D}(\varepsilon)$. By Proposition 5 and the strict increasingness of R&D cost function ψ_f (Assumption 2), this implies that for either firm, the equilibrium R&D effort satisfies $a_f^*(d_f, d_{-f}) > a_f^*(d_f+k, d_{-f}+k)$ for k large enough.

□

B.6 Proof of Proposition 8

Without loss of generality, suppose $\lim\limits_{s\to t^-} d_A(s) = 0$. Firstly discuss the case where $\lim\limits_{s\to t^-} d_B(s) < \overline{m}$. By updating rule (28) and first-order condition (33), $\lim\limits_{s\to t^-} a_A^*(s) = \psi_A^{-1}\left(\lambda\left[v_A(0, d_B+1) - v_A(0, d_B)\right]\right)$, and $a_A^*(t) = \psi_A^{-1}\left(\lambda\left[v_A(0, d_B+1) - v_A(1, d_B+1)\right]\right)$. By Lemma 4, $v_A(0, d_B) > v_A(1, d_B+1)$. By the strict Monotonicity of ψ_A as in Assumption 2, $a_A^*(t) > \lim\limits_{s\to t^-} a_A^*(s)$.

If $\lim\limits_{s\to t^-} d_B(s) = \overline{m}$. By boundary condtion (34), $\lim\limits_{s\to t^-} a_A^*(s) = 0$. Again by first-order condtion (33), $a_A^*(t) > 0$. Thus $a_A^*(t) > \lim_{s \to t^-} a_A^*(s)$ holds.

$\square$

C Model Parameterization

C.1 Calibration

To determine parameter κ_f, $f \in \{A, B\}$ as in the R&D cost function (36), I use equation (37) derived from the first-order condition with respect to the R&D effort. The R&D cost $\psi_f(a)$ is approximated empirically by R&D intensity from the data, and the change in firm value induced by the technological progress is approximated by the firm-year innovation value scaled by sale. The latter is justified by the way patent values are estimated – the change in firm value on the stock market explained by the news of new patent. It is scaled by firm-year sale because it is to which R&D expenditure is scaled to get the RDI.

In this way, denote firm-year innovation value, RDE and sale by $tsm_{i,j,t}$, $rde_{i,j,t}$ and $sale_{i,j,t}$, respectively, and κ_f is calibrated by

$$\kappa_A = \frac{1}{JT} \sum_{j,t} \frac{1}{N_{jt}} \sum_i \frac{\left(tsm_{i,j,t}/sale_{i,j,t}\right)^2}{rde_{i,j,t}/sale_{i,j,t}} \times \mathbb{1}\{c^8_{i,j,t} = 1\}; \tag{C.1}$$

$$\kappa_B = \frac{1}{JT} \sum_{j,t} \frac{1}{N_{jt}} \sum_i \frac{\left(tsm_{i,j,t}/sale_{i,j,t}\right)^2}{rde_{i,j,t}/sale_{i,j,t}} \times \mathbb{1}\{c^3_{i,j,t} = 1\}, \tag{C.2}$$

where J are T are the numbers of industries and years in the sample, and N_{jt} the number of firms in industry-year cell (j, t). Indicators $\mathbb{1}\{c^8_{i,j',t} = 1\}$ and $\mathbb{1}\{c^3_{i,j',t} = 1\}$ are defined in (46), and is used to distinguish low- and high-cost firms in data. Here it means that I'm treating the firms with log-RDI between the second and third deciles in its industry-year cell as representative for the low-cost firm in my model; and those between the seventh and eighth deciles for the high-cost firm. In the model, the low-cost firm, in expectation, exerts more R&D effort than the high-cost firm under the stationary distribution.

The sample mean is firstly taken within each industry-year cell, then averaged again across all such cells. This is to be consistent with the approach for the GMM estimation which will be introduced as follows.

C.2 GMM Estimation

With any arbitrary set of model parameters θ, the numerical solution of the policy functions $\{a^*_A(\Delta n_A|\theta), a^*_B(\Delta n_B|\theta)\}_{\Delta n_A=-\Delta n_B\in\mathcal{M}}$ is conditional on θ. By Proposition 2, the stationary distribution $\mu(m|\theta)$, $m \in \mathcal{M}$ is also a function on parameter set θ. Letting $\theta = (\overline{\pi}, h)$, the two parameters remains undetermined, I construct selected moments using the stationary distribution, and calculate their weighted distance to their counterparts in the data. This distance is then minimized over the parameter space to pin down the value of θ. The following is a step-by-step summary of my approach regarding the GMM estimation.

1. There needs to be two targeted moments to estimate the two parameters: total profit $\overline{\pi} > 0$ and imitation rate $h > 0$. One of the chosen moment is the expected ratio of the low-cost firm's R&D effort to that of the high-cost firm's:

$$\phi_1 = \mathbb{E}\left[\frac{a^*_A(\Delta n_A)}{a^*_B(\Delta n_B)}\right]. \tag{C.3}$$

And the other one is the expected ratio of the low-cost firm's innovation value to that of the low-cost firm's:

$$\phi_2 = \mathbb{E}\left[\frac{(v_A(\Delta n_A) - v_A(\Delta n_A - 1)) \times \mathbb{1}\{\Delta n_A > -\overline{m}\}}{(v_B(\Delta n_B) - v_B(\Delta n_B - 1)) \times \mathbb{1}\{\Delta n_B > -\overline{m}\}}\right]. \tag{C.4}$$

2. With a fixed set of parameters θ, I solve the baseline model numerically, and obtain the stationary distribution $\mu(\theta)$ using the transition rate matrix Q and jump matrix P as in (21). Then I compute the moments using $\mu(m|\theta)$ as follows:

$$\overline{\phi}_1(\theta) = \sum_{m=-\overline{m}}^{\overline{m}} \mu(m|\theta)\frac{a^*_A(m|\theta)}{a^*_B(-m|\theta)}; \tag{C.5}$$

$$\overline{\phi}_2(\theta) = \sum_{m=-\overline{m}+1}^{\overline{m}} \mu(m|\theta)\frac{v_A(m|\theta) - v_A(m-1|\theta)}{v_B(-m+1|\theta) - v_B(-m|\theta)}. \tag{C.6}$$

3. Now calculate the corresponding moments from data, denoted $\Phi = (\Phi_1, \Phi_2)$. While doing so I use R&D intensity to approximate the R&D cost ψ, and innovation value scaled by book value of asset, v7[32], as the proxy of the change in firm value induced by

[32] The variable name v7 is from the dataset of Kogan et al. (2017).

innovation.

$$\Phi_1 = \frac{1}{JT} \sum_{j,t} \frac{\sum_i rdi_{i,j,t} \times \mathbb{1}\{c^8_{i,j,t} = 1\}}{\sum_i rdi_{i,j,t} \times \mathbb{1}\{c^3_{i,j,t} = 1\}}; \tag{C.7}$$

$$\Phi_2 = \frac{1}{JT} \sum_{j,t} \frac{\sum_i v7_{i,j,t} \times \mathbb{1}\{c^8_{i,j,t} = 1\}}{\sum_i v7_{i,j,t} \times \mathbb{1}\{c^3_{i,j,t} = 1\}};. \tag{C.8}$$

Take the calculation of Φ_1 for example. Firstly for each industry-year cell (j, t), take the sum of RDI over the eighth and third decile groups, respectively. Secondly, compute the ratio of these two sums, which is equivalent to computing the ratio of the means of the RDI in the eighth group to that in the third group, because the number of observation in each decile group is by definition equal. Finally, compute the mean of this ratio across all industry-year cells. This is the sample counterpart of moment (C.3), the expected ratio of R&D costs. The similar can be said about (C.8).

4. Let $\widehat{W}$ be the covariance matrix of the two ratios $\frac{\sum_i rdi_{i,j,t} \times \mathbb{1}\{c^8_{i,j,t} = 1\}}{\sum_i rdi_{i,j,t} \times \mathbb{1}\{c^3_{i,j,t} = 1\}}$ and $\frac{\sum_i v7_{i,j,t} \times \mathbb{1}\{c^8_{i,j,t} = 1\}}{\sum_i v7_{i,j,t} \times \mathbb{1}\{c^3_{i,j,t} = 1\}}$, invert it to get $\widehat{W}^{-1}$. The distance between $\overline{\phi}(\theta) = \left(\overline{\phi}_1(\theta), \overline{\phi}_2(\theta)\right)$ and Φ is a function of the parameters to be estimated:

$$Q(\theta) = \left(\overline{\phi}(\theta) - \Phi\right)' \widehat{W}^{-1} \left(\overline{\phi}(\theta) - \Phi\right) \tag{C.9}$$

which is to be minimized over the parameter space $(\overline{\pi}, h)$, and the minimizer is the estimated values. Since the dimension is low, I use the grid search method to find the local minimizer reported in the paper.