A Cost-benefit Analysis of R&D and Patents: Firm-level Evidence from China*

Zhiyuan Chen† Jie Zhang‡ Yuan Zi§

December 19, 2020

Abstract

We extend the empirical framework by Peters et al. (2017) to include both R&D and patents in the productivity evolution. We provide a decomposition of the benefits of R&D into the patent and non-patent components, and a novel measure of the patent value conditioning on the firm's R&D investment. Using a sample of Chinese high-tech manufacturing firms, we find that (1) 47.8% to 67% of the benefits of R&D investment comes from non-patent R&D activities; (2) On average an invention (a utility model) patent causes around 0.76 (0.66) percent increase in the firm value; (3) The start-up costs of R&D are around ten times as large as the maintenance R&D costs. The counterfactual analysis shows that the lump-sum subsidy is

*We thank the insightful views from the editors and two anonymous referees. Zhiyuan Chen is grateful to Mark Roberts and Jonathan Eaton for their intellectual support when this project was initiated as a chapter of his PhD thesis. This paper also benefits from the helpful comments by Paul Grieco and Daniel Grodzicki at its very early stage. Zhiyuan Chen acknowledges the Starting Research Fund from the Renmin University of China. Jie Zhang acknowledges the support from the National Natural Science Foundation of China (No. 71973139). Yuan Zi acknowledges the funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement 715147). All remaining errors are ours.

†School of Business, Renmin University of China. Email: chenzhiyuan@rmbs.ruc.edu.cn.

‡*Corresponding author.* School of Economic Reform and Development, Renmin University of China. Email: zhanjie0402@ruc.edu.cn.

§Department of Economics, University of Oslo. Email:yuanzi.economics@gmail.com

more effective than the proportional subsidy in increasing the expected firm value and innovation probability. R&D continuers respond more actively than the R&D starters to the R&D subsidy.

Keywords: dynamic R&D investment; patent; productivity; R&D subsidy

1 Introduction

Innovation is a key engine of productivity growth. Quantifying the costs and benefits of innovation activities is essential to understanding the firm's incentives to innovate. Considering the innovation as a process of producing knowledge, R&D is the input while patents are part of the output. This is formally analyzed in the previous econometric models linking innovation outcome to the R&D investment and allowing the productivity to be only affected by the innovation output (Crépon et al., 1998; Mairesse et al., 2005; Raymond et al., 2015). In these models, all of the returns to R&D is captured by the innovation output, which is usually measured by patents.

Patents, however, are not a perfect measure for the outcome of innovation. In many developing countries, the protection of intellectual property rights is often limited, making firms in these countries less motivated to patent their invention. Even when the patent law is perfectly implemented, not all inventions are patentable. In reality, a large part of the innovation output is the accumulation of tacit knowledge that is difficult to measure and/or patent.[1] Therefore it is unlikely for firms to apply for patents to codify all of their inventions. Regarding this situation, treating the patent as the sole innovation output may lead to biased estimates for the returns to either R&D or patents. By construction, including R&D or the patent alone in the evolution process of productivity prevents us from analyzing the structure of R&D benefits in terms of patent and non-patent channels.

To address these limitations, in this paper we treat the patent as a part of the innovation outcome and allow for other unobserved non-patent knowledge to affect the firm's future productivity. In particular, we treat R&D as the fundamental source of productiv-

[1] As illustrated by Hall et al. (2010), R&D generates an intangible asset, the firm's knowledge base, which contributes to profits in future years. The outcome of R&D is usually embedded in the human capital of the firm's employees, and is often hard to be codified.

ity growth, but the marginal effect of R&D investment is affected by patent outcomes. Our specification enriches the existing literature by incorporating both the impact of innovation input (R&D investment) and imperfectly observed innovation output (patents) in the process of productivity evolution. This seemingly small change brings us economic insights in understanding the internal structure of returns to R&D investment through the patent and non-patent channels.

Our model is based on Peters et al. (2017) (PRVF hereafter) which provides a structural analysis of the costs and benefits of R&D investment in a unified empirical framework containing both the innovation inputs (R&D) and outputs (observed process and product inventions). In the PRVF model, R&D affects the probability of realizing innovation outcomes, and the realized innovation outcome leads to productivity change. One limitation of their framework is the researchers have to obtain accurate measures of the innovation outcomes, which are usually not available in the commonly used datasets.[2] To make things more difficult, in the survey firms are likely to hide their true information on inventions from the interviewer to prevent imitation from their competitors.

We consider a more realistic setting where the R&D investment and some imperfect measures of innovation output jointly affect the productivity evolution. We follow the main setup of the PRVF model, but introduce two changes: (1) instead of relying on accurately measured innovation outcomes, we use patent counts as an indicator for innovation outcomes; (2) we let productivity be dependent on R&D even after conditioning on patents. In our framework, R&D is the fundamental source of productivity change; its impact is determined by the realization of innovation outcomes. In PRVF, they estimate two primitives regarding the innovation process. The first is the probability of realizing innovation outcome conditional on the state of R&D investment, and the

[2] PRVF use the Mannheim Innovation Panel survey which contains the information on the introduction of a new product or a new production process by each firm.

second is the distribution of future productivity conditioning on the realized innovation outcome. In our case, we replace the innovation outcome with the firm's patent applications and estimate probability distribution similar to PRVF. Slightly differently, we estimate the distribution of future productivity conditional on both R&D and patents.

Because the productivity gains depend on both R&D and realized patent applications, an advantage of our framework is the decomposition of returns to R&D investment into components related to patents (new codified knowledge) and non-patents (new tacit knowledge) innovations. Another contribution of our empirical framework is a novel estimator of the private patent value inferred from the increase in the firm value. Different from Pakes (1986) where the value of holding a patent is determined by solving an optimal stopping problem, we measure the realized value of a patent by treating patent applications as a random outcome of R&D investment and conditioning on the R&D investment. Our approach can be applied to any data sets with firm-level information on R&D, patents, and production.

We estimate the model using a sample of Chinese high-tech manufacturing firms. Our main findings from the structural estimates are as follows. First, on average R&D investment increases the expected firm value by around 0.223% (or around 0.109 million USD) in the firm value. This value is much lower than that obtained by PRVF who found that R&D investment increases firm value by 6.70 percent (around 4.41 million Euros) for the median high-tech firm in a sample of German firms. This shows that Chinese firms obtain much smaller returns to R&D than German firms. Second, a decomposition of the return to R&D shows that non-patent innovation accounts for a substantial part of the total returns to R&D. This implies that a large part of the R&D benefits come from the accumulation of tacit knowledge. Third, conditioning on R&D investment, the average value of an invention (a utility model) patent is around .374 (.326) million USD.

Lastly, the average start-up cost of R&D is over ten times as large as the average maintenance cost. This suggests that R&D investment has high adjustment costs: starting an innovation project is much more costly than maintaining an ongoing research project. Lastly, R&D costs vary substantially across different industries, with the electronics industry being the highest.

Using the estimated model, we perform a series of counterfactual exercises to evaluate the effectiveness of different types of R&D subsidy policies. In particular, we compare the effect per dollar lump-sum subsidy and proportional subsidy.[3] To make different subsidy programs comparable, we restrict that their expected payments are the same before the implementation. The counterfactual exercises yield some interesting results. First, for both types of subsidy programs, a reduction in maintenance costs causes a greater increase in the firm value and innovation probability. Second, in terms of the effect of one-dollar spending, the lump-sum subsidy is more efficient than the proportional subsidy.

We explore the robustness of our counterfactual results by providing both theoretical justifications and empirical analyses on the effects of subsidy programs. First, we prove theoretically that when R&D costs follow an Exponential distribution, the lump-sum subsidy always performs better than the proportional subsidy in increasing the firm value at any subsidy rate. Second, we demonstrate that the lump-sum subsidy is more effective than proportional subsidy in increasing the firm value when: (1) the probability density function of R&D costs is a decreasing function of its argument; and (2) the right tail of the probability density function is sufficiently thin. Third, guided by the theoretical justification, we choose the Weibull distribution as a generalization of the

[3]The lump-sum subsidy is a government transfer that reduces R&D costs paid by the firm, while the proportional subsidy is a proportional reduction in R&D costs at a fixed rate. If the amount of R&D costs is c, the lump-sum subsidy reduces R&D costs to $c - s_l$, and the proportional subsidy changes the R&D costs to be $(1 - s_p)c$.

Exponential distribution. We re-estimate our model for a set of different values for the parameter governing the shape of the distribution. The counterfactual results confirm that our benchmark results are robust to certain alternative distributions for the R&D costs.

In another extension, we adapt our empirical framework to include firm ownership as a determinant of R&D costs (and benefits). We find no significant difference between SOEs and non-SOEs in terms of the start-up R&D costs, but SOEs tend to have significantly higher maintenance R&D costs in some industries.

This study is closely related to the literature quantifying returns to R&D. The knowledge capital model of Griliches (1979) has been a cornerstone of this literature. In their framework, the investment in innovation by the firm creates knowledge stock, which is similar to physical capital in the way that they enter into the production function. The most important extension related to this paper is the econometric framework proposed by Crépon, Duguet, and Mairesse (1998) (CDM hereafter) which estimates a reduced-form model incorporates R&D, patents, and productivity. Recently, Raymond, Mairesse, Mohnen, and Palm (2015) extends this framework to a dynamic setting. However, the knowledge-stock approach faces the problem of estimating the firm's knowledge stock. It also rules out the high degree of intrinsic uncertainty facing innovation investment.

Our work is also related to the endogenous productivity approach in the literature on R&D investment. In this strand of literature, the evolution of productivity is a controlled Markov process with the current R&D investment increases the future productivity (Hall and Hayashi, 1989; Klette, 1996; Aw et al., 2011; Doraszelski and Jaumandreu, 2013; Peters et al., 2017).

We contribute to these two strands of literature in two main aspects. First, we enrich the literature on quantifying returns to R&D by providing a decomposition of the benefits of R&D into the patent and non-patent channels. The empirical finding suggests that non-patent R&D investment plays a major role in returns to R&D. Second, we provide a new method of estimating the private value of patents. The empirical analysis in this paper also provides a first structural analysis of the cost-benefits structure of R&D and patents in China, thus contributing to the literature on the characteristics of innovation by Chinese firms.[4] Our estimation results indicate that non-patent innovation accounts for most of the returns to R&D. We point out that the relatively small benefits and high start-up R&D costs jointly contribute to the firm's low participation in R&D investment in the sample period. Lastly, we provide a theoretical analysis of the relative effectiveness of two popular types of subsidies and confirm the superiority of lump-sum subsidies in scenarios of general distributions for the R&D costs.

The rest of this paper is organized as follows. In Section 2, we outline the R&D model, illustrate the method of decomposing the R&D benefits as well as estimating the patent value. Data are introduced in Section 3. We display the empirical results in Section 4. Section 5 presents the counterfactual analyses. In Section 6, we discuss two main extensions of the model. Section 7 concludes the paper.

2 The Empirical Framework

In this section, we first briefly describe a standard dynamic model of R&D investment and patent outcomes. The basic structure of the model is identical to PRVF, the only difference is that we allow both R&D and patents play a role in shifting the distribution of future productivity. Based on the model, we propose a new method of decomposing the returns to R&D to patent and non-patent channels, as well as estimating the patent value.

[4] See, for example, Hu and Jefferson (2009), Hu et al. (2017), and Chen et al. (2017).

2.1 Model

In PRVF, a firm's R&D decision changes the probability of realizing product or process innovations, which affect the firm's future productivity and expected profits. In their specification, only the innovation outcome (process or product innovation) generated by R&D activities has an impact on the firm's future productivity and hence future profits. In contrast, we allow both R&D investment and patents enter into the productivity evolution equation while the creation of new ideas represented by observed patent applications affects the marginal effect of R&D on future productivity.

The model comprises four parts. The first part is the production function of patents that links the firm's distribution of patents to their R&D investment. The second component is the cost function of investment in R&D, which is influenced by the firm's previous experience in R&D. The third component of the model links a firm's patent activities with the process of productivity evolution, in which patents and R&D alter the distribution of the firm's future productivity. The last component of the model determines the firm value as a function of current R&D activities and future productivity. In equilibrium, each firm chooses the optimal level of investment in R&D to maximize its firm value.[5]

R&D-patents linkage. We define two binary variables n_{it} and b_{it} , where $n_{it}, b_{it} \in \{0,1\}$, to represent patents for inventions and utility models, respectively. Detailed descriptions of these two types of patents are provided in Section 3. $n_{it} = 1$ $(b_{it} = 1)$ if firm i produces any invention (utility model) patents in year t. Let rd_{it} be the firm's R&D choice; we use $P(n_{it+1}, b_{it+1}|rd_{it})$ to represent the joint distribution of invention patents and utility model patents conditional on the past decision on R&D, .

By formulating the R&D-patent linkage as a conditional joint distribution, we implicitly take the correlation between invention patents and utility model patents into con-

[5]As we show in the online appendix, this framework can easily be extended to accommodate the adjustment of R&D investment at the intensive margin.

sideration. This can be caused by the idea diffusion inside the firm. We also expect that firms engaging in R&D activities are more likely to produce invention or utility patents. We abstract the possibility that different firms have different inclination to protect their ideas by filing patent applications.

Revenue and profits. The demand is CES with an elasticity of substitution being σ. The logged firm's short-run production revenue is given by:

$$r_{it} = (\sigma - 1)(\beta_k k_{it} + \beta_a a_{it} + \phi_{it}) + \mu_0 + \mu_t \tag{1}$$

where k_{it} is the log of the firm's capital stock, and a_{it} is the firm age.[6] μ_0 is a constant term. μ_t is a year-specific variable common to all firms, which contains information on factor prices. k_{it} is treated as a fixed factor in the short-run. ϕ_{it} represents the revenue productivity, which includes the firm's production efficiency as well as the idiosyncratic demand shifter. The firm's short-run profits is:

$$\pi_{it} = \frac{1}{\sigma}\exp(r_{it}) \tag{2}$$

Productivity evolution. We extend the process of productivity evolution in PRVF by assuming that both current R&D activity and future patent counts affect the firm's future productivity. In our specification, the effect of R&D investment on future productivity depends on the patents. Specifically, the distribution of future productivity is affected by a firm's current productivity (ϕ_{it}), R&D activities (rd_{it}), invention patents (n_{it+1}), and utility model patents (b_{it+1}) in the next period. The evolution equation of the firm-level productivity is given by:

[6]See Appendix B for the detailed derivation.

$$\phi_{it+1} = h(\phi_{it}, rd_{it}, rd_{it} \times n_{it+1}, rd_{it} \times b_{it+1}) + \varepsilon_{it+1}, \tag{3}$$

where $h(\phi_{it}, rd_{it}, rd_{it} \times n_{it+1}, rd_{it} \times b_{it+1})$ is the conditional mean of future productivity and ε_{it+1} is an i.i.d stochastic shock normally distributed with zero mean and variance σ_{ε}^2. This formulation assumes that (1) a firm's productivity is persistent over time, implying that the future productivity will be correlated with its current productivity; (2) R&D and patent counts jointly shift the mean of future productivity, with R&D being the fundamental source of endogenous productivity change; and (3) productivity change is influenced by stochastic shocks ε_{it+1}. More importantly, we allow the impact of R&D on future productivity depends on the patent outcomes. To account for the difference between invention patents and utility model patents in affecting the firm's future productivity, we allow that $\partial h / \partial n_{it+1}$ and $\partial h / \partial b_{it+1}$ to be different. It is worth noting that the formulation of the process of productivity is different from that considered in PRVF. In the specification of PRVF, only the innovation outcomes enter into the evolution process of productivity. In other words, R&D investment can only affect productivity evolution through the measured innovation outcome. In our setting, it is clear that R&D is the fundamental source of productivity growth, but the magnitude of its impact is determined by the innovation outcome. Therefore, R&D investment has both direct and indirect effects on productivity growth.

R&D costs and equilibrium. Following PRVF, the innovation cost is assumed to be dependent on prior R&D experience and current capital stock. For firm i in year t, its innovation cost C_{it} is given as:

$$C_{it} \sim \exp(\kappa_m \times rd_{it-1} \times k_{it} + \kappa_s \times (1 - rd_{it-1}) \times k_{it}), \tag{4}$$

where $\exp(\cdot)$ represents the exponential distribution. Hence the cost of investing in R&D follows an exponential distribution with a mean of $\kappa_m k_{it}$ when $rd_{it-1} = 1$, and with a mean of $\kappa_s k_{it}$ when $rd_{it-1} = 0$. The innovation cost is observed by the firm, but not to us as the econometricians. Therefore the innovation cost is an additional factor that affects the firm's behavior of investing in R&D. κ_m and κ_s can be different, implying that the distribution of maintenance costs differ from start-up costs. $\kappa_s k_{it}$ captures the start-up costs when a firm did not participate in R&D activities in the previous period but plan to undertake R&D investment in the current period. In contrast, $\kappa_m k_{it}$ reflects the maintenance costs when a firm was active in R&D investment in the previous period and continue to invest in R&D. k_{it} enters the distribution of R&D costs because of the scale effect that a firm with larger capital stock are required to hire more researchers and build larger research labs. Throughout the estimation, we treat k_{it} to be exogenous. The state variables are $s_{it} = (\phi_{it}, rd_{it-1})$. The firm's decision on R&D will affect the evolution of s_{it}. The firm's expected value function $V(s_{it})$ can be expressed as:[7]

$$\begin{aligned} V(s_{it}) = \pi(\phi_{it}) + & \\ & \int_0^\infty \max_{rd_{it}} \{\beta \mathbf{E} V(s_{it+1}|\phi_{it}, rd_{it} = 1) - C_{it}, \beta \mathbf{E} V(s_{it+1}|\phi_{it}, rd_{it} = 0)\}\, dG(C_{it}), \end{aligned} \tag{5}$$

where β is the discount factor. We define γ_{it} as the parameter for the exponential distribution:

$$\gamma_{it} \equiv \kappa_m \times rd_{it-1} \times k_{it} + \kappa_s \times (1 - rd_{it-1}) \times k_{it},$$

then $G(C_{it}) = 1 - \exp(-C_{it}/\gamma_{it})$ for $C_{it} \geq 0$ and zero otherwise. The expected future value

[7] Note that the firm's value function is

$$\tilde{V}(s_{it}) = \max_{rd_{it}} \{\pi(\phi_{it}) - rd_{it} C_{it} + \mathbf{E}\tilde{V}(s_{it+1}|s_{it}, rd_{it})\}$$

The integrated Bellman equation (5) is obtained by taking the expectation of the value function above with respect to C_{it}.

of the firm is an expectation over the future productivity levels and the count of patent applications:

$$\mathbf{E}V(s_{it+1}|s_{it}) = \sum_{n_{it+1}} \sum_{b_{it+1}} \int_{\phi'} V(s_{it+1}) dF(\phi'|\phi_{it}, n_{it+1}, b_{it+1}, rd_{it}) P(n_{it+1}, b_{it+1}|rd_{it}) \tag{6}$$

Note that (6) is composed of two parts representing two kinds of uncertainties facing innovation. The first uncertainty comes from the creating of applicable patents (or creating new ideas); the second uncertainty comes from the response of future productivity to future patenting activities and R&D decision in the previous year. The firm maximized its firm value, which implies that a firm will choose to invest in R&D if and only if

$$\Delta EV(\phi_{it}) \equiv \mathbf{E}V(s_{it+1}|\phi_{it}, rd_{it} = 1) - \mathbf{E}V(s_{it+1}|\phi_{it}, rd_{it} = 0) \geq C_{it} \tag{7}$$

In equilibrium, a firm will only invest in R&D as long as the expected net benefit from R&D is greater than the costs.

2.2 R&D benefits decomposition and patent value

R&D benefits and its decomposition. Following PRVF, the long-run benefits of R&D are measured as the relative change in the expected firm value caused by R&D investment:

$$LB(\phi_{it}) = \frac{\mathbf{E}V(s_{it+1}|\phi_{it}, rd_{it} = 1) - \mathbf{E}V(s_{it+1}|\phi_{it}, rd_{it} = 0)}{\mathbf{E}V(s_{it+1}|\phi_{it}, rd_{it} = 0)} \tag{8}$$

One novelty of the current paper is a decomposition of R&D's returns into the patent and non-patent channels. Conditioning on that a firm is undertaking R&D investment, its expected value is a weighted average of different states of the realization of patents (n_{it+1}, b_{it+1}). When the patent count is zero and the firm is active in R&D investment, that is $(n_{it+1}, b_{it+1}, rd_{it}) = (0,0,1)$, the firm benefits solely from the non-patent channel, which is given by:

$$LB_N(\phi_{it}) = P(0,0|1)\int_{\phi'} V(s_{it+1})dF(\phi'|\phi_{it},0,0,1), \tag{9}$$

Similarly, the R&D benefits through the realized patent is:

$$LB_P(\phi_{it}) = \sum_{\{n',b':n'+b'>0\}} P(n',b'|1)\int_{\phi'} V(s_{it+1})dF(\phi'|\phi_{it},n',b',1) \tag{10}$$

While the benefits of R&D (LB) are interesting by themselves, the decomposition of LB into the patent and non-patent channel provides a deeper view of the internal structure of R&D benefits. Since LB_p contains information on the firm's patent value that is reflected in the productivity growth, the relative importance of LB_p in total R&D benefits is closely related to the quality of the patent system. When LB_N plays a dominant role in the total benefits of R&D, we can expect that the value of the patent is not so valuable and most of the R&D benefits are realized simply through non-patent activities such as knowledge accumulation.

Patent value. In Pakes (1986), the distribution of returns from holding patents is estimated by solving the patentee's optimal stopping problem of whether renewing the patent or not. Introducing patent counts into the productivity evolution enables us to analyze the patent value using a new approach. To obtain the value of patents, we need to condition on the firm's R&D investment since R&D investment is the fundamental source of patents and productivity growth. For firms not participating in R&D, the patent plays no role in stimulating the firm's future productivity growth. After conditioning on the R&D investment, the realized patent will influence the productivity effects of R&D, which ultimately affects the firm value. Following this logic, we define the long-run value of an invention patent as:

$$VP_{inv}(\phi_{it}) = \underbrace{\ln\left[\sum_{b'\in\{0,1\}} \Pr(b'|n_{it+1}=1)\mathbf{E}V(s_{it+1}|\phi_{it},1,b',1)\right]}_{\text{firm value when an invention occurs: } VP^1_{inv}(\phi_{it})} - \underbrace{\ln\left[\sum_{b'\in\{0,1\}} \Pr(b'|n_{it+1}=0)\mathbf{E}V(s_{it+1}|\phi_{it},0,b',1)\right]}_{\text{firm value when no invention occurs: } VP^0_{inv}(\phi_{it})}, \tag{11}$$

where $\Pr(b'|n_{it+1}=n')$ is the probability of the event $b_{it+1}=b'$ conditional on $n_{it+1}=n'$. In a similar way, we can compute the value of a patent of utility model:

$$VP_{uti}(\phi_{it}) = \underbrace{\ln\left[\sum_{n'\in\{0,1\}} Pr(n'|b_{it+1}=1)\mathbf{E}V(s_{it+1}|\phi_{it},n',1,1)\right]}_{\text{firm value when a utility model occurs: } VP^1_{uti}(\phi_{it})} - \underbrace{\ln\left[\sum_{n'\in\{0,1\}} Pr(n'|b_{it+1}=0)\mathbf{E}V(s_{it+1}|\phi_{it},n',0,1)\right]}_{\text{firm value when no utility model occurs: } VP^0_{uti}(\phi_{it})}. \tag{12}$$

Then the expected firm value when a firm invests in R&D can be decomposed as:

$$\begin{aligned}\mathbf{E}V(s_{it+1}|\phi_{it}, rd_{it}=1) &= \Pr(n_{it+1}=1)VP^1_{inv} + (1-\Pr(n_{it+1}=1))VP^0_{inv} \\ &= \Pr(b_{it+1}=1)VP^1_{uti} + (1-\Pr(b_{it+1}=1))VP^0_{uti},\end{aligned} \tag{13}$$

where the unconditional probabilities $\Pr(n_{it+1}=1) = P(1,0|1)+P(1,1|1)$ and $\Pr(b_{it+1}=1) = P(0,1|1)+P(1,1|1)$. In principle, the patent value is defined for each firm. Even if this firm does not submit any patent applications, the formulae (11) and (12) delivers the shadow value of a potential patent. To make the results comparable with existing literature, one can estimate the value of patent focusing on observations with positive counts of patent.

In what follows, we employ the empirical framework to analyze a sample of Chinese high-tech manufacturing firms. We first introduce the data source, then we explain the estimation procedures and the estimation results.

3 Data

3.1 Data sources

Firm-level production data. The first data set contains information on the large and medium-sized Chinese manufacturing firms from 2001 to 2007 complied by China's National Bureau of Statistics (CNBS hereafter). This data set is widely used in studies on Chinese firms (See Hsieh and Klenow (2009), Song et al. (2011), and Brandt et al. (2012) for example). This data set includes all Chinese State-Owned Enterprises (SOEs hereafter) and non-SOEs with annual sales no less than five million *Renminbi* (equivalent to about 700,000 US dollars). These firms account for 98% of the manufacturing exports. This data set contains all the information about the firm's major accounting sheets, which includes more than 100 financial variables. Serving for this study, it includes firm sales, number of employees, material input, fixed assets, R&D expenditures, and other firm characteristics like firm age and its industrial code. In summary, this rich data set provides information on firm-level production activities. We follow Brandt et al. (2012) to clean the dataset and focus a sample of high-tech manufacturing firms from this dataset.[8]

Patent data. The second database is on patent statistics collected by the State Intel-

[8]Focusing on high-tech firms help alleviate the concern that many firms may fail to report their R&D expenditures. This is because high-tech firms usually undertake a large amount of R&D investment, and the possibility of not reporting small R&D expenditures is much lower than the non-high-tech industries.

lectual Property Office (henceforth SIPO) of China. It contains all the patents that are applied by Chinese firms and granted in mainland China. For each patent, the database provides information on its type (invention, utility model, and design), owner, application time, certification time, the agent of application, abstract, location, and expiration time during 1985 and 2012. But it should be noted that there is no information on citations for patents in the database, which makes it difficult to measure the patent quality directly. According to China's Patent Law, the utility model refers to a new technical solution suitable for practical use proposed for the shape, construction or combination of the products.[9] Generally, an invention patent is also related to a new technical solution proposed for the product, method, or its improvement. But the patenting process for invention patent consists of a "substantive review" which specifically emphasizes on the novelty and originality of the breakthrough in technical upgrading. Lower creativity standards are enforced for utility model patents.As for design patents, they represent more rudimentary types of innovation and are considered to be of lower quality than invention patents and utility model patents.[10] Therefore we anticipate that design patents is less related to the firm's productivity. Considering this, we focus on invention patents and utility patents in the empirical investigation.

Final combined database. We follow He et al. (2018) to match the aforementioned two datasets.[11] In the last row of Table 1, we show that across all years, the total num-

[9]According to WIPO, utility models are sometimes referred to as "short-term patents", "utility innovations" or "innovation patents"......In general, utility models are considered particularly suited for protecting inventions that make small improvements to, and adaptions of, existing products or that have a short commercial life.

[10]According to Article 22 of the Patent Law of the P.R.C.: any invention or utility model for which patent right may be granted must possess novelty, inventiveness and practical applicability. In comparison, the requirement for the approval of design patents is in Article 24 of the Patent Law of the P.R.C as "...... must not be identical with or similar to any design which, before the date of filing, has been publicly disclosed in publications in the country or abroad or has been publicly used in the country, and must not collide with any legal prior rights obtained by any other person."

[11]Unfortunately name is the only information that appears in both the NBS survey and the patent data. He et al. (2018) designed data parsing and pre-processing routines to clean and stem firm and assignee

ber of invention patents in our merged data accounts for over 50% of the total number of patent reported in the China Statistical Yearbook on Science and Technology.[12] This suggests that the merged dataset captures the majority of the patenting activities in China. In Table 1, we also report the total number of invention- and utility-patents across years. Note that there is a strong upward trend for both patent types over our sample period. The upward trend of patent counts consistently reflects the explosive growth of patents during this period, which is well documented in the literature.[13].

Table 1: Patent counts in the merged database and matching efficiency

year	2001	2002	2003	2004	2005	2006	2007
invention	1982	4462	5333	7993	10100	17033	19750
utility model	4202	5649	7496	7798	10720	15324	18212
matching	57.10%	81.26%	96.35%	87.20%	57.58%	67.33%	55.67%

Note: matching efficiency refers the ratio of number of invention patents in the merged dataset to the published figures in China Statistical Yearbook on Science and Technology 2001-2007.

3.2 Descriptive statistics

China's high-tech industries mainly cover four 2-digit industries: pharmaceutical manufacturing, special equipment, electric machinery, and electronics. In Table 2 we report the summary statistics for the R&D and patenting activities in the final dataset. The average R&D expenditure for high-tech manufacturing firms is 218.095 thousand yuan

names for the Chinese manufacturing firms from 1998-2009, and created the matching algorithm tailored for the NBS and patent data. The algorithm maintains a balance between matching accuracy and workload of the manual check. We also implement a systematic manual check process to filter out false positives matches. As we only focus on high-tech industries between 2001 and 2007, our sample size is much smaller, so the workload of manual checking is manageable.

[12]The ratio varies across years, with 55.57% in 2007 and 96.35% in 2003. This is probably because the census data only contains large and medium-sized Chinese manufacturing firms, while the aggregate statistics on patent counts are for all Chinese firms including many other non-manufacturing sectors.

[13]According to the World Intellectual Property Organization (WIPO), The number of domestic invention patent filings with the Chinese patent office (SIPO) has increased at an average annual rate of 32% during the period 1999–2013. Many different explanations for the stunning growth has been discussed (e.g., Hu and Jefferson (2009); Eberhardt et al. (2016); Hu et al. (2017); Chen and Zhang (2019))

(equivalent to around 31,584 US dollars). The R&D intensity measured by total R&D expenditures over total sales is lower than that reported for developed countries. Lastly, compared to R&D participation, we observe that the probability of generating a patent is relatively low. Even we focus on high-tech firms, the pooled probability of generating an invention patent application is only 5.6%.

Table 2: Summary statistics for high-tech and non-high-tech industries

	High-tech		Non-high-tech	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
R&D expenditures	218.095	963.152	34.886	326.677
R&D/employees	1.794	8.327	0.282	4.686
R&D/sales	0.007	0.024	0.001	0.008
Pr(R&D>0)	0.289	0.454	0.106	0.307
Inventions	0.056	0.776	0.010	0.306
Utility models	0.083	1.058	0.028	0.376

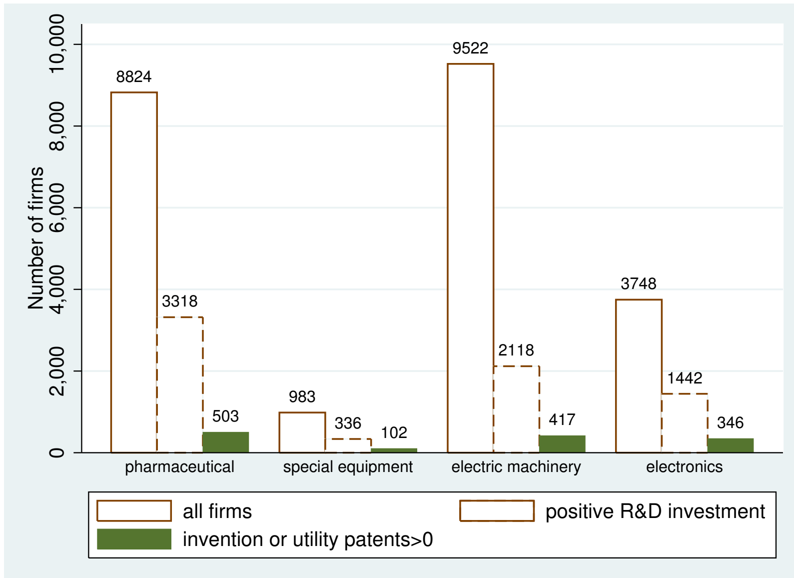
Note: the unit of R&D expenditures is 1,000 yuan (around 150 US dollars).

In Figure 1, we present the number of firms for each 2-digit high-tech industry by their innovative activities. We find consistent patterns across industries: only a small portion of firms in these high-tech industries invest in R&D, an even smaller fraction of them file for patents. Moreover, the difference between the number of firms filing for patents and the number of firms having R&D is substantial across all industries, suggesting it may be important to distinguish innovation inputs and innovation outcomes. The smaller fraction of patenting firms also indicates that innovative firms may face great uncertainty in generating patents.

3.3 Quality of R&D data

Mairesse et al. (2005) provides evidence on the substantial measurement error of using R&D expenditures to predict the innovation probability for the firm-level data from

Figure 1: Number of Firms by Innovation Activities

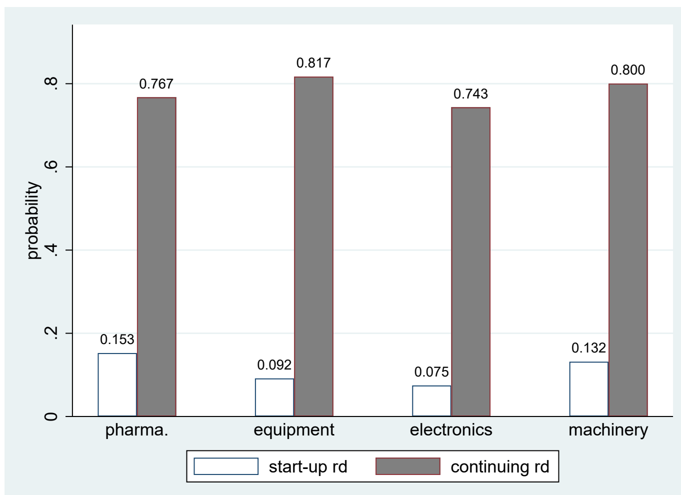


Note: numbers are from the final database

the French innovation survey. More related, Chen et al. (2017) document that corporate income tax reductions induce Chinese firms to relabel administrative expenses as expenditures on R&D. These results suggest that the levels of R&D may not accurately reflect the firm's actual investment in innovation. In appendix A.2, we also show that the extensive margin of R&D and patents captures most of the innovation effects. Therefore we prefer using the information on the R&D choice at the extensive margin as PRVF did. The estimation methodology does not require the researchers to have information on the intensive margin of R&D investment. Instead, it assumes that the econometrician is ignorant about the true R&D costs paid each firm, therefore to some extent it is robust to measurement errors in R&D data at the intensive margin. As long as the extensive-margin status of R&D is measured correctly, the PRVF method still holds when the distribution of R&D costs is well assumed.

In our model, a firm's current state includes its R&D choice in the previous period. Firm i is undertaking start-up R&D if $(rd_{it-1}, rd_{it}) = (0,1)$, and is participating maintenance R&D if $(rd_{it-1}, rd_{it}) = (1,1)$. The start-up R&D is not the innovation expenses for a new project as we do not have information on the operation of research projects by each firm. In this sense, our definition of different types of R&D is loose and only reflects the continuity of R&D investment in a two-period fashion. For later estimation, we require a firm to exist for at least two consecutive years to identify the firm's R&D status in the past and current periods.

Figure 2: Transition probability for start-up and continuing R&D



Note: the probability of start-up R&D is calculated as $\Pr(rd_{it} = 1|rd_{it-1} = 0) = \sum_{i,t} \mathbf{1}(rd_{it} = 1)/\sum_{i,t-1} \mathbf{1}(rd_{it-1} = 0)$, and the probability of continuing R&D is computed as $\Pr(rd_{it} = 1|rd_{it-1} = 1) = \sum_{i,t} \mathbf{1}(rd_{it} = 1)/\sum_{i,t-1} \mathbf{1}(rd_{it-1} = 1)$.

In Figure 2, we show the probabilities for start-up and maintain R&D. In the sample, 80% of firms that undertake R&D investment in the current period tend to continue their R&D investment in the next period, but only around 10% of non-R&D participants would ever start new R&D investment in the next year. This implies that firms face a high adjustment cost for R&D investment, which is consistent with the R&D literature documenting that firms tend to smooth R&D spending over time because of a long time between conception and commercialization (Hall et al., 1986; Lach and Schankerman, 1989; Hall et al., 2010). Our finding is also consistent with the pattern of the transition dynamics of R&D status in recent studies using firm-level datasets from other countries or regions, such as Aw et al. (2011) for Taiwanese firms and Peters et al. (2017) for German firms. We also notice that according to our definition of R&D starters and continuers, the R&D starters are smaller than R&D continuers in terms of capital stock, the total number of employees. Also, the R&D starters are, on average, two years younger than the R&D continuers.[14] Unfortunately, we are not able to identify whether firms stop R&D or fail to report it. If firms fail to report it, we may over-estimate the probability of R&D starters and hence the estimates of start-up costs of R&D are biased downward.

4 Estimation and Results

The empirical method we use follows PRVF closely with only one exception: we include both R&D and patents in the productivity evolution. In PRVF, they estimate to primitives regarding the impact of R&D investment: (1) Pr(innovation|R&D), and (2) Pr(productivity|innovation). In our case, we still estimate (1), but (2) becomes the distribution of productivity conditional on innovation and R&D investment. Instead of using a direct measure of innovation outcome, we use patent counts as the indicator for innovation. In addition to computing the total benefits of R&D, we provide a decomposition of R&D benefits into the patent and non-patent channels. We also estimate the average value of a patent.

[14] See the data appendix for relevant statistics.

4.1 Productivity estimation

Revenue equation. We parameterize the productivity evolution process as a cubic function of lagged productivity:

$$\begin{aligned}\phi_{it+1} = &\rho_0 + \rho_1 \phi_{it} + \rho_2 \phi_{it}^2 + \rho_3 \phi_{it}^3 \\ &+ \rho_4 rd_{it} + \rho_5 (n_{it+1} \times rd_{it}) + \rho_6 (b_{it+1} \times rd_{it}) + \varepsilon_{it+1},\end{aligned} \tag{14}$$

where the first line on the right-hand side captures the persistence in productivity trajectory, the second line describes the impacts of R&D and patents on the evolution of productivity. It is clear that past R&D activity and patent counts jointly affect the future productivity. In particular, we have $\frac{\partial^2 \phi_{it+1}}{\partial rd_{it} \partial n_{it+1}} = \rho_5$ and $\frac{\partial^2 \phi_{it+1}}{\partial rd_{it} \partial b_{it+1}} = \rho_6$. These two second-order partial derivatives clearly state that the impact of R&D investment (the innovation input) on productivity is affected by patents (the innovation outcome). We expect that ρ_5 and ρ_6 are different from each other because different types of patents represent different forms of realized innovation. These two parameters also give us information on the quality of patents and the effectiveness of the patenting system. When they are positive, patents help strengthen the productivity effect of R&D investment. It is also possible that they are positive, implying that submitting patent applications weakens the productivity premium caused by R&D investment.[15] We estimate the productivity using the first-order condition of materials.The demand for materials is dependent on the observed capital stock, age, and unobserved productivity. This gives us an expression for productivity:

$$\phi_{it-1} = \left(\frac{1}{1-\sigma}\right) \beta_{t-1} + \beta_k k_{it-1} + \beta_a a_{it-1} - \frac{1}{1-\sigma} m_{it-1} \tag{15}$$

[15] This may be because of the leakage of key information on production technologies to the firm's competitors, which ultimately pulls down the demand facing the firm and reduces the revenue productivity.

where β_t represents the intercept of the CES demand function and the price of variable inputs common to all firms. Combining (14) and (15), then plugging them into (1) yields an empirical equation for the firm revenue:

$$
\begin{aligned}
r_{it} = &(1-\sigma)\beta_k k_{it} + (1-\sigma)\beta_a a_{it} \\
&-\rho_1\left[\beta_{t-1} + \beta_k(1-\sigma)k_{it-1} + \beta_a(1-\sigma)a_{it-1} - m_{it-1}\right] \\
&-\frac{\rho_2}{1-\sigma}\left[\beta_{t-1} + \beta_k(1-\sigma)k_{it-1} + \beta_a(1-\sigma)a_{it-1} - m_{it-1}\right]^2 \\
&-\frac{\rho_3}{(1-\sigma)^2}\left[\beta_{t-1} + \beta_k(1-\sigma)k_{it-1} + \beta_a(1-\sigma)a_{it-1} - m_{it-1}\right]^3 \\
&-(1-\sigma)\left[\rho_4 rd_{it-1} + \rho_5(n_{it}\times rd_{it-1}) + \rho_6(b_{it}\times rd_{it-1})\right] + \mu_0 + \mu_t + v_{it}
\end{aligned} \tag{16}
$$

where $v_{it} = u_{it} - (1-\sigma)\varepsilon_{it}$, with u_{it} being the measurement error to the revenue and exogenous to the firm's decisions on choosing variable inputs or investment in R&D. The estimation of (16) relies on the condition that the composite error v_{it} is uncorrelated with all the explanatory variables on the right-hand side. μ_0 is an intercept which combines constants from the revenue function and the productivity process. μ_t and β_{t-1} are functions of the common time-varying variables including the demand intercept and factor prices. The higher-order powers on ϕ_{it-1} enables us to distinguish β_{t-1} from μ_t and identify up to a base-year normalization. We follow PRVF and employ a two-step estimation strategy. In the first step, we estimate the demand elasticity. In the second step, we replace σ with its estimates and estimate (16) using the Non-linear Least Square estimator.

Demand elasticity. For each industry, note that the ratio of total variable costs to firm revenue VC/R is equivalent to $(1-1/\sigma)$. Therefore, for industry j, we can estimate σ by using the average of the ratio of variable costs to revenue. Table 3 reports the estimation results. Notice that σ varies across industries. For the electronics industry, the estimate of σ is 6.34, the corresponding markup is 1.187. In comparison, for the machinery industry, the demand elasticity is estimated to be -5.043, implying a markup of 1.247. We can also find that the estimates of σ are smaller than that obtained by PRVF using German data. This indicates that Chinese high-tech firms have a lower markup than German high-tech firms.

Table 3: Estimates of the demand elasticities

industry	pharmaceutical	equipment	electronics	machinery
$\hat{\sigma}$	5.926	5.043	6.341	5.415
$\frac{\hat{\sigma}}{\hat{\sigma}-1}$	1.203	1.247	1.187	1.227

Productivity evolution equation. We plug the estimates of σ into (16) to estimate the parameters for the productivity evolution equation. Table 4 reports the full estimation results. Column (1) shows the estimation results of including rd_t, $n_{t+1} \times rd_t$, and $b_{t+1} \times rd_t$ in addition to 3d order polynomials of current productivity. The estimation results show the impact of R&D on productivity hinges on the patenting activities. Note that the marginal effect of rd_t on the expectation of future productivity is $\rho_4 + \rho_5 n_{t+1} + \rho_6 b_{t+1}$, the estimates of which are as follows:

$$\frac{\Delta \mathbf{E}\left(\phi_{t+1} | \phi_t, rd_t\right)}{\Delta rd_t} = .00435 + .0145 \times n_{t+1} + .0137 \times b_{t+1}$$

This indicates that patents play an important role in enhancing the productivity effect of R&D. If we think of a firm with positive investment in R&D in the current period, then the expected increase in productivity would be $\frac{.0145}{.00435} \approx 3.33$ times greater if it produces an invention patent at the end of this period and $\frac{.0137}{.00435} \approx 3.15$ times greater if it generates a utility model patent. In the current model, patents are channels through which R&D spurs productivity growth. This is different from the PRVF model in which the impact of

R&D on future productivity is fully captured by realized process or product innovation.[16]

Table 4: Estimates of productivity evolution equation and cost function

	cubic parameterization	
Productivity evolution:		
rd_t	.00435**	(2.90)
$n_{t+1} \times rd_t$	.0145**	(2.90)
$b_{t+1} \times rd_t$	.0137*	(2.76)
ϕ_t	.824**	(14.96)
ϕ_t^2	.503**	(2.69)
ϕ_t^3	-1.135**	(-14.26)
ρ_0:		
common part	.0295**	(8.34)
Pharmaceutical	-.0144*	(-3.90)
Electronics	-.0132**	(-3.61)
Electric Machinery	-.0116**	(-2.92)
σ_ε	.10	
Cost function:		
k	-.0299**	(-25.82)
$a \in (10, 19)$	.0740**	(12.62)
$a \in (20, 49)$	0.111**	(12.88)
$a \geq 50$	.149**	(7.84)
sample size	22492	

Note: T statistics are in parentheses; * p<0.05, ** p<0.01.

For the endogenous productivity approach, an implicit condition for a firm to be active in innovative activities is that the productivity cannot increase or decrease too fast in order for it to innovate. Otherwise, the productivity is unbounded in the future, which discourages firms from investing in R&D. The estimation results show that the revenue productivity is between -0.454 and 0.817, which implies that the absolute value of the

[16]PRVF find that the coefficient of realized process innovation is 0.029 and that of realized product innovation is 0.036 for German high-tech firms. However, we find it difficult to directly compare our results with theirs. This is because, according to the Chinese patent law, the classification of inventions and utility models is not based on that whether they are related to process innovation or product innovation, but on the level of originality, the examination procedures, and the speed of granting.

first-order derivative of expected future productivity with respect to current productivity is less than one, thus satisfying the requirement for the value function estimation. In the appendix, we display the range of this slope. We also try a quadratic specification in which only the first- and second-order of ϕ_t are included. However, in this case, the fraction of observations that violates this assumption is not negligible; the corresponding results are displayed in appendix A.3.

4.2 R&D-patents relation

By formulating the R&D-patent linkage as a conditional joint cumulative distribution function, our model accounts for the correlation between patent applications for inventions and for utility models. This can be caused by knowledge spillovers across different research projects within the firm. We also expect that firms engaging in R&D activities are more likely to produce patentable inventions and utility models. Lastly, we do not model the possibility that different firms have different inclination to protect their ideas by creating patents. By selecting high-tech industries, we try to alleviate the concern that some firms may not want to protect their innovation via patenting because it is likely that high-tech firms file patent applications when they create new ideas. In addition, our sample period starts from 2002, before which China has implemented several amendments to the patent laws aimed to strength the protection of intellectual property rights (Hu and Jefferson, 2009). As a result, our measure is an average of the industry-specific propensity of submitting patent applications.

We estimate the probability of producing applicable patents conditional on the firm's past R&D status. For notation simplicity, $P\left(n_{t+1}=n', b_{t+1}=b' | rd_t\right)$ is denoted as $P\left(n', b' | rd_t\right)$.

For each industry, these conditional probabilities are estimated by

$$\hat{P}\left(n', b'|d\right) = \frac{\sum_i \sum_t \mathbb{I}\left(n_{it+1} = n'\right) \mathbb{I}\left(b_{it+1} = b'\right)}{\sum_i \sum_t \mathbb{I}\left(rd_{it} = d\right)} \tag{17}$$

where $\mathbb{I}(\cdot)$ is the indicator function and $n', b', d \in \{0, 1\}$. This procedure imposes that the probability of filing patent applications only depends on the firm's past R&D activity. Moreover, the technology of generating patent applications is common to all firms within the same industry.

Table 5: Distribution of patent applications conditional on R&D investment

Industries	$p(0,0)$	$p(1,0)$	$p(0,1)$	$p(1,1)$
pharmaceutical	0.903	0.085	0.007	0.005
equipment	0.826	0.012	0.115	0.047
electronics	0.899	0.009	0.069	0.023
machinery	0.857	0.015	0.094	0.035

Note: $p(x, y) = \Pr(n_{t+1} = x, b_{t+1} = y | rd_t = 1)$.

The results are displayed in Table 5. Innovation probabilities differ across industries. The overall probability of generating an invention or a utility model is very low; most firms undertaking R&D investment are not able to generate any patentable innovation. While the pharmaceutical industry is better at producing invention patents, the other three high-tech industries create more utility models. This may imply that "major innovation" is more prevalent in the pharmaceutical industry, but "minor innovation" is more common in other high-tech industries. This may reflect that Chinese high-tech firms are technological lagged behind and concentrate more on utility models that are of a short commercial life. Last but not the least, there is a certain probability that firms simultaneously generate invention patents and utility models.

4.3 R&D costs and benefits

The firm's probability of investing in R&D is

$$\begin{aligned}\Pr(rd_{it}=1|\phi_{it},rd_{it-1})=&\Pr\left[\Delta EV(\phi_{it},rd_{it-1})\geq C_{it}\right] \\ =&1-\exp\left[-\frac{\beta(\mathbf{E}V_1-\mathbf{E}V_0)}{\gamma_{it}}\right],\end{aligned} \tag{18}$$

where the second equality is based on the assumption that R&D costs follow an exponential distribution. Recall that $\gamma_{it}=\kappa_m\times rd_{it-1}\times k_{it}+\kappa_s\times(1-rd_{it-1})\times k_{it}$. Given the firm's capital stock and past R&D choice, κ_s and κ_m determines the distribution of R&D costs. The cost function links R&D expenditures to capital stock. The identification of κ's relies on the status of R&D investment in the previous period as well as in the current period. Conditional on the firm's past R&D status, the current productivity, and the capital stock, the R&D investment in the current period is associated with R&D costs which are shaped by κ_s and κ_m. The relative variation in current and past R&D status allows us to identify κ's.

Equation (18) also indicates that β can not be separated from γ_{it}. Therefore we employ the annual deposit rate to set the value for β. Let $\bar{R}$ be the average real annual deposit rate. We follow Song et al. (2011) to choose the annual real deposit rate to be $\bar{R}=1.1075$, and hence $\beta=1/1.0175=0.983$.

We follow Rust (1987) to apply the nested fixed-point algorithm to estimate the dynamic discrete choice model. To implement this algorithm, we discretize the productivity space into 100 grid points, the capital stock into 50 grid points. Remember that we have 4 categories of ages and two states for past R&D experience. Therefore, we estimate the value function for $100\times50\times4\times2=40000$ types of firms. We use the methodology proposed by Farmer and Toda (2017) to discretize the non-linear Markov process specified for the productivity evolution. Finally, we assume that the costs are i.i.d across all firms and periods, then the cost parameters can be estimated using the Maximum Likelihood Estimator (MLE) obtained by solving the following problem:

$$\max_{(\kappa_m,\kappa_s)} \left\{ \sum_{i}^{N} \sum_{t}^{T_i} \log \left[rd_{it} \Pr(rd_{it}=1|\phi_{it}, rd_{it-1}) + (1-rd_{it}) \Pr(rd_{it}=0|\phi_{it}, rd_{it-1}) \right] \right\} \tag{19}$$

where N is the sample size of the firm, T_i is the number of periods in which firm i exists in the data. The details of computation are presented in appendix C.

In Panel A of Table 6, we display the estimation results of (κ_s, κ_m). For all Chinese high-tech industries, we find that the start-up costs of investing in R&D are over ten times larger than maintenance costs. The estimates also show substantial variation in expenditures on maintaining and continuing R&D for different high-tech industries. The electronics industry has the largest start-up costs and maintenance costs. This suggests that R&D investment in developing new technologies and ideas on producing electronic products is more costly. On the other hand, the pharmaceutical industry has the lowest start-up costs while the machinery industry has the least maintenance costs.

To see these results more clearly, we translate these estimates into average R&D costs. The average R&D cost is calculated by plugging the industrial average capital stock into the mean value of the specified distribution of R&D costs.[17] We report the results in Panel B of Table 6. The average start-up costs lie between .810 million US dollars for the pharmaceutical industry to 2.331 million US dollars for the electronics industry. While the maintenance costs range from 86 to 140 thousand US dollars. The difference in magnitudes of start-up costs and maintenance costs help explain the high persistence in the R&D investment. We have shown in the data section that firms tend to continue their

[17] Recall that the mean of R&D costs distribution is $\gamma_i = (1-rd_i)\kappa_s k_i + rd_i \kappa_m k_i$ for a firm i. Let $\bar{k}_j$ be the industrial average capital stock, then the average start-up (maintenance) R&D costs is $\bar{\gamma}_j^s = \kappa_s \bar{k}_j$ ($\bar{\gamma}_j^m = \kappa_m \bar{k}_j$).

Table 6: Estimation results of R&D costs

Panel A: Estimates for the costs parameters				
Industries	Pharmaceutical	Equipment	Electronics	Machinery
κ_s	.9125	1.7752	2.8111	1.1743
	(.4669)	(.0294)	(.1233)	(.0877)
κ_m	.1116	.1069	.1685	.1052
	(.0165)	(.0005)	(.0026)	(.002)
LLF	-4163.4	-344	-3209.3	-1554.1
sample size	8603	939	9308	3604
Panel B: Average R&D costs				
Start-up costs	0.810	1.470	2.331	0.961
Maintenance costs	0.099	0.089	0.140	0.086

Note: Standard errors in the parenthesis are obtained by bootstrapping 100 times. The currency unit for R&D costs is million US dollars.

previously started R&D investment at a high probability. The relatively low maintenance costs and high start-up costs provide a good match to the data, showing that firms need to pay a large adjustment cost for R&D investment.

Because our estimation does not require the econometrician to know the actual R&D spending, the unobserved R&D costs are backed out as a parameterized distribution. The uncovered R&D costs parameters (κ_m, κ_s) are common to all firms within the industry, with the firm-level capital stock being the only idiosyncratic component affecting the mean value of the R&D costs. Consider two firms operating in the same industry and are of the same level of capital stock, they would face the same R&D costs according to our setting. Though the current assumption is restrictive, it can easily be relaxed to include other observable firm characteristics that also affect R&D costs.[18] In the case of China, firm ownership may also be an important determinant for the firm's R&D costs as State-Owned firms receive preferential R&D subsidies. In the section for extension

[18]For example, Peters et al. (2016) added a financial strength variable to the specification of the distribution of R&D costs and analyzed the importance of financial strength on the costs and benefits of R&D investment.

and robustness, we extend the current empirical framework to accommodate such situations.

The model contains several pieces and are estimated in different stages. To check how the estimated model fits the data, we first check the closeness between the model-predicted revenue and the data. Then we compare the model-generated R&D activities with the data from two angles: (1) as a cross-section check, we consider the pooled probability of investing in R&D; (2) We also examine if transition dynamics for R&D generated by the model fits the data well. Overall, the estimated model provides a good match for these moments in the data, giving us the confidence to perform further structural analysis on the benefits of R&D and patents. The details of these results are presented in Appendix B.

4.4 Benefits of R&D investment

4.4.1 Aggregate results

The short-run benefits of R&D investment is directly reflected by the changes in productivity, which ultimately influences sales and profits in subsequent periods. In comparison, the long-run gains of R&D can be captured by the changes in the firm's expected future value.[19] We also report the absolute change in firm value to evaluate the long-run benefits of R&D more completely. Note that the measure of benefits is independent of past R&D activities. However, past R&D activities will affect the current innovation choice jointly with the expected benefits from investing in R&D. We present the estimation results in Table 7. We can find that the percentage change in the firm value caused by R&D investment ranges from 0.0287 % to 0.0330 %. On average R&D investment

[19]Note that under the CES demand structure, the proportional change in the profits is the same as that in revenue.

causes around 0.031 % increase in the annual revenue.

Table 7: Short-run and long-run benefits of R&D investment

sectors	Pharm.	Equip.	Elect.	Mach.	Average
Panel A: Short-run					
Δ pct.	0.0287	0.0300	0.0330	0.0302	0.031
Panel B: Long-run					
Δ pct.					
mean	0.205	0.208	0.239	0.230	0.223
median	0.205	0.218	0.242	0.238	
std	0.071	0.051	0.080	0.068	
Δ abs.					
mean	0.101	0.086	0.127	0.089	0.109
median	0.095	0.084	0.115	0.085	
std	0.046	0.031	0.062	0.038	

Note: the absolute change is measured in million USD; the average is a weighted average using the sample size.

Despite that the increase in the firm's annual sales is relatively small, the effect of R&D is amplified in the long-run. This is because historical R&D spending can exert an impact on current productivity and hence the firm value. The estimation results show that innovation spurs around a 0.223% increase in the firm value, with the electronics industry the highest (0.239%) and the pharmaceutical industry the lowest (0.205%). The median of the long-run benefits is close to the mean value, indicating that the distribution is not very skewed. Inspecting the absolute change in firm value, we know that on average the investment in innovation increases the firm value around 0.109 million USD for Chinese high-tech firms. Different high-tech industries have different returns to R&D. On average, firms operating in the electronics sector increases their firm value by 0.127 million USD from R&D investment, while this number is 0.089 in the machinery industry. We also notice that the median value is slightly lower than the mean, implying that the distribution is slightly right-skewed. Interestingly but not surprisingly, the benefits of R&D investment in Chinese high-tech industries are much lower than that obtained for German high-tech firms. In PRVF, the median of the absolute change in firm value for high-tech industries in Germany ranges from 2.331 million euros to 6.770 million euros. The lower private return to investment in R&D speaks partly for the less willingness for Chinese firms to participate in innovation activities.

4.4.2 Decomposing the benefits of R&D

Based on (9) and (10), we decompose the benefits of R&D into four components: (1) no patent; (2) only invention patents; (3) only utility model patents; (4) co-existence of inventions and utility models. In Table 8 we present the average of R&D benefits for each high-tech industry measured by proportional change and absolute change in the firm value, respectively. As a reference, we also show the mean value of total R&D benefits in the row titled 'total'. The results consistently show that creating patents increases the benefits of innovation dramatically. In the pharmaceutical industry, when there is no patent application, the proportional change in firm value is only 0.152%, and the absolute change in firm value is 0.075 million USD. In sharp contrast, when invention patents and utility patents occur, the corresponding change becomes 1.186% and 0.588 million USD. This large difference implies that the patent contributes to private returns to the R&D investment.

Our previous results, however, do not account for the uncertainty of the realization of different states. To understand more about the relative importance of each component of the innovation activities, we multiply each component of R&D benefits by their probability of realization in Table 9. Note that by considering these probabilities, we can calculate the actual contribution of each component in the benefits of R&D investment. Not surprisingly, for all high-tech industries the case of no patent application is

Table 8: Decomposition of the Long-run benefits of R&D investment

	Pharmaceutical	Equipment	Electronics	Machinery
proportional change:				
no patent	0.152	0.120	0.169	0.144
invention	0.672	0.530	0.749	0.638
utility model	0.643	0.507	0.717	0.610
both	1.186	0.930	1.323	1.121
total	0.205	0.208	0.239	0.230
absolute change:				
no patent	0.075	0.050	0.090	0.056
invention	0.333	0.220	0.398	0.247
utility model	0.319	0.210	0.380	0.237
both	0.588	0.385	0.702	0.435
total	0.101	0.086	0.127	0.089

Note: the absolute change is measured in million USD.

the largest component in the benefits of R&D because of the low probability of generating patents. As for the importance of creating invention patents or utility patents, their relative importance varies over industries. This is mainly driven by the difference in innovation probabilities $\Pr(n_{t+1} = n', b_{t+1} = b')$. On average, we find that non-patent R&D investment accounts for between 48% to 67% of returns to R&D, implying that the realization of a large part of the R&D benefits comes from non-patent activities such as the accumulation of tacit knowledge that are not patentable. For the pharmaceutical industry, the relative importance of invention patents is 2.0%, while for other high-tech industries, the contribution of invention patents is only around 27.9%. In contrast, the contribution of utility model patents in these industries is over 20%, much larger than the 2.2% in the pharmaceutical industry. This is because firms in the pharmaceutical industry have a higher chance of creating valuable invention patents than other high-tech industries.

Table 9: Decomposition of the long-run R&D benefits: relative importance

	Pharmaceutical	Equipment	Electronics	Machinery
no patent	0.670	0.478	0.637	0.538
utility	0.022	0.281	0.207	0.250
invention	0.279	0.031	0.028	0.042
both	0.029	0.210	0.128	0.171

Note: each column adds up to one.

4.5 Value of patents

Now we employ (11) and (12) to calculate the value of patents. In principle, the value of patents is defined for each firm. Even if this firm does not file patent applications, our formula gives the shadow value (or expected value) of a patent. To make the results comparable with the literature, we only estimate the value of invention (utility model) patents focusing on the observations with positive invention (utility model) patents. That is, the patent value is reported only when the firm files some patent applications. The estimation results are displayed in Table 10. We can see that invention patents and utility model patents play a significant role in increasing the firm value. Take the pharmaceutical industry as an example, the mean value of a proportional increase in firm value caused by creating an applicable invention patent is 0.544%, and the associated average absolute change is 0.283 million USD. In comparison, the average of the proportional increase in the firm value caused by creating a utility model patent is 0.671%, which is associated with an increase of 0.333 million USD in the firm value. On average, an invention patent causes a 0.764% increase in the firm value, while a utility model leads to a 0.661% in the firm value. This indicates that the value of a patent is about twice as much as the benefits of R&D investment. Also, note that the value of invention patents is smaller than the utility model in the pharmaceutical industry, while the situation is reversed in the other three high-tech industries. Since firms enjoy the largest gain from R&D investment by generating inventions and utility models simultaneously, the conditional probability $\Pr(b_{t+1} = 1 | n_{t+1} = 1, rd_t = 1)$ is also an important factor in explaining the patent value of an invention. A lower $\Pr(b_{t+1} = 1 | n_{t+1} = 1, rd_t = 1)$ can also lead to a lower value for invention patent. In the pharmaceutical industry, the conditional probability is only 0.056, being much lower than the other high-tech industries. As a result, the value of invention patent is estimated to be lower than other high-tech industries.

Table 10: Estimates of patent value

	invention			utility		
	mean	median	std	mean	median	std
proportional change:						
Pharmaceutical	0.544	0.546	0.184	0.671	0.673	0.225
Equipment	0.680	0.713	0.162	0.503	0.528	0.121
Electronics	0.951	0.965	0.308	0.692	0.702	0.226
Machinery	0.784	0.813	0.224	0.595	0.617	0.172
average	0.757			0.661		
absolute change:						
Pharmaceutical	0.270	0.253	0.122	0.333	0.312	0.149
Equipment	0.282	0.275	0.101	0.209	0.204	0.075
Electronics	0.506	0.461	0.242	0.368	0.335	0.177
Machinery	0.305	0.292	0.126	0.231	0.221	0.096
average	0.374			0.326		

Note: value of invention (utility) patents is only reported for observations with invention (utility) patents; the absolute change is measured in million USD.

The patent value measured by our model captures the proportional changes in the firm's value conditioning on that the firm has an investment in R&D in the current period. Notice that the estimated patent value is much higher than the benefits of R&D. This is because the estimated patent value is for the realized patent while the benefits of R&D is related to the expected value of the patent. In this sense, the production-based measure is more closely related to the private value of patents instead of their social benefits, which are realized through knowledge spillovers across industries and firms.[20]

5 Counterfactual Analysis

In this section, we analyze the impact of two different R&D subsidy policies that are currently implemented in China. The first is reducing R&D costs by lowering the firm's borrowing interest rate. We view this as a cost reduction proportional to expenditures on R&D (proportional subsidy). The second policy is a lump-sum transfer to firms planning to undertake R&D investment (lump-sum subsidy).

These two R&D subsidy policies potentially affect firms' value and innovation probabilities differently. We investigate both policies' impact via (1) reducing maintenance costs and (2) reducing start-up costs. In total, we consider four subsidy programs in total. We also quantify the per-unit subsidy's impact and compare its effects with the above two subsidy policies.

5.1 Setup for counterfactuals

5.1.1 Definition of different R&D subsidies

Proportional subsidy. We first consider a scenario in which the government can perfectly evaluate the innovation projects to know each firm's innovation cost C_{it}. Let $(1-\delta_s)$ and $(1-\delta_m)$ be the subsidy rate for start-up costs and maintenance costs, separately. Under the scheme of proportional subsidy, the amount for subsidy received by each firm is simply $\tau_{Tit} = (1-\delta_T)C_{it}$, where $T \in \{s,m\}$ denote the type of firms subsidized by the

[20] Dang and Motohashi (2015) proposes to use the measure of knowledge breath as a proxy for the quality of patents. In appendix C.4, we show that this method may not be a good indicator of the patent quality. At least, it does not reflect the private value of patents measured by the increase in the firm's value.

proportional subsidy; $T = s$ represents that the subsidy is targeted at the R&D starters and $T = m$ for R&D continuers. In period t, upon receiving a proportional subsidy, firm i's R&D costs become $C(\tau_{Tit}) = C_{it} - \tau_{it} = \delta_T C_{it}$. Then the expected expenditure on proportional subsidies conditional on the firm's capital stock k_{it} is

$$\mathbf{E}(\tau_{Tit}|k_{it}) = (1-\delta_T)\mathbf{E}(C_{it}|k_{it}) = (1-\delta_T)\gamma_{it}, \tag{20}$$

where $\gamma_{it} = \delta_s \kappa_s k_{it}$ for R&D starters and $\gamma_{it} = \delta_m \kappa_m k_{it}$ for R&D continuers. We consider the impact of one-period subsidy. After introducing the proportional subsidy, we can express the expected firm value as:

$$W_{Tit}^{prop} = \pi(\phi_{it}) + \int_0^\infty \max_{rd_t \in \{0,1\}} \{\beta \mathbf{E} V_0, \beta \mathbf{E} V_1 - \delta_T C_{it}\}\, dG(C_{it}) \tag{21}$$

where $G(C_{it})$ is the cumulative density function for the firm's innovation cost C_{it}. Based on the estimates of $\mathbf{E}V_0$ and $\mathbf{E}V_1$ from our previous analysis, we can calculate W_{Tit}^{prop}. Then the long-run effect of a one-period proportional subsidy is defined as the increase in the firm value, which can be expressed as:[21]

$$\begin{aligned} LB_{Tit}^{prop} &= W_{Tit}^{prop} - V(\phi_{it}, rd_{it-1}) \\ &= \gamma_{it}\left[\delta_T e^{-\frac{\beta \Delta EV_{it}}{\delta_T \gamma_{it}}} - e^{-\frac{\beta \Delta EV_{it}}{\gamma_{it}}} + (1-\delta_T)\right] \end{aligned} \tag{22}$$

Lump-sum subsidy. In reality, the government can hardly observe each firm's innovation costs. Noticing this, we consider the case when the government only knows the distribution of R&D costs, as the econometrician does. Since the government does not know the realization of C_{it}, we assume that the innovation subsidy is implemented by targeting the

[21] See the Lemma in Appendix B.2.

observable–the expected R&D costs, $\mathbf{E}(C_{it}) = \gamma_{it}$. In particular, the innovation subsidy is a lump-sum transfer to the firm with an amount of F_{Tit} for type-T firms. We impose that $F_{Tit} = \mathbf{E}(\tau_{Tit}|k_{it})$ so that the expected value of lump-sum transfer received by the firm is equal to the expected proportional subsidies.[22] Under one-period lump-sum subsidy, the firm's expected value function becomes:

$$W_{Tit}^{lump} = \pi(\phi_{it}) + \int_0^{\infty} \max_{rd_{it} \in \{0,1\}} \{\beta \mathbf{E}V_0, \beta \mathbf{E}V_1 + (1-\delta_T)\gamma_{it} - C_{it}\}\, dG(C_{it}) \tag{23}$$

Similar to the proportional subsidy analyzed just now, the long-run effect of the one-period lump-sum transfer on firm value is estimated as:

$$\begin{aligned} LB_{Tit}^{lump} =& W_{Tit}^{lump} - V(\phi_{it}, rd_{it-1}) \\ =& \gamma_{it} \left[e^{-\frac{\beta \Delta EV_{it} + (1-\delta_T)\gamma_{it}}{\gamma_{it}}} - e^{-\frac{\beta \Delta EV_{it}}{\gamma_{it}}} + (1-\delta_T) \right] \end{aligned} \tag{24}$$

Similarly, we can calculate the average probability of innovation after receiving subsidy as P_{Tit}^{prop} and P_{Tit}^{lump} and compute the innovation-enhancing effect of different subsidy programs:

$$P_{Tit}^{prop} = \int_0^{\beta \Delta EV_{it}} dG(c/\delta_T) = 1 - \exp\left(-\frac{\beta \Delta EV_{it}}{\delta_T \gamma_{it}}\right) \tag{25}$$

$$P_{Tit}^{lump} = \int_0^{\beta \Delta EV_{it} + (1-\delta_T)\gamma_{it}} dG(c) = 1 - \exp\left(-\frac{\beta \Delta EV_{it} + (1-\delta_T)\gamma_{it}}{\gamma_{it}}\right) \tag{26}$$

Start-costs subsidy and maintenance-costs subsidy. For both of the proportional subsidy and lump-sum subsidy, the government need to subsidize two types of firms: the R&D starters ($T_{it} = s$) and continuers ($T_{it} = m$). Because these two types of firms have different distributions of innovation costs, the expected government expenditures on subsidizing

[22] This is because $\mathbf{E}(F_{Tit}) = \mathbf{E}(\mathbf{E}(\tau_{Tit}|k_{it})) = \mathbf{E}(\tau_{Tit})$ by the law of iterated expectations.

them can still vary even if $\delta_s = \delta_m$. To be able to compare their effectiveness, we require that their expected expenditures are the same. For R&D continuers and starters, the expected subsidies they receive are:

$$\mathbf{E}(\tau_{it}^s) = (1-\delta_s)\kappa_s \mathbf{E}(k_{it}|T_{it}=s) \text{ (R\&D starters);}$$
$$\mathbf{E}(\tau_{it}^m) = (1-\delta_m)\kappa_m \mathbf{E}(k_{it}|T_{it}=m) \text{ (R\&D continuers).}$$

Equalizing $E(\tau_{it}^s)$ and $E(\tau_{it}^m)$ leads to

$$\delta_s = 1-(1-\delta_m)\frac{\kappa_m \mathbf{E}(k_{it}|T_{it}=m)}{\kappa_s \mathbf{E}(k_{it}|T_{it}=s)}. \tag{27}$$

For each year, we can estimate $\mathbf{E}(k_{it}|T_{it}=m)$ and $\mathbf{E}(k_{it}|T_{it}=s)$ as the sample average of the firm's capital stocks for R&D continuers and starters, separately. Therefore we can use Equation(27) to determine δ_s given any chosen δ_m. This enables us to rule out that the different impacts generated by the subsidies are simply because that R&D starters and R&D continuers receive different levels of financial supports from the government.

Table 11: Amount of subsidies for different policies

Subsidy types	Proportional subsidy	Lump-sum subsidy
R&D starters	$(1-\delta_s)C_{it}$	$C_{it}-(1-\delta_s)\gamma_{it}$
R&D continuers	$(1-\delta_m)C_{it}$	$C_{it}-(1-\delta_m)\gamma_{it}$

In total, we consider two types of subsidy policies and their effects on two types of firms (R&D continuers and R&D starters). We summarize these four different regimes of subsidies in Table 11. In total, we consider two types of subsidy policies (proportional subsidy and lump-sum subsidy). For each type of subsidy, we consider their effects for two types of firms (R&D continuers and R&D starters), separately. With the parameter restriction in Equation 27), the expectation of all four different types of subsidies are equal to $(1-\delta_m)\gamma_{it}$.

5.1.2 Theoretical analysis of different R&D subsidy programs

Proportional versus lump-sum subsidy. Based on the model's setting, we can characterize the effects of the introduced subsidy programs. We focus on two firm margins in evaluating the impact of R&D policies: the expected firm value and the innovation probability. We first compare the outcomes of proportional subsidy with that of the lump-sum subsidy by fixing the firm type $T \in \{s,m\}$, then we consider the outcomes across different types of firms (R&D starters versus R&D continuers) for proportional subsidy and lump-sum subsidy, respectively. We have the following proposition:

Proposition 1. *If the R&D costs follow an exponential distribution, for $T \in \{s,m\}$:*

1. $LB_{it}^{lump} > LB_{Tit}^{prop}$ *for any* $\delta_T \in (0,1)$;
2. *Each firm has a cut-off value* $\bar{\delta}_{Tit} \equiv \beta\Delta EV_{it}/\gamma_{it}$ *such that* $P_{Tit}^{lump} < P_{Tit}^{prop}$ *if* $\delta_T < \bar{\delta}_{Tit}$ *and* $P_{Tit}^{lump} > P_{Tit}^{prop}$ *whenever* $\delta_T > \bar{\delta}_{Tit}$.

Proof. See Appendix B.2. □

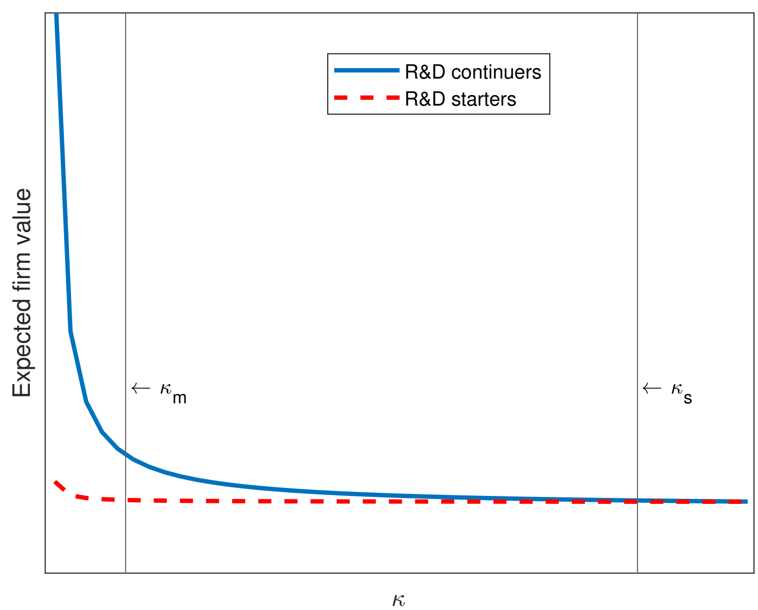
The first result of Proposition 1 shows that conditional on the expected amount of subsidy, lump-sum subsidy always increases the firm value more than the proportional subsidy when the R&D costs follow an exponential distribution.[23].

The second result of Proposition 1 states that the effect of enhancing innovation probability depends on the magnitude of the subsidy. Note that $\beta\Delta EV_{it}/\gamma_{it}$ measures the expected net benefits of innovation, and is positively related to the firm's innovation

[23]We consider the case of a more generalized distribution and its implication for counterfactual analysis in Subsection 6.1

probability in the absence of innovation subsidies (see Equation (18)). According to result 2 of Proposition 1, the lump-sum subsidy stimulates the innovation more than the proportional subsidies when the firm's intrinsic innovation probability is relatively low. This implies that the lump-sum subsidy is more effective than the proportional subsidy for the group of firms who are unlikely to participate in innovation prior to receiving any subsidies.

Figure 3: Shape of the value function



R&D starters versus continuers. R&D starters and R&D continuers differ not only in their R&D status in the previous period, but also in their productivity and capital stock. Conditioning on the type of subsidy policy (proportional or lump-sum), it is hard for us to reach a general conclusion regarding which firms benefit more from the R&D subsidy. However, if we condition on the firm's capital stock and productivity, we can numerically show the R&D continuers will increase their firm value more upon the same reduction in the R&D costs. Figure 3 shows the expected firm value with respect to κ for both R&D continuers and R&D starters. When κ is low, the expected firm value of R&D starters increases more than the R&D continuers upon receiving the same decrease in the R&D costs. But when κ is large, the curve is flattened and the expected firm value is insensitive to a small change in κ. Because the cost parameters we have obtained universally indicate that $\kappa_m < \kappa_s$, we can anticipate that the expected firm value of R&D continuers will increase more after receiving the R&D subsidy. It is clear that the innovation probability is a convex function of κ's, and has a flattened right tail. Similar to our analysis of the firm value, we anticipate the R&D continuers will respond more actively to the R&D subsidies regardless of the subsidy type.

5.1.3 Ex-post effects per unit R&D subsidy

In analyzing the policy effects of proportional subsidy and lump-sum subsidy, we equalize the expected payments of the government for different subsidy programs. However, the "actual" total subsidy can still be different because firms have different incentives to innovate when facing different R&D subsidy programs. As the consequence, the distinct effects of different subsidies on the expected firm value and innovation probability may simply be driven by the differences in their ex-post amount.

To address this concern, we now normalize the change in firm value using the ex-post amount of R&D subsidy. For different subsidy policies, the change in firm value caused by one unit subsidy is defined as:

$$\chi_T^{prop} = \frac{\sum_i \sum_t P_{Tit}^{prop} LB_{Tit}^{prop}}{\sum_i \sum_t P_{Tit}^{prop}(1-\delta_T)\gamma_{it}} \tag{28}$$

$$\chi_T^{lump} = \frac{\sum_i \sum_t P_{Tit}^{lump} LB_{Tit}^{lump}}{\sum_i \sum_t P_{Tit}^{lump}(1-\delta_T)\gamma_{it}} \tag{29}$$

Equations (28) and (29) deserve some explanation. On the right-hand side of these equations, the numerator is the summation of the expectation of the increase in firm value, and the denominator is the expected ex-post subsidies distributed to firms.[24] We need to consider the innovation probability to back out the actual subsidy received by firms, because only firms that are active in R&D investment are able to receive the R&D subsidy. χ_T^{prop} and χ_T^{lump} measure the change in firm value caused by one unit proportional subsidy or lump-sum subsidy, respectively. It is clear that $\chi_T^{prop} > \chi_T^{lump}$ (or $< \chi_T^{lump}$) indicates the proportional subsidy is more (or less) efficient than lump-sum subsidy in increasing the firm value. Similarly, we can define the average efficiency of one-unit subsidy in increasing the innovation probability.

5.2 Counterfactual results

We conduct three groups of experiments by choosing $\delta_m = 90\%$, 85%, and 80%. These experiments display similar results in terms of the rankings of the effectiveness of different R&D subsidy policies. For the sake of brevity, we report the results of the experiment in which $\delta_m = 80\%$.[25] The results for this counterfactual analysis are displayed in Table 12.

In terms of increasing the firm value, there are several interesting findings from this experiment. First, the lump-sum subsidy is much more effective than the proportional subsidy in enhancing the firm value for all Chinese high-tech industries. Second, the efficiency of lump-sum subsidy (effects per unit subsidy) also dominates the proportional subsidy. Third, reductions in maintenance costs increase the firm value more than the start-up costs. This implies that these policies are more effective for R&D continuers.

[24] For the proportional subsidy, the expectation of ex-post subsidy is $\int_0^\infty \Pr(C_{it} < \beta\Delta EV_{it}/\gamma_{it})(1-\delta_T)C_{it}dG(C_{it}) = (1-\delta_T)\gamma_{it}$; for the lump-sum subsidy, we have $\int_0^\infty \Pr(C_{it} < \beta\Delta EV_{it}/\gamma_{it})(1-\delta_T)\mathbf{E}(C_{it})dG(C_{it}) = (1-\delta_T)\gamma_{it}$

[25] Other results are available upon request.

Inspecting the results for increasing the firm's probability of innovation. We note that the proportional subsidy is more effective than the lump-sum subsidy in enhancing the firms' innovation probability. However, the lump-sum subsidy has a higher ex-post efficiency than the proportional subsidy. For both the lump-sum and proportional subsidy, subsidizing the R&D continuers is more effective in stimulating firms to participate R&D investment.

It is worthing noting that our counterfactual results are probably dependent on the parametric assumption of the distribution of R&D costs. In the next section, we further show that most of these results are stable to other distributional assumptions that are more generalized than the exponential distribution. Another key factor that may affects the counterfactual results are the joint distribution of state variables (ϕ_t, rd_{t-1}, k_t). This distribution further determines the cut-off value $\bar{\delta}_{Tit}$ for each firm. When the cut-off value is relatively high, the stimulating effect of proportional subsidy dominates the lump-sum subsidy. Following the same logic, when comparing the R&D continuers with the R&D starters, the difference in their distribution of states and their R&D cost parameters both play a role in shaping the firm's response to the same subsidy program.

6 Discussion and Extensions

6.1 Distribution of R&D costs

In our previous analysis, idiosyncratic R&D costs are assumed to be unobserved by the econometrician and follow an exponential distribution. The exponential distribution has brought us analytical forms in calculating the expected value function, which further reduces the computational burden. Our theoretical characterization also hinges on this distributional assumption. Though this assumption is well accepted in the structural

Table 12: Counterfactual results for the benchmark model

Sectors	Phamaceutical		Equiment		Electronics		Machiney	
R&D continuers:	prop.	lump.	prop.	lump.	prop.	lump.	prop.	lump.
firm value ↑:	0.0665	0.1352	0.0545	0.1169	0.0796	0.1773	0.0576	0.1171
per unit	0.0884	0.6816	0.0801	0.6645	0.0772	0.6430	0.0880	0.6832
innovation Prob. ↑:	0.0752	0.0686	0.0781	0.0701	0.0754	0.0772	0.0759	0.0680
per unit	0.0911	0.3128	0.1086	0.3754	0.0657	0.2507	0.1069	0.3623
R&D starters:								
firm value ↑:	0.0016	0.0252	0.0003	0.0109	0.0005	0.0163	0.0007	0.0164
per unit	0.0022	0.1376	0.0004	0.0648	0.0004	0.0639	0.0011	0.1016
innovation Prob. ↑:	0.0027	0.0227	0.0007	0.0116	0.0006	0.0117	0.0015	0.0169
per unit	0.0032	0.1079	0.0008	0.0637	0.0005	0.0399	0.0020	0.0942

Note: $\delta_m = 0.8$ for this counterfactual experiment; monetary units are in 10,000 US dollars analysis of R&D investment (see Aw et al. (2011); Peters et al. (2017), among others), one may argue that the results, especially the counterfactual analysis, can be sensitive to the chosen parametric form for the distribution of R&D costs. To relieve this concern, we now relax this assumption and consider $G(c)$ being a general cumulative distribution function. Proposition 2 states that the lump-sum subsidy dominates the proportional subsidy in increasing the firm value for a more general class of distributions.

Proposition 2. *If $G(c;\gamma_{it})$ is continuous and twice differentiable with respect to c, with $g(c;\gamma_{it}) = \partial G(c;\gamma_{it})/\partial\gamma_{it}$ being the probability density function and $\bar{C}_{it}$ being the expected value. The proportional and lump-sum subsidy are $(1-\delta_T)C_{it}$ and $(1-\delta_T)\bar{C}_{it}$, respectively. Then for $T \in \{s,m\}$ and $\delta_T \in \{0,1\}$:*

1. *There always exists a cut-off value $\bar{\delta}_{Tit} \equiv \beta\Delta EV_{it}/\gamma_{it}$ for each firm such that $P^{lump}_{Tit} < P^{prop}_{Tit}$ if $\delta_T < \bar{\delta}_{Tit}$, and $P^{lump}_{Tit} > P^{prop}_{Tit}$ whenever $\delta_T > \bar{\delta}_{Tit}$;*
2. *For $\delta_T > \bar{\delta}_{Tit}$, LB^{lump}_{Tit} is greater than LB^{prop}_{Tit}; for $\delta_T < \bar{\delta}_{Tit}$, a sufficient condition for $LB^{lump}_{Tit} > LB^{prop}_{Tit}$ is: (1) $\partial g(c;\gamma_{it})/\partial c < 0$ and (2) $\lim_{x\to\infty} c^3 g(c;\gamma_{it}) = 0$.*

Proof. See Appendix B.2. □

In the result 2 of Proposition 2, condition (1) requires that the probability mass accumulates more for low R&D costs, condition (2) restricts that the probability density function of the R&D costs must have a thin right tail. Both of these conditions point out that the distribution of R&D costs should be skewed towards low R&D costs. It can be verified that many distributions in the exponential family satisfy these conditions under some restrictions. As robustness analysis for our quantitative exercise, we let the R&D costs follow a Weibull distribution with a shape parameter θ and a scale parameter γ_{it}.

In this case, the cumulative density function of C_{it} is given by

$$G(c) = 1 - e^{-\left(\frac{c}{\gamma_{it}}\right)^{\theta}}. \tag{30}$$

Here θ is common to all firms and is related to the variance of the R&D costs. We can verify that $\lim_{c\to\infty} c^3 g(c) = 0$ using the L'Hospital's rule. Note that when $\theta = 1$, the Weibull distribution degenerates into the exponential distribution. The density function is $g(c) = \frac{\theta c^{\theta-1}}{\gamma_{it}^{\theta}} e^{-\left(\frac{c}{\gamma_{it}}\right)^{\theta}}$. When $\theta \leq 1$, the probability density function is decreasing in c. According to Proposition (2), the lump-sum subsidy increases the firm value more than the proportional subsidy. If $\theta > 1$, this density function has a bell shape.[26] In this case, the relative effectiveness of proportional and lump-sum subsidy become less clear. Hence we re-estimate our model with θ being 0.5, 1.5, and 2 and repeat our counterfactual analyses.[27]

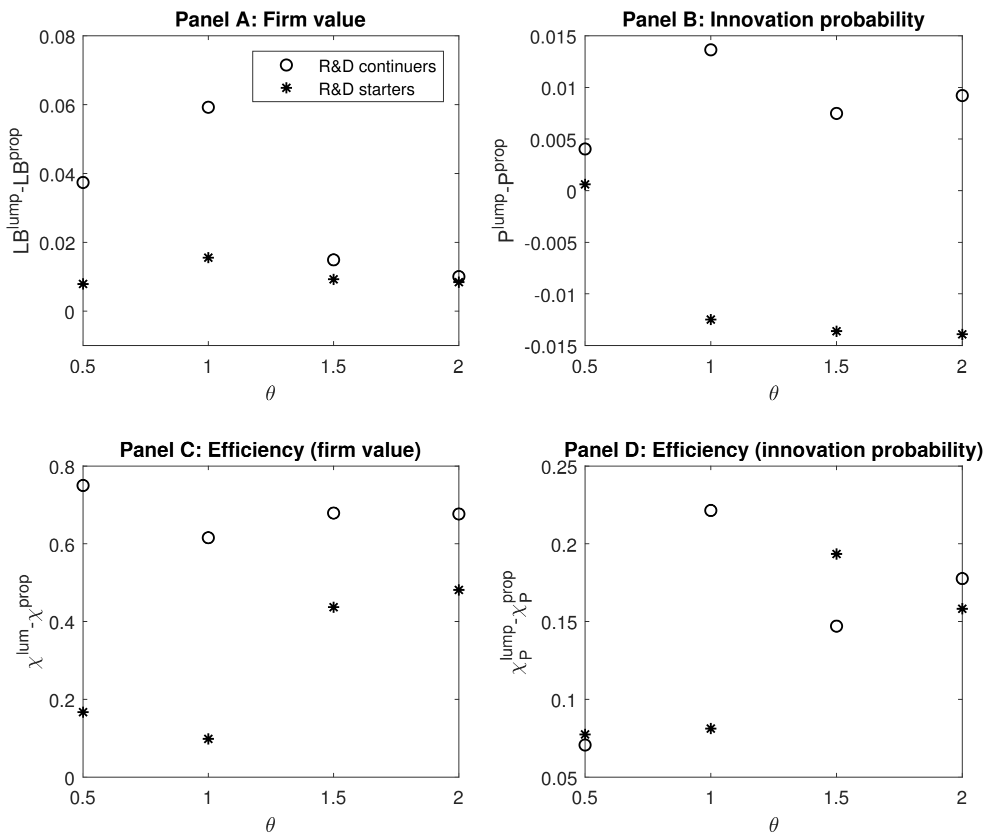
The estimates for κ's and the full counterfactual results are presented in Appendix A. We find that, for the average costs of R&D, the benefits of R&D and its decomposition into different channels, as well as the value of the patent, the estimation results stay close to our benchmark results.[28] We summarize the counterfactual results in Figure 4. For different values of θ, we take the difference between the outcome of lump-sum subsidy and proportional subsidy and average them by pooling firms in different industries. From Panel A, we know that the lump-sum subsidy performs better in increasing the

[26] Another potential generalization of the Exponential distribution is the Gamma distribution of which the probability density function is $g(c) = \frac{x^{\theta-1} e^{-\frac{c}{\gamma_{it}}}}{\gamma_{it}^{\alpha}\Gamma(\alpha)}$ for $c > 0$ and zero otherwise, with θ being the shape parameter and γ_{it} characterizes the mean. When $\theta = 1$, the Gamma distribution turns to be the exponential distribution. Using Proposition 2, it is easy to show that when $\theta \leq 1$, the lump-sum subsidy dominates proportional subsidy in terms of increasing the firm value.

[27] When the R&D costs have a Weibull distribution, the expected value of R&D costs is given by $\mathbf{E}(C_{it}|k_{it}) = \Gamma(1+\frac{1}{\theta})\gamma_{it}$. We choose the subsidy rate $(1-\tilde{\delta}_m) = \frac{1-\delta_m}{\Gamma(1+\frac{1}{\theta})}$ so that the expectation of the amount of innovation subsidy is equalized for different projects.

[28] To save space, we do not provide these results here; they are available upon request.

Figure 4: Counterfactual results for Weibull- Distributed R&D Costs



Note: θ is the shape parameter for the Weibull distribution.

firm value as a total. The difference between the lump-sum subsidy and proportional subsidy diminishes as we improve the value of θ. Panel C shows that the lump-sum subsidy causes a larger increase in firm value per-unit subsidy than the proportional subsidy. But their difference shrinks as we increase θ. This pattern also exists for the R&D starters. As for the change in innovation probability, compared to the benchmark model, we observe a smaller gap between the proportional and lump-sum subsidy when we adjust θ to be 0.5. As we increase the value of θ, this gap becomes wider. When we look at the relative efficiency of stimulating innovation probability, depending on the value of θ, there is no clear pattern for which subsidy policy should work better. Overall, these results confirm that our benchmark results are not unique to the exponentially distributed R&D costs and are stable to more generalized distributions of R&D costs.

6.2 Firm ownership

One concern about using Chinese data is the preferential subsidy towards State-Owned Enterprises (SOEs). The co-existence of SOEs and non-SOEs means that firms may have different R&D costs simply because of their differences in the ownership. If this is the case, we should consider the effects of firm ownership on R&D costs in the specification of R&D costs. Recall that in our benchmark model, we relate the mean value of R&D costs to the firm size (measured by capital stock) in equation (4). We can easily incorporate firm ownership in the specification of our model.[29] Let o_{it} be the variable indicating firm ownership: o_{it} equals one for SOEs and zero otherwise. We assume the distribution of R&D costs is given by:

$$C_{it} \sim \exp\{rd_{it-1} \times (\kappa_m + \kappa_{mo} o_{it})k_{it} + (1 - rd_{it-1}) \times (\kappa_s + \kappa_{so} o_{it})k_{it}\}, \tag{31}$$

where κ_{mo} and κ_{so} measure the effects of firm ownership on continuing and start-up R&D costs, respectively. A negative κ_{mo} (κ_{so}) implies that SOEs have lower maintenance (start-up) costs than non-SOEs conditional on their capital stocks. The estimation of the extended model is similar to the benchmark model; the only difference is that we now have two additional parameters $(\kappa_{mo}, \kappa_{so})$ to characterize the distribution of R&D costs.

[29] A similar extension can be found in Peters et al. (2016) which treats financial strength as a determinant of the firm's R&D costs.

Table 13: Estimation results for the extended model with firm ownership

	κ_s	κ_m	κ_{so}	κ_{mo}	N	LLF
Pharmaceutical	.904	.111	1.101	.085	8538	-4087.80
	(.081)	(.003)	(2.343)	(.038)		
Equipment	1.776	.107	2.210	.038	930	-335.70
	(.086)	(.002)	(1.546)	(.017)		
Electronics	2.786	.170	1.358	.143	9248	-315.50
	(.121)	(.004)	(1.327)	(.031)		
Machinery	1.209	.105	-.063	.033	3565	-1526.90
	(.079)	(.002)	(40.40)	(.030)		

Note: Standard errors in the parenthesis are obtained by bootstrapping 50 times. The currency unit for R&D costs is million US dollars.

In Table 13, we present the estimation results for the extended model with firm ownership. We note that the estimates for κ_s and κ_m are very close to that obtained from the benchmark model. The estimate for κ_{so} is not significant for all of the four industries. This indicates that SOEs do not differ substantially from non-SOEs when starting their innovation activities. In the column for κ_{mo}, the estimates are significantly positive except for the industry of machinery, which suggests that SOEs tend to pay more maintenance costs conditional on the capital stock. This might be because of the preferential subsidies that allocated to SOEs have stimulated their R&D investment.

We have also tried to include the ownership variable o_{it} and its interactions with rd_{it}, $rd_{it} \times n_{it+1}$, and $rd_{it} \times b_{it+1}$ in the productivity evolution equation, but the estimates of these variables show no significance. This implies that the endogenous productivity equation for SOEs is not significantly different from the non-SOEs. The estimation results for productivity are reported in Appendix A.3. Overall, adding the indicator for SOEs in the model does not change our results regarding the benefits of R&D, though it implies higher maintenance costs for SOEs.

7 Conclusion

Understanding the costs and benefits of R&D investment is crucial for the designing of innovation policy in spurring R&D investment and enhancing the firm value. It is usually difficult to measure the innovation outcomes because not all inventions are patentable and the accumulating tacit knowledge is hard to measure. Our empirical framework includes both the innovation input (the R&D investment) and imperfect measures for innovation outcome (patents) in the productivity evolution equation and does not require that researchers perfect observe the innovation output caused by R&D investment. Moreover, based on the model, we propose an empirical framework to decompose the benefits of R&D into the patent and non-patent channels, as well as a novel estimator of the patent value conditioning on the firm's R&D choice.

We apply the empirical model to a sample of Chinese high-tech manufacturing firms between 2001 and 2007. We find that Chinese high-tech firms generate much lower benefits from innovation than estimates of high-tech firms in Germany obtained by PRVF. More interestingly, we document that most of the benefits of R&D investment originate from the non-patent channel. We show that the lump-sum transfer performs better than the proportional subsidy in increasing the firm value both empirically and theoretically for a large class of distributions.

In the current empirical framework, we treat the R&D-patent relation to be exogenous and do not model the firm's patenting choice. We also do not consider the quality differences in patents produced by different firms in the sense that each patent causes the same percentage change in the firm's productivity. Richer models incorporating endogenous patenting choice and patent quality differences will be an important avenue for future research.

Appendices

A Data and Supplementary Estimation Results

A.1 Characteristics of R&D Starters and Continuers

In the table below, we display several firm characteristics for R&D starters and continuers, respectively.

Table A.1: Firm characteristics: Starters vs. Continuers

variables	log(capital)	log(employees)	age
R&D starters	8.38	4.75	16.16
R&D continuers	8.77	4.93	18.10
Total	8.50	4.80	16.76

A.2 Data features of R&D-Patents Relation

In this appendix, We present more features of the data for Chinese high-tech manufacturing firms. In particular, We will discuss the distributional characteristics for patents and the R&D-patents relation. These discussions suggest that the extensive margins of R&D and patents captures most part of the innovation activities.

Distribution of patents. Table A.2 reports the distribution of patents. We can see that the distribution of patents is highly concentrated at zero for all three types of patents. The share of firm observations with zero invention (utility) patents in the final sample is 98.01% (95.72%). This implies that only a small fraction of firms file patent applications. Focusing on the positive part of the distribution, the percentage of firm observations filing only an (a) invention-(utility-)patent is 1.22% (2.22%), and that of firm observations filing two invention (utility) patents is 0.39% (0.96%). Moreover, the number of firm observations submitting no less than three invention- (utility-) patents account for 0.37% (1.10%) in the sample. Overall, these observations imply that the variation in the outcome of the patent in positive part (extensive margin) is much less significant than the change from zero to one (i.e. the extensive margin) for invention and utility patents. In other words, among firms that have positive patent applications, most of the firms (over 50%) have only one patent. Overall, these results suggest that we have to rely on the variation in the extensive margin to identify the productivity effects of the patents.

Table A.2: Distribution of patents for high-tech manufacturing firms

Patents counts	0	≥ 1	1	2	≥ 3
Invention	98.01%	1.99%	1.22%	0.39%	0.37%
Utility	95.72%	4.28%	2.22%	0.96%	1.10%

Note:the percentage represents the share of observations in the specified cohort.

R&D-patents relation. The R&D-patents linkage is an important part in the structural model to be explained in the next section. Since we do not have a direct measure for innovation, we rely on the patents to measure the outcome of innovation. Different from the indicators of process and product innovation used in PRVF, we observe the number of invention patents and/or utility patents filed by each firm. To check the validity of using patents as indicators for innovation, we report the correlation between patents and R&D at both intensive and extensive margins in Table A.3. We estimate a linear model relating patents applications to R&D investment controlling for industry, and year fixed effects. We also control for firm size by use R&D intensity defined as the ratio of R&D expenditures to the firm's sales.

The matrix of regression coefficients in Table A.3 show that only the correlation between the extensive margin of invention patents and R&D investment is positive and highly significant. This implies that the variation in patents outcome at the intensive margin is not well explained by the firm's R&D effort. Because R&D is the fundamen-

Table A.3: Correlation between different margins of R&D and patents

	Invention patents			Utility patents		
	(1)	(2)	(3)	(4)	(5)	(6)
margins	extensive	intensive	both	extensive	intensive	both
extensive	.0402***	-.216	.0732***	.0444***	.165	.121***
	(.003)	(.284)	(.009)	(.003)	(.374)	(.025)
intensive	0.690***	3.214	1.631***	.403***	3.319	1.350**
	(.120)	(2.196)	(.364)	(.109)	(3.060)	(.442)
both	.883***	1.015	1.883***	.780***	2.870	2.302***
	(.110)	(2.497)	(.327)	(.107)	(3.109)	(.476)

Note: all regression contain industry and year fixed effects. Results in columns (1) and (3) are obtained using all the sample; columns (2) and (4) display the results using observations with positive patent applications. Standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

tal source of innovation, we will expect that the variation in patents along the intensive margin will not have a significant impact on the firm's growth. Another important observation from the table is that the intensive margin of R&D is an important explanatory variable for the extensive margin of patents outcome. However, the correlation coefficient becomes much larger if we consider both margins of R&D and use R&D intensity as the indicator for R&D. This suggests that the change of R&D from zero to positive has a much larger marginal impact in generating patents. In the data, the fraction of firms with zero patent is around 30%, while the fraction is increased to be 60% for firms with at least one invention or utility patent. In contrast, the mean value of R&D for firms with at least one patent is only slightly higher than firms without any patent. This confirms that the variation in R&D along the extensive margin is the main driver in explaining the outcome of the patent.

A.3 NLLS estimation

A.3.1 Quadratic $h(\cdot)$

As a robustness check, we also try to parameterize $h(\cdot)$ as a quadratic function. The estimation results are reported in Table A.4. In order to use the first-stage estimates to calculate the value function, we need to check the first-order derivative of $h(\cdot)$ with respect to ϕ_{it}. To ensure that firms have some incentives to invest in R&D and their value functions are bounded, we require this derivative is between 0 and 1. In figure A.1, we show the empirical $\partial h(\cdot)/\partial \phi_t$ against productivity for different forms of h. It is clear that the quadratic form would provide a poor prediction of R&D investment in the data because the non-stationary productivity process will discourage firms from undertaking R&D investment in the infinite-horizon model.

Table A.4: Estimated productivity evolution equation with quadratic form

productivity evolution equation		
ϕ_t	0.826**	(17.13)
ϕ_t^2	0.300**	(15.87)
rd_t	0.00505**	(3.35)
$rd_t \times n_{t+1}$	0.0153**	(3.04)
$rd_t \times b_{t+1}$	0.0142*	(2.85)
β_k	-0.308**	(-25.71)
N	22492	

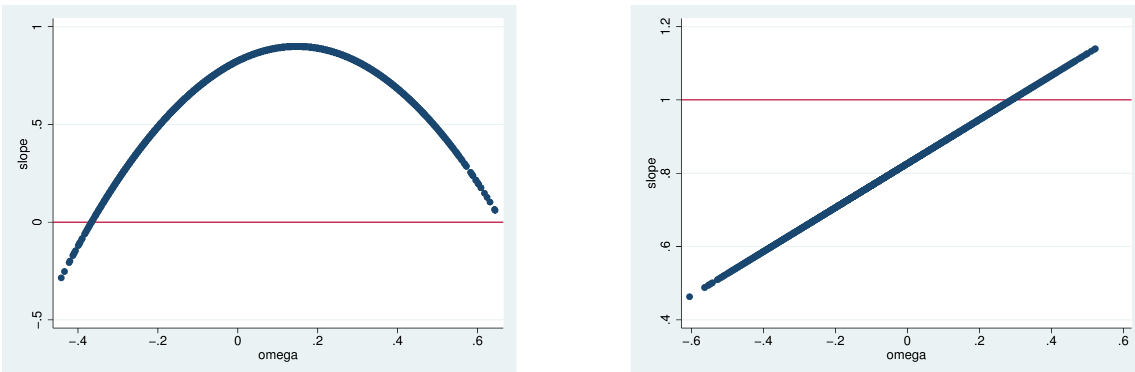
Note:T statistics are in parentheses; * p<0.05, ** p<0.01.

A.3.2 Productivity evolution equation with firm ownership

To account for the ownership effects in the productivity evolution, we extend the benchmark productivity process to be as follows:

Figure A.1: First-order derivative of $h(\cdot)$ with respect to ϕ_{it}

Panel A: Cubic form | Panel B: Quadratic form



Note: In the cubic form, the number of observations that has a slope below zero is 89; in the quadratic equation, the number of observations with a slope greater than one is 517.

$$\phi_{t+1} = \rho_1\phi_t + \rho_2\phi_t^2 + \rho_3\phi_t^3 + \rho_4 rd_t + \rho_5 rd_t \times n_t + \rho_6 rd_t \times b_t \tag{A.1}$$

$$+\rho_7 rd_t \times o_t + \rho_8 rd_t \times n_t \times o_t + \rho_9 rd_t \times b_t \times o_t + \rho_{10} o_t + \varepsilon_{it+1}$$

Under this setting, the marginal effect of R&D on future productivity is given by:

$$\frac{\Delta\phi_{t+1}}{\Delta rd_t} = \rho_4 + \rho_5 n_t + \rho_6 b_t + \rho_7 o_t + \rho_8 n_t \times o_t + \rho_9 b_t \times o_t \tag{A.2}$$

Therefore the firm ownership plays a role in affecting the productivity effects of R&D and patents. The difference in the R&D's productivity effects between SOEs and non-SOEs is given by $\rho_7 + \rho_8 n_t + \rho_9 b_t$. Negative values for ρ_8 and ρ_9 indicate that the R&D investment of SOEs is less effective in stimulating future productivity. Moreover, o_t controls the difference in productivity levels between SOEs and non-SOEs. The estimation results are displayed in Table A.5. In columns (1) to (3), the estimates for ρ_1 to ρ_6 are very similar. In Column (3) in which we control the difference in productivity levels, the estimation results of ρ_7 to ρ_9 show no significant difference in the productivity effects for SOEs and non-SOEs.

Table A.5: Estimated productivity evolution with firm ownership

	(1)	(2)	(3)
$\rho_1(\phi_t)$	0.823**	0.823**	0.824**
	(0.06)	(0.06)	(0.06)
$\rho_2(\phi_t^2)$	0.508**	0.508**	0.508**
	(0.19)	(0.19)	(0.19)
$\rho_3(\phi_t^3)$	-1.133**	-1.150**	-1.150**
	(0.08)	(0.08)	(0.08)
$\rho_4(rd_t)$	0.00436**	0.00378*	0.00373*
	(0.00)	(0.00)	(0.00)
$\rho_5(rd_t \times n_t)$	0.0145**	0.0144**	0.0155**
	(0.01)	(0.01)	(0.01)
$\rho_6(rd_t \times b_t)$	0.0137**	0.0140**	0.0139**
	(0.00)	(0.00)	(0.01)
$\rho_7(rd_t \times o_t)$		0.0217	0.0240
		(0.02)	(0.02)
$\rho_8(rd_t \times n_t \times o_t)$		-0.0709**	-0.0511
		(0.01)	(0.08)
$\rho_9(rd_t \times b_t \times o_t)$			0.00710
			(0.08)
$\rho_{10}(o_t)$			-0.0708**
			(0.01)
N	22492	22492	22492

Note: * p<0.05 ** p<0.01
Standard errors in parentheses

Table A.6: Cost parameters for the case of the Weibull Distribution

		Pharm	Equip.	Elect.	Mach.
	κ_s	5.560	23.040	3.863	0.962
$\theta = 0.5$	κ_m	0.074	0.066	0.011	0.007
	LLF	-3969.4	-332.6	-3011.5	-1508.7
	κ_s	0.368	0.612	0.086	0.043
$\theta = 1.5$	κ_m	0.090	0.094	0.013	0.009
	LLF	-4487.9	-352.4	-3446.1	-1658.5
	κ_s	0.254	0.395	0.049	0.028
$\theta = 2$	κ_m	0.100	0.091	0.014	0.010
	LLF	-4843.0	-390.1	-3790.3	-1824.6

A.4 The Weibull Distribution for R&D Costs

A.4.1 Cost parameters

For the identification reason, we set the shape parameter θ externally, and estimate the model using the Nested Fixed Point algorithm. The estimation results are displayed in Table A.6.

A.4.2 Counterfactual results

For externally chosen θ, the subsidy rate is determined by:

$$1 - \tilde{\delta}_m = \frac{1 - \delta_m}{\Gamma(1 + \frac{1}{\theta})} \tag{A.3}$$

Based on this normalization, we choose δ_s using Equation 27; full results are displayed in Table A.7.

Table A.7: Full counterfactual results using Weibull distribution for R&D costs

θ	type	Pharm.				Equip.				Elect.				Mach.			
		$\uparrow v$	$\uparrow P$	$\frac{\uparrow v}{\mathbf{E}(\tau)}$	$\frac{\uparrow P}{\mathbf{E}(\tau)}$	$\uparrow v$	$\uparrow P$	$\frac{\uparrow v}{\mathbf{E}(\tau)}$	$\frac{\uparrow P}{\mathbf{E}(\tau)}$	$\uparrow v$	$\uparrow P$	$\frac{\uparrow v}{\mathbf{E}(\tau)}$	$\frac{\uparrow P}{\mathbf{E}(\tau)}$	$\uparrow v$	$\uparrow P$	$\frac{\uparrow v}{\mathbf{E}(\tau)}$	$\frac{\uparrow P}{\mathbf{E}(\tau)}$
.5	*conti.*: prop.	.019	.018	.033	.030	.016	.019	.033	.037	.014	.005	.156	.061	.009	.004	.171	.079
	lump.	.097	.024	.721	.173	.080	.023	.726	.199	.019	.001	.967	.058	.011	.001	.978	.060
	start: prop.	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.002
	lump.	.018	.009	.137	.064	.007	.004	.067	.037	.003	.001	.182	.068	.003	.002	.285	.144
1.5	*conti.*: prop.	.097	.100	.164	.147	.082	.116	.142	.186	.023	.000	.257	.000	.016	.000	.257	.000
	lump.	.126	.075	.792	.404	.113	.087	.744	.518	.023	.000	1.000	.000	.016	.000	1.000	.000
	start: prop.	.006	.012	.013	.020	.001	.003	.002	.005	.015	.012	.148	.103	.013	.010	.179	.126
	lump.	.028	.035	.215	.216	.012	.018	.087	.118	.018	.009	.854	.347	.015	.006	.935	.346
2	*conti.*: prop.	.113	.124	.178	.164	.097	.135	.176	.219	.024	.000	.258	.000	.016	.000	.258	.000
	lump.	.135	.101	.787	.482	.116	.102	.790	.612	.024	.000	1.000	.000	.016	.000	1.000	.000
	start: prop.	.016	.032	.031	.051	.003	.007	.005	.013	.022	.008	.197	.063	.016	.004	.205	.042
	lump.	.039	.057	.296	.338	.012	.024	.101	.170	.023	.005	.975	.182	.016	.002	.992	.111

B Proofs

B.1 Profit maximization and revenue equation

The firm's profits maximization problem is

$$\max_{L_{it}, M_{it}} \{P_{it}Q_{it} - P_{Lt}L_{it} - P_{Mt}M_{it}\}$$
$$s.t.\, Q_{it} = P_{it}^{-\sigma}P_t^{\sigma}Q_t$$
$$Q_{it} = \Phi_{it}K_{it}^{\beta_k}L_{it}^{\beta_l}M_{it}^{\beta_m}\exp\left(\beta_a a_{it}\right)$$

Write the revenue as a function of the output, the first-order conditions are:

$$\left(1-\frac{1}{\sigma}\right)\left(P_t^{\sigma}Q_t\right)^{\frac{1}{\sigma}}Q_{it}^{-\frac{1}{\sigma}}\frac{\partial Q_{it}}{\partial L_{it}} = P_{Lt} \tag{A.4}$$

$$\left(1-\frac{1}{\sigma}\right)\left(P_t^{\sigma}Q_t\right)^{\frac{1}{\sigma}}Q_{it}^{-\frac{1}{\sigma}}\frac{\partial Q_{it}}{\partial M_{it}} = P_{Mt} \tag{A.5}$$

where $\frac{\partial Q_{it}}{\partial M_{it}} = \beta_m\frac{Q_{it}}{M_{it}}$ and $\frac{\partial Q_{it}}{\partial L_{it}} = \beta_l\frac{Q_{it}}{L_{it}}$. This implies that

$$M_{it} = \frac{P_{Lt}}{P_{Mt}}\frac{\beta_m}{\beta_l}L_{it} \tag{A.6}$$

Plugging this back to the foc for L_{it}, we obtain

$$\begin{aligned}
P_{Lt} &= \beta_l\left(1-\frac{1}{\sigma}\right)\left(P_t^{\sigma}Q_t\right)^{\frac{1}{\sigma}}\frac{Q_{it}^{1-\frac{1}{\sigma}}}{L_{it}} \\
&= \beta_l\left(1-\frac{1}{\sigma}\right)\left(P_t^{\sigma}Q_t\right)^{\frac{1}{\sigma}}\left(\frac{P_{Lt}}{P_{Mt}}\frac{\beta_m}{\beta_l}\right)^{\frac{\beta_m(\sigma-1)}{\sigma}}\left(\Phi_{it}K_{it}^{\beta_k}e^{\beta_a a_{it}}\right)^{1-\frac{1}{\sigma}}L_{it}^{\frac{(\beta_l+\beta_m)(\sigma-1)}{\sigma}-1}
\end{aligned}$$

$$\Rightarrow L_{it} = \left[\frac{P_{Lt}\left(\frac{P_{Lt}}{P_{Mt}}\frac{\beta_m}{\beta_l}\right)^{\frac{\beta_m(1-\sigma)}{\sigma}}}{\beta_l\left(1-\frac{1}{\sigma}\right)\left(P_t^{\sigma}Q_t\right)^{\frac{1}{\sigma}}}\left(\Phi_{it}K_{it}^{\beta_k}\exp\left(\beta_a a_{it}\right)\right)^{\frac{1-\sigma}{\sigma}}\right]^{\frac{\sigma}{(\beta_l+\beta_m)(\sigma-1)-\sigma}} \tag{A.7}$$

Also note that the foc for L_{it} also implies that the revenue can be expressed as

$$R_{it} = \frac{\sigma P_{Lt} L_{it}}{\beta_l (\sigma - 1)} \tag{A.8}$$

When $\beta_l + \beta_m = 1$, combining these two expressions and take logs yields:

$$r_{it} = \mu_0 + \mu_t + (\sigma - 1)(\beta_k k_{it} + \beta_a a_{it} + \phi_{it}) \tag{A.9}$$

where

$$\mu_0 = (\sigma - 1) \ln \left(\frac{\sigma - 1}{\sigma} \beta_l^{\beta_l} \beta_m^{\beta_m} \right) \tag{A.10}$$

$$\mu_t = (1 - \sigma) \ln \left(P_{Lt}^{\beta_l} P_{Mt}^{\beta_m} \right) + \ln (P_t^{\sigma} Q_t) \tag{A.11}$$

B.2 Effects of Various R&D Subsidies

B.2.1 Proof of Proposition 1

We first show that the following lemma to simplify the algebra.

Lemma. *Define that $\Delta_{it} \equiv EV_1(\phi_{it}, rd_{it-1}) - EV_0(\phi_{it}, rd_{it-1})$The expected firm value under different subsidy programs can be expressed as:*

$$W_{Tit}^{prop} = \pi(\phi_{it}) + \beta EV_0(\phi_{it}, rd_{it-1}) + \int_0^{\Delta_{it}} G(\frac{c}{\delta_T}) dc \tag{A.12}$$

$$W_{Tit}^{lump} = \pi(\phi_{it}) + \beta EV_0(\phi_{it}, rd_{it-1}) + \int_0^{\beta \Delta_{it} + (1-\delta_T)\gamma_{it}} G(c) dc \tag{A.13}$$

Proof. Since the proof for obtaining W_T^{prop} is similar to that for W_T^{lump}, here we only show the proof for W_T^{prop} as below:

$$\begin{aligned}
W_{Tit}^{prop} &= \pi(\phi_{it}) + \int_0^\infty \max\{\beta EV_0(\phi_{it}, rd_{it-1}), \beta EV_1(\phi_{it}, rd_{it-1}) - \delta_T c\} dG(c) \\
&= \pi(\phi_{it}) + \int_0^\infty \max\{\beta EV_0(\phi_{it}, rd_{it-1}), \beta EV_1(\phi_{it}, rd_{it-1}) - c\} dG(c/\delta_T) \\
&= \pi(\phi_{it}) + \int_0^{\beta\Delta_{it}} (\beta EV_1 - c) dG(c/\delta_T) + \int_{\beta\Delta_{it}}^\infty \beta EV_0 dG(c/\delta_T) \\
&= \pi(\phi_{it}) + \beta EV_1 G(\beta\Delta_{it}/\delta_T) - \int_0^{\beta\Delta_{it}} c dG(c/\delta_T) + \beta EV_0(1 - G(\beta\Delta_{it}/\delta_T)) \\
&= \pi(\phi_{it}) + \beta\Delta_{it} G(\beta\Delta_{it}/\delta_T) + \beta EV_0 - cG(c/\delta_T)\Big|_{c=0}^{\beta\Delta_{it}} + \int_0^{\beta\Delta_{it}} G(c/\delta_T) dc \\
&= \pi(\phi_{it}) + \beta EV_0(\phi_{it}, rd_{it-1}) + \int_0^{\beta\Delta_{it}} G(c/\delta_T) dc
\end{aligned}$$

□

From now on, we omit the state variables in the value function and the subscripts *it* as long as no confusion arises. We prove Proposition 1 as follows:

Proof. 1. Using this Lemma, the difference between W_T^{lump} and W_T^{prop} is:

$$\begin{aligned}
\Delta W(\delta_T) &= W_T^{lump} - W_T^{prop} \\
&= \int_0^{\beta\Delta + (1-\delta_T)\gamma} G(c) dc - \int_0^{\beta\Delta} G(c/\delta_T) dc
\end{aligned}$$

Because $G(c) = 1 - \exp(c/\gamma)$, $\Delta W(\delta_T)$ can be expressed as:

$$\begin{aligned}
\Delta W(\delta_T) &= \int_0^{\beta\Delta + (1-\delta_T)\gamma} \left[1 - \exp\left(-\frac{c}{\gamma}\right)\right] dc - \int_0^{\beta\Delta} \left[1 - \exp\left(-\frac{c}{\delta_T \gamma}\right)\right] dc \\
&= \gamma \left[\exp\left(-\frac{\beta\Delta + (1-\delta_T)\gamma}{\gamma}\right) - \delta_T \exp\left(-\frac{\beta\Delta}{\delta_T \gamma}\right)\right]
\end{aligned}$$

This implies that

$$\Delta W(\delta_T) > 0 \Leftrightarrow \exp\left(-\frac{\beta\Delta + (1-\delta_T)\gamma}{\gamma}\right) - \delta_T \exp\left(-\frac{\beta\Delta}{\delta_T\gamma}\right) > 0$$
$$\Leftrightarrow \exp\left(\frac{\beta\Delta}{\gamma}\left(\frac{1}{\delta_T} - 1\right) + \delta_T - 1\right) > \delta_T$$

Consider function $h(\delta_T) = \exp(\delta_T - 1) - \delta_T$, then $h(\delta_T)' = \exp(\delta_T - 1) - 1 \leq 0$ when $\delta_T \in (0,1]$. $h(\delta_T)$ attains its minimum value at $\delta_T = 1$, $h(\delta_T)_{\min} = h(1) = 0$ for $\delta_T \in (0,1]$. This shows that $\exp(\delta_T - 1) > \delta_T$ for $\delta_T \in (0,1)$. Combining that $e^{(\frac{\beta\Delta}{\gamma}(1/\delta_T - 1))} > 1$, we have $\exp\left(\frac{\beta\Delta}{\gamma}\left(\frac{1}{\delta_T} - 1\right) + \delta_T - 1\right) > \delta_T$.

2. Take the difference between P_{Tit}^{lump} and P_{Tit}^{prop}, we obtain that

$$P_{Tit}^{lump} - P_{Tit}^{prop} = \exp\left(-\frac{\beta\Delta EV_{it}}{\delta_T\gamma_{it}}\right) - \exp\left(-\frac{\beta\Delta EV_{it} + (1-\delta_T)\gamma_{it}}{\gamma_{it}}\right)$$
$$= \exp\left(-\frac{\beta\Delta EV_{it}}{\delta_T\gamma_{it}}\right)\left\{1 - \exp\left(\frac{\beta\Delta EV_{it}}{\gamma}\frac{1-\delta_T}{\delta_T} - (1-\delta_T)\right)\right\}.$$

Hence the condition that $P_{Tit}^{lump} - P_{Tit}^{prop} > 0$ is equivalent to $\exp\left[\frac{\beta\Delta EV_{it}}{\gamma}\frac{1-\delta_T}{\delta_T} - (1-\delta_T)\right] < 1$, which implies $\delta_T > \beta\Delta EV_{it}/\gamma_{it}$.

□

B.2.2 Proof of Proposition 2

The proof of result 1 in Proposition 2 is very similar to what we have shown in the proof of the same result in Proposition 1; we only show the proof for result 2 as below.

Proof. The difference of expected firm value between the lump-sum subsidy and proportional subsidy is

$$\Delta W(\delta_T) = \int_0^{\beta\Delta+(1-\delta_T)\bar{C}} G(c)dc - \int_0^{\beta\Delta} G(c/\delta_T)dc$$

(1) $\delta_T > \beta\Delta/\bar{C}$: By change of variable $c_1 = c/\Delta_T$ for the second term of $\Delta W(\delta_T)$, we have

$$\begin{aligned}
\Delta W(\Delta_T) &= \int_0^{\beta\Delta+(1-\delta_T)\bar{C}} G(c)dc - \delta_T \int_0^{\beta\Delta/\delta_T} G(c)dc \\
&> \int_0^{\beta\Delta+(1-\delta_T)\beta\Delta/\delta_T} G(c)dc - \delta_T \int_0^{\beta\Delta/\delta_T} G(c)dc \\
&= (1-\delta_T)\int_0^{\beta\Delta/\delta_T} G(c)dc > 0
\end{aligned}$$

(2) $\delta_T < \beta\Delta/\bar{C}$: Taking the derivative with respect to δ_T yields

$$\begin{aligned}
\Delta W(\delta_T)' &= -\bar{C}G\left(\beta\Delta + (1-\delta_T)\bar{C}\right) + \int_0^{\beta\Delta} \frac{c}{\delta_T^2} g(c/\delta_T)dc \\
&= -\bar{C}G\left(\beta\Delta + (1-\delta_T)\bar{C}\right) + \int_0^{\frac{\beta\Delta}{\delta_T}} cg(c)dc,
\end{aligned}$$

where the second equation is by the change of variable. Note that

$$\lim_{\delta_T\to 0} \Delta W(\delta_T)' = -\bar{C}G(\beta\Delta+\bar{C}) + \int_0^{\infty} cg(c)dc = \bar{C}(1-G(\beta\Delta+\bar{C})) > .$$

Moreover, observe that the second-order derivative

$$\Delta W(\delta_T)'' = \bar{C}^2 g(\beta\Delta + (1-\delta_T)\bar{C}) - \frac{(\beta\Delta)^2}{\delta_T^3} g(\frac{\beta\Delta}{\delta_T})$$

is an increasing function of δ_T whenever $g'(c) < 0$. This implies that $\Delta W(\delta_T)'' > \Delta W(\delta_T = \varepsilon)''$ for $0 < \varepsilon < \delta_T$. And that

$$\lim_{\varepsilon \to 0^+} \Delta W(\varepsilon)'' = \bar{C}^2 g(\beta\Delta + \bar{C}) - (\beta\Delta)^{-1} \lim_{c\to\infty} \{c^3 g(c)\} = \bar{C}^2 g(\beta\Delta + \bar{C}) > 0$$

Because $\Delta W(\delta_T)''$ is right-continuous at 0, we have $\Delta W(\delta_T)'' > \lim_{\varepsilon\to 0^+} \Delta W(\varepsilon)'' > 0$. This further implies that $\Delta W(\delta_T)' > 0$ and $\Delta W(\delta_T) > \lim_{\delta_T \to 0^+ \Delta W(\delta_T)}$. Note that

$$\begin{aligned}
\lim_{\delta_T\to 0^+} \Delta W(\delta_T) &= \int_0^{\beta\Delta+\bar{C}} G(c)dc - \lim_{\delta_T\to 0^+} \int_0^{\beta\Delta} G(c/\delta_T)dc \\
&= \int_0^{\beta\Delta+\bar{C}} G(c)dc - \beta\Delta
\end{aligned}$$

Consider function $h(x) = \int_0^{x+\bar{C}} G(c)dc - x$, and its first-order derivative $h(x)' = G(x+\bar{C}) - 1 \leq 0$, hence $h(x)$ is weakly decreasing in x. This indicates that

$$\begin{aligned}
\Delta W(\delta_T) &> \lim_{\Delta\to\infty} \left\{ \int_0^{\beta\Delta+\bar{C}} G(c)dc - \beta\Delta \right\} \\
&= \lim_{\Delta\to\infty} \left\{ (\beta\Delta+\bar{C})G(\beta\Delta+\bar{C}) - \int_0^{\beta\Delta+\bar{C}} cg(c)dc - \beta\Delta \right\} \\
&= \lim_{\Delta\to\infty} \left\{ -\beta\Delta\left[1 - G(\beta\Delta+\bar{C})\right] + \bar{C}G(\beta\Delta+\bar{C}) - \int_0^{\beta\Delta+\bar{C}} cg(c)dc \right\} \\
&= \lim_{\Delta\to\infty} \left\{ -\beta\Delta\left[1 - G(\beta\Delta+\bar{C})\right] \right\}
\end{aligned}$$

Let $x = \beta\Delta$. Applying the L'Hospital's rule to evaluate the limit, we obtain that

$$\begin{aligned}
\lim_{x\to\infty} \left\{ x\left[1 - G(x+\bar{C})\right] \right\} &= \lim_{x\to\infty} \left\{ x^2 g(x+\bar{C}) \right\} \\
&= \lim_{x\to\infty} \left\{ \frac{x^2}{(x+\bar{C})^3} (x+\bar{C})^3 g(x+\bar{C}) \right\} \\
&= \lim_{x\to\infty} \left\{ (x+\bar{C})^3 g(x+\bar{C}) \right\} = 0
\end{aligned}$$

Therefore $\Delta W(\delta_T) > 0$ when $\lim_{x\to\infty} \left\{ t^3 g(t) \right\} > 0$.

□

Online Appendix (Not for Publication)

A The Model Extension with Continuous R&D Investment

In this subsection, we first briefly lay out an extended dynamic model with R&D investment and patents. The basic structure of the model is similar to that considered in Aw et al. (2011); Doraszelski and Jaumandreu (2013); Peters et al. (2016, 2017), with the exception that both R&D and patents play a role in shifting the future productivity.

Production and Profits A firm has a Cobb-Douglas production function

$$Q_{it} = \Phi_{it} K_{it}^{\beta_k} L_{it}^{\beta_l} M_{it}^{\beta_m} \exp\left(\beta_a a_{it}\right) \tag{A.14}$$

where Q_{it} is the physical output of firm i in period t, Φ_{it} is the total factor productivity, K_{it} is the capital, L_{it} is the labor, M_{it} is the material, a_{it} is the firm's age. Consider a well-behaved inverse demand equation

$$p_{it} = D(Q_{it}) \tag{A.15}$$

where p_{it} is the output price. To simplify the analysis, we assume that a firm treats capital and productivity as predetermined when choosing labor and materials in each period. Let $\Pi(\phi_{it}, \mathbf{S}_{it})$ be the optimal profits, and $R(\phi_{it}, \mathbf{S}_{it})$ be the revenue, where $\phi_{it} = \ln(\Phi_{it})$, $\mathbf{S}_{it} = (K_{it}, P_{Lit}, P_{Mit}, a_{it})$ is a vector of exogenous state variables. The cost minimization implies that

$$\Pi(\phi_{it}, \mathbf{S}_{it}) = (1 - \frac{\beta_l + \beta_m}{\theta_{it}}) R(\phi_{it}, \mathbf{S}_{it}) \tag{A.16}$$

where θ_{it} is the markup. Note that when $\beta_l + \beta_m = 1$, the production function is of constant return to scale in terms of L_{it} and M_{it}.

Productivity evolution The firm's productivity ϕ_{it} is unobserved by the econometrician. R&D investment and patenting enter the Markov process governing the productivity evolution. In particular, the dynamics of productivity is given by

$$\phi_{it+1} = h(\phi_{it}, d_{it}, \mathbf{o}_{it+1}) + \varepsilon_{it+1} \tag{A.17}$$

where d_{it} represents the R&D investment and $\mathbf{o}_{it+1}$ is a vector summarizing the innovation outcome next period, and ε_{it+1} is an iid shock with a mean-zero normal distribution. Considering different types of innovation output, $\mathbf{o}_{it}$ can be a vector of process innovation and product innovation measured by the patents or other observed indicators. The marginal effects of R&D investment and innovation output are captured by three partial derivatives $\partial h/\partial d_{it}$, $\partial h/\partial \mathbf{o}_{it+1}$, and a cross derivative $\partial^2 h/\partial d_{it}\partial \mathbf{o}_{it+1}$. Because R&D is the fundamental source of productivity change, we impose that $\partial h(\phi_{it}, 0, \mathbf{o}_{it+1})/\partial \mathbf{o}_{it+1} = \mathbf{0}$. This implies that without R&D investment, we should expect no endogenous productivity growth, though we can see productivity growth through the channel of exogenous shocks. The cross-derivative $\partial^2 h/\partial d_{it}\partial \mathbf{o}_{it+1}$ deserves some discussion. When positive, it means that the stimulating impact of R&D on productivity is strengthened through the patenting. This indicates that the patenting system help firms protect their inventions. When negative, we anticipate that the effect of knowledge spillovers dominates so the firm's productivity improves less by patenting. This may be due to the weak patenting system.

Following CDM, we assume that patents are random variables of which the distribution is determined by R&D investment. This assumption greatly simplifies the analysis by only considering the R&D investment choice. Specifically, patents outcome in next period is assumed to be a distribution depending on past R&D. The distribution of $\mathbf{o}_{it+1}$ is given by $Pr(\mathbf{o}_{it} \leq \mathbf{o}) = G(\mathbf{o}; d_{it})$.[30] This layer of uncertainty is similar to that considered in PRVF. Note the Markovian property implies that the conditional expectation of future productivity is

$$\mathbf{E}(\phi_{it+1}|\phi_{it}, d_{it}) = \int h(\phi_{it}, d_{it}, \mathbf{o}) dG(\mathbf{o}; d_{it})$$

Therefore R&D investment can influence future productivity through affecting $h(\phi_{it}, d_{it}, \mathbf{o})$ and the distribution of innovation output $G(\mathbf{o}; d_{it})$. This allows me to decompose the impact of R&D into patenting and non-patenting channels.

Recursive formulation To consider a general setting, denote $C(d_{it}, \mathbf{X}_{it})$ as the variable costs of R&D investment. Here $\mathbf{X}_{it} = (\mathbf{S}_{it}, \mathbf{Z}_{it})$, $\mathbf{Z}_{it}$ is the additional exogenous states that influence the costs of R&D investment.[31] In addition, there is a fixed cost of R&D investment, denoted as $f(\mathbf{X}_{it})$. With these fixed costs, the model can capture the innovation choice at an extensive margin. Note that we allow the exogenous state variables to affect the costs of R&D. Omitting the subscripts, the firm's dynamic programming problem can be written in a recursive formulation:

$$V(\phi, \mathbf{X}) = \max_{d} \left\{ V^{0}(\phi, \mathbf{X}), V^{d}(\phi, \mathbf{X}) \right\} \tag{A.18}$$

[30] In the reduced-form analysis, this process is usually estimated using count data models. See Hall and Hayashi (1989); Hall et al. (2010).

[31] For example, $\mathbf{Z}_{it}$ may contain past R&D investment decisions so the R&D costs also include adjustment costs.

where the value functions for different choices of d are

$$V^0(\phi) = \Pi(\phi, \mathbf{X}) + \beta \mathbf{E}\left[V(\phi', \mathbf{X}')|d=0\right], \tag{A.19}$$

$$V^d(\phi) = \max_d \left\{\Pi(\phi, \mathbf{X}) - C(d, \mathbf{X}) - f(\mathbf{X}) + \beta \mathbf{E}\left[V(\phi', \mathbf{X}')|d\right]\right\}, \tag{A.20}$$

where β is the discounting factor. We assume that firms have perfect foresight for the exogenous state variables. This allows us to calculate the expected firm value as

$$\mathbf{E}\left[V(\phi', \mathbf{X}')|d\right] = \int\int V(h(\phi, d, \mathbf{o}) + \varepsilon', \mathbf{X}')dG(\mathbf{o}; d)dF(\varepsilon') \tag{A.21}$$

where $F(\cdot)$ is the distribution of ε'. A stochastic equilibrium of the model is a decision rule $d(\phi, \mathbf{X}) \geq 0$ such that the recursive problem (A.18) is solved.

B Model fit

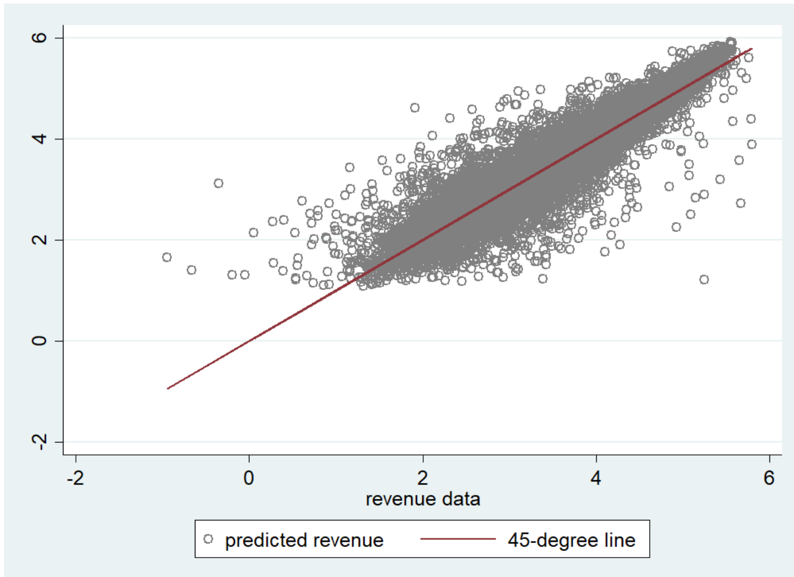
B.1 Predicted revenues

In Figure A.2, we present a scatter plot to check the relationship between the model predicted revenue and the revenue information in the data. We can see that the predicted revenues concentrates around the 45-degree line, which indicates the revenue equation fits the data well.

B.2 Pooled probability of investing in R&D

Given current state, we can solve for the probability of undertaking R&D,$Pr(d=1|\phi, d_{-1}, \mathbf{S})$ using equation (18). Therefore the aggregate hazad function for R&D choice can be cal-

Figure A.2: Model fit for the revenue data



Note: sales are in logs of revenues in 100,000 USD.

culated as:

$$\mathscr{H} = \frac{1}{NT} \sum_{i}^{N} \sum_{t}^{T} Pr(d_{it} = 1 | \phi_{it}, d_{it-1}, \mathbf{S}_{it}) \tag{A.22}$$

On the other hand, in the data the probability of investing in R&D is given as

$$\tilde{\mathscr{H}} = \frac{1}{NT} \sum_{i}^{N} \sum_{t}^{T} d_{it} \tag{A.23}$$

We calculate this Hazard function for each sector. The results are displayed in Table A.8. Overall, the estimated model predicts the probability of innovation similar to that exhibited in the data. The model-predicted pooled probability of choosing R&D is slightly higher, but the difference from the data is around 0.04. Implying that the estimated model captures the decision of innovating reasonably well in terms of the probability of choosing innovation for the pooled sample.

Table A.8: Pooled probability of investing in R&D

Prob. of innovation	Pharmaceutical	Equipment	Electronics	Machinery
model	0.300	0.256	0.169	0.296
data	0.379	0.349	0.222	0.389

B.2.1 Transition dynamics of R&D choice

The transition probability characterizes the dynamics of transition for past R&D choice to current R&D decision. Let $rd_{-1} \in \{0,1\}$ be the R&D status in previous year and $rd \in \{0,1\}$ be the current R&D decision, then the transition probabilities are $Q(rd|rd_{-1})$. We calculate these probabilities in data using a formula as follows:

$$\tilde{Q}(rd = d'|rd_{-1} = d) = \frac{\sum_i^N \sum_t^T \mathbb{I}\{rd_{it} = d', rd_{it-1} = d\}}{\sum_i^N \sum_t^T \mathbb{I}\{rd_{it-1} = d\}} \tag{A.24}$$

where $d, d' \in \{0,1\}$. The model predicts the transition probability for each given state as:

$$Q(rd = d'|rd_{-1} = d) = \frac{1}{NT} \sum_i^N \sum_t^T Pr(rd_{it} = d'|rd_{it-1} = d, k_{it}) \tag{A.25}$$

where $Pr(rd_{it} = d'|rd_{it-1} = d, k_{it})$ can be obtained using Equation (18). We calculate these two transition probabilities for each sector and present it in Table A.9. First, the general patterns of the relative magnitudes of transition probabilities in the data and that predicted by the model are quite similar. In all four industries, the probability of staying in the previous state is much higher than transiting to a new state. In other words, $Q(0,0)$ is larger than $Q(0,1)$, and $Q(1,1)$ is greater than $Q(1,0)$. Second, the probabilities predicted by the estimated model are very close to that observed in the data. These results suggest that the estimated model captures the transition dynamics in the R&D activities well.

Table A.9: Transition dynamics of R&D choice

sectors	source	$Q(0,0)$	$Q(0,1)$	$Q(1,0)$	$Q(1,1)$
Pharmaceutical	data	0.851	0.149	0.233	0.767
	model	0.890	0.110	0.379	0.621
Equipment	data	0.915	0.085	0.183	0.817
	model	0.946	0.054	0.387	0.614
Electronics	data	0.930	0.070	0.257	0.743
	model	0.950	0.050	0.426	0.574
Machinery	data	0.878	0.122	0.200	0.800
	model	0.916	0.084	0.375	0.625

Note: $Q(d',d) = \Pr(rd = d'|rd_{-1} = d)$

C Computation

C.1 Discretizing non-linear Markov process

The productivity process in the model is non-linear and deserves special treatment. To improve the accuracy of the approximation, we refer to Farmer and Toda (2017) to discretize the non-linear Markov process for productivity by matching low order moments of the conditional distribution using maximum entropy.

C.2 Nested fixed-point algorithm

Profit function We normalize the productivity with the constant ψ_0 in the empirical model. From the estimation equation, we can write the profit as:

$$\hat{r}_{it}\left(\hat{\phi}_{it}\right) = \hat{\psi}_0 + \hat{\psi}_t + \left(1+\hat{\theta}_j\right)\hat{\rho}_0 + \left(1+\hat{\theta}_j\right)\hat{\beta}_k k_{it} + \left(1+\hat{\theta}_j\right)\hat{\beta}_a a_{it} - \left(1+\hat{\theta}_j\right)\hat{\phi}_{it} \tag{A.26}$$

After the normalization, it follows that the firm's profits can be calculated as

$$\pi_{it}\left(\hat{\phi}_{it}\right) = -\frac{1}{\hat{\theta}_j}\exp\left(\hat{r}_{it}\left(\hat{\phi}_{it}\right)\right) \tag{A.27}$$

Choose the grid points We compute by industry. Some coefficients are specific to each industry: ψ_0, ρ_0, and θ_j. Note that we discretize the productivity into 200 grid points, and age into 4 groups. To implement the estimation, we use the trapezoid method to discretize the capital space in evenly distributed 100 points. Therefore, we are encountered with $100 \times 4 \times 4 = 1600$ types of firms. For each type of firm, we solve the value function for $100 \times 2 = 200$ states. In the end, we solve $1600 \times 200 = 320,000$ value functions. We compute the value function by industry.

For the extended model with firm ownership, we add two points to specify the type of ownership. This requires us to solve $1600 \times 200 \times 2 = 640,000$ value functions. The value function is also computed for each industry.

C.3 Inner and outer loops

Inner loop: Value function iteration Given $(\phi_{it}, k_{it}, a_{it})$, we use V_d to denote $V(\phi_{it}, rd_{it-1} = d; k_{it}, a_{it})$. We also define $\gamma_{it}^d \equiv \gamma(rd_{it-1} = d, k_{it}; \kappa_m, \kappa_s)$, for $d \in \{0,1\}$, and $\kappa \equiv (\kappa_m, \kappa_s)$ be the parameter to be estimated. Under this definition, we have $\gamma_{it}^0 = \kappa_s k_{it}$ and $\gamma_{it}^1 = \kappa_m k_{it}$.[32] Employing

[32] We only illustrate the benchmark case here. for the extended model with firm ownership, we only need to add parameters κ_{mo} and κ_{so} into the vector of parameters such that $\gamma_{it}^d = \gamma(rd_{it-1} = d, k_{it}; \kappa_m, \kappa_s, \kappa_{mo}, \kappa_{so})$ and $\kappa \equiv (\kappa_m, \kappa_s, \kappa_{mo}, \kappa_{so})$. There we have $\gamma_{it}^0 = \kappa_m k_{it} + \kappa_{mo} o_{it}$ and $\gamma_{it}^1 = \kappa_s k_{it} + \kappa_{so} o_{it}$. The computation algorithm for the benchmark model still applies.

the exponential distribution, we can express the value function as

$$
\begin{aligned}
V_d &= \pi_{it}(\phi_{it}) + \beta \int_0^{\Delta EV} (EV_1 - c)\, dG(c) + \beta \int_{\Delta EV}^{\infty} EV_0 dG(c) \\
&= \pi_{it}(\phi_{it}) + \beta EV_1 \left(1 - e^{-\frac{\Delta \mathbb{E}V}{\gamma_{it}^d}}\right) + \beta \left(\Delta EV + \gamma_{it}^d\right) e^{-\frac{\Delta \mathbb{E}V}{\gamma_{it}^d}} \\
&- \beta \gamma_{it}^d + \beta EV_0 e^{-\frac{\Delta \mathbb{E}V}{\gamma_{it}^d}}
\end{aligned}
$$

where $\Delta EV = EV_1 - EV_0$. Therefore the expression of V_d can be simplified as

$$
V_d = \pi_{it}(\phi_{it}) + \beta \gamma_{it}^d \left(\exp\left(-\frac{\Delta EV}{\gamma_{it}^d}\right) - 1\right) + \beta EV_1, \text{ for } d \in \{0, 1\} \tag{A.28}
$$

In computing the value functions, V_d is a 200 by 1 vector given capital, age, and industry. Let $p_{mn} = \Pr(n_{t+1} = m, b_{t+1} = n | rd_t = 1)$, for $m, n \in \{0, 1\}$, and further denote P_{mn} as the corresponding transition matrix of the productivity and P_0 as the transition matrix of productivity when $rd_t = 0$. then Equation (A.28) can be transformed as

$$
V_1 = \pi(\phi) - \beta \gamma^1 \left(1 - \exp\left(-\frac{\Delta EV}{\gamma^1}\right)\right) + \beta \left(\sum_m \sum_n p_{mn} P_{mn}\right) V_1 \tag{A.29}
$$

$$
V_0 = \pi(\phi) - \beta \gamma^0 \left(1 - \exp\left(-\frac{\Delta EV}{\gamma^0}\right)\right) + \beta \left(\sum_m \sum_n p_{mn} P_{mn}\right) V_1 \tag{A.30}
$$

Denote $P_1 = (\sum_m \sum_n p_{mn} P_{mn})$, then

$$
V_1 = (I - \beta P_1)^{-1} \left[\pi(\phi) - \beta \gamma^1 \left(1 - \exp\left(-\frac{\Delta EV}{\gamma^1}\right)\right)\right]
$$

because $\Delta EV = P_1 V_1 - P_0 V_0$, it follows that:

$$\Delta EV = (I - \beta P_0) P_1 (I - \beta P_1)^{-1} \left[\pi(\phi) - \beta \gamma^1 \left(1 - \exp\left(-\frac{\Delta EV}{\gamma^1} \right) \right) \right] \tag{A.31}$$
$$- P_0 \left[\pi(\phi) - \beta \gamma^0 \left(1 - \exp\left(-\frac{\Delta EV}{\gamma^0} \right) \right) \right]$$

We use equation (A.31) to solve for ΔEV and then we use equations (A.29) and (A.30) to solve for V_1 and V_0. Use T_κ as the linear operator applied to ΔEV, it is easy to show that T_κ is a contraction mapping. Then ΔEV is a fixed point such that

$$T_\kappa(\Delta EV) = \Delta EV.$$

Now we are in the position to use Newton-Kantorovich iterations. First note that the Frechét derivative of T_κ with respect to ΔEV is:

$$\begin{aligned} T'_\kappa &= \frac{\partial T_\kappa(\Delta EV)}{\partial \Delta EV} \\ &= \beta(\beta P_0 - I) P_1 (I - \beta P_1)^{-1} \exp\left[\operatorname{diag}\left\{ -\frac{\Delta EV_i}{\gamma^1} \right\} \right] \\ &+ \beta P_0 \exp\left[\operatorname{diag}\left\{ -\frac{\Delta EV_i}{\gamma^0} \right\} \right] \end{aligned} \tag{A.32}$$

where

$$\operatorname{diag}\left\{ -\frac{\Delta EV_i}{\gamma^d} \right\} = \begin{bmatrix} -\frac{\Delta EV_1}{\gamma^d} & 0 & \cdots & 0 \\ 0 & -\frac{\Delta EV_2}{\gamma^d} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\frac{\Delta EV_n}{\gamma^d} \end{bmatrix}$$

Using the invertibility of $[I - T'_\kappa]$ the nth iteration in the Newton-Kantorovich algorithm is

$$\Delta EV_{n+1} = \Delta EV_n - \left[I - T'_\kappa\right]^{-1} (I - T_\kappa)(\Delta EV_n)$$

We set the tolerance as e^{-6}; the iteration stops when $\|\Delta EV_{n+1} - \Delta EV_n\| \leq e^{-6}$.

Outer loop: BHHH optimization algorithm The outer loop solves the following problem:

$$\max_{\kappa_s, \kappa_m} \sum_i \sum_t l_{it}(\kappa, \Delta EV_{it})$$

where

$$\begin{aligned} l_{it}(\kappa, \Delta EV_{it}) &= \log\{rd_{it} \Pr(rd_{it} = 1|\kappa, rd_{it-1}) + (1 - rd_{it}) \Pr(rd_{it} = 0|\kappa, rd_{it-1})\} \\ &= \log\{(2rd_{it} - 1) \Pr(rd_{it} = 1|\kappa, rd_{it-1}) + 1 - rd_{it}\} \end{aligned} \tag{A.33}$$

where $\Delta EV_{it} = \Delta EV(\phi_{it})$ and[33]

$$\begin{aligned} \Pr(rd_{it} = 0|\kappa, rd_{it-1}) &= 1 - \Pr(rd_{it} = 1|\kappa, rd_{it-1}) \\ &= \exp\left\{\frac{-\beta \Delta EV(\phi_{it})}{\kappa_m rd_{it-1} k_{it} + \kappa_s (1 - rd_{it-1}) k_{it}}\right\}. \end{aligned} \tag{A.34}$$

The basic parameter iteration under the BHHH algorithm is:

$$\kappa^{n+1} = \kappa^n + \lambda \underbrace{\left[\sum_{i,t} \left(\frac{\partial l_{it}(\kappa^n, \Delta EV_{it})}{\partial \kappa^n}\right)\left(\frac{\partial l_{it}(\kappa^n, \Delta EV_{it})}{\partial (\kappa^n)'}\right)\right]^{-1} \left(\sum_{i,t} \frac{\partial l_{it}(\kappa^n, \Delta)}{\partial \kappa^n}\right)}_{\equiv D(\kappa^n)}$$

[33] In the extended model with firm ownership, this equation becomes

$$\Pr(rd_{it} = 0|\kappa, rd_{it-1}) = \exp\left\{\frac{-\beta \Delta EV(\phi_{it})}{\kappa_m rd_{it-1} k_{it} + \kappa_s (1 - rd_{it-1}) k_{it} + \kappa_{so} (1 - rd_{it-1}) o_{it} + \kappa_{mo} rd_{it-1} o_{it}}\right\}$$

From (A.33) we know that

$$\frac{\partial l_{it}\left(\kappa^n,\Delta EV_{it}\right)}{\partial\left(\kappa^n\right)'}=w_{it}\left[\frac{\partial Pr\left(rd_{it}=1|\kappa,rd_{it-1}\right)}{\partial\kappa_s^n},\frac{\partial Pr\left(rd_{it}=1|\kappa,rd_{it-1}\right)}{\partial\kappa_m^n}\right] \tag{A.35}$$

where

$$w_{it}=\frac{\left(2rd_{it}-1\right)}{\left(2rd_{it}-1\right)\Pr\left(rd_{it}=1|\kappa,rd_{it-1}\right)+1-rd_{it}}$$

$$\frac{\partial Pr\left(rd_{it}=1|\kappa,rd_{it-1}\right)}{\partial\kappa_s^n}=\beta\frac{\frac{\partial\Delta EV_{it}}{\partial\kappa_s^n}\gamma_{it}^{rd_{it-1}}-\left(1-rd_{it-1}\right)k_{it}\Delta EV_{it}}{\left(\gamma_{it}^{rd_{it-1}}\right)^2\exp\left(\frac{\beta\Delta EV_{it}}{\gamma_{it}^{rd_{it-1}}}\right)}$$

$$\frac{\partial Pr\left(rd_{it}=1|\kappa,rd_{it-1}\right)}{\partial\kappa_m^n}=\beta\frac{\frac{\partial\Delta EV_{it}}{\partial\kappa_n^m}\gamma_{it}^{rd_{it-1}}-rd_{it-1}k_{it}\Delta EV_{it}}{\left(\gamma_{it}^{rd_{it-1}}\right)^2\exp\left(\frac{\beta\Delta EV_{it}}{\gamma_{it}^{rd_{it-1}}}\right)}$$

where $\frac{\partial\Delta EV_{it}}{\partial\kappa_s}$ ($\frac{\partial\Delta EV_{it}}{\partial\kappa_m}$) is the element in the 1st (2nd) column such that the corresponding productivity in the row is ϕ_{it}. To finish the nested fixed-point algorithm, we need to compute the derivatives of the expected value function, $\partial\Delta EV/\partial\gamma$. Applying the implicit theorem to $T_\kappa\left(\Delta EV\right)=\Delta EV$, we get

$$\frac{\partial\Delta EV}{\partial\kappa}=\left[I-T_\kappa'\right]^{-1}\frac{\partial T_\kappa\left(\Delta EV\right)}{\partial\kappa}$$

From (A.31), we know that

$$\frac{\partial T_\kappa\left(\Delta EV\right)}{\partial\kappa}=\left[\frac{\partial T_\kappa(\Delta EV)}{\partial\kappa_s},\frac{\partial T_\kappa(\Delta EV)}{\partial\kappa_m}\right]$$

where

$$\frac{\partial T_\kappa(\Delta EV)}{\partial \kappa_m} = \beta k(\beta P_0 - I)P_1(I-\beta P_1)^{-1}\left[1-\exp\left(\frac{-\Delta EV}{\gamma^1}\right)-\frac{\Delta EV}{\gamma^1}\odot\exp\left(\frac{-\Delta EV}{\gamma^1}\right)\right]$$
$$\frac{\partial T_\kappa(\Delta EV)}{\partial \kappa_s} = \beta k P_0\left[1-\exp\left(\frac{-\Delta EV}{\gamma^0}\right)-\frac{\Delta EV}{\gamma^0}\odot\exp\left(\frac{-\Delta EV}{\gamma^0}\right)\right]$$

are both 200-by-1 vectors and k is the exogenous state variable: capital stock. We use $\odot$ to denote the element-wise product. To determine the step size λ, we use secant iteration to find the solution to $\partial f(\lambda)/\partial\kappa = 0$, where $f(\lambda)\equiv L(\kappa+\lambda D(\kappa))$. The iteration is given as:

$$\lambda^{m+1} = \lambda^m - \frac{\left(\lambda^m - \lambda^{m-1}\right)f'(\lambda^m)}{f'(\lambda^m)-f'(\lambda^{m-1})} \tag{A.36}$$

where

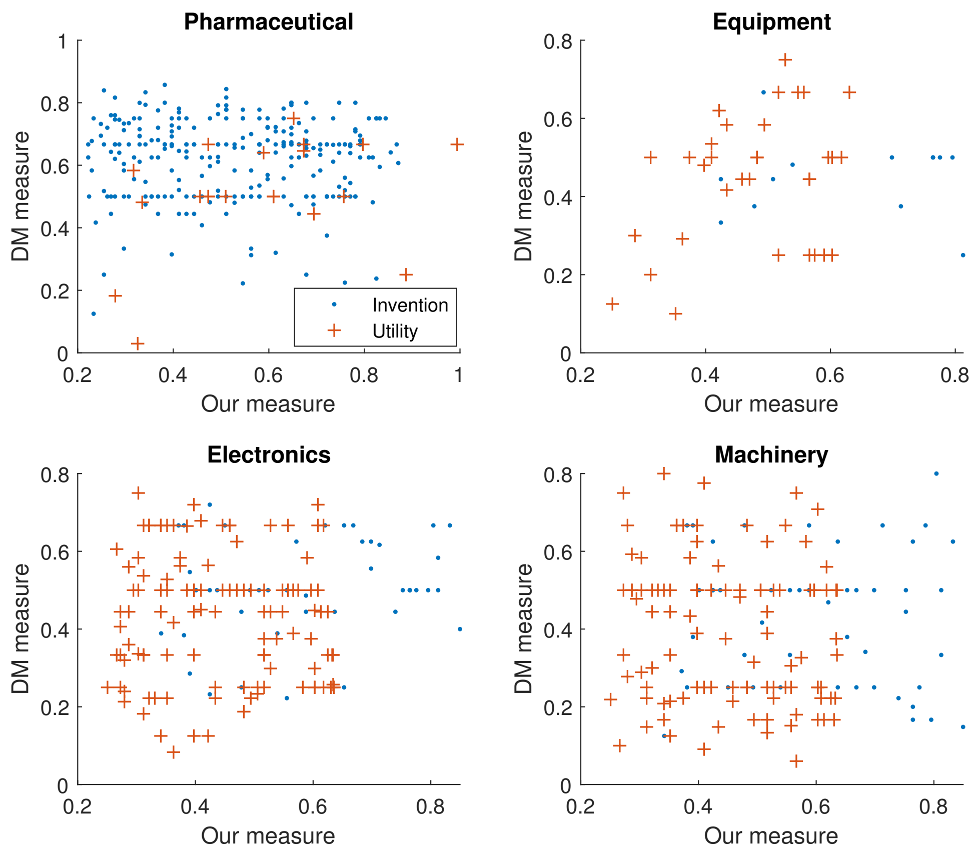
$$f'(\lambda^m) = \sum_{i,t}\frac{\partial l_{it}\left(\kappa+\lambda^m D(\kappa_n),\Delta EV_{it}\right)}{\partial\kappa'}D(\kappa_n) \tag{A.37}$$

This iteration determines the optimal step size λ^*. Finally, the iteration stops when $\|\kappa^{n+1}-\kappa^n\|\le 1e^{-6}$.

C.4 Alternative indicator for patent quality

A widely used indicator for patent quality is patent citations. However, China's patent data are lack of patent citations. Dang and Motohashi (2015) propose to use the measure of knowledge breath as a proxy for the quality of patents. It is questionable whether this measure is a good indicator for the quality of patents. In Figure A.3, we display the correlation between the estimated value of patents to the indicator based on Dang and Motohashi (2015). Interestingly, we find barely no correlation between these two indicators. This may suggest that the knowledge breadth measure is not a good indicator for representing the quality of the patents. At least, it does not reflect the private value of patents measured as increasing the firm's value.

Figure A.3: Estimated patent value and knowledge breadth-based measure



Note: DM measure for the patent quality is based on Dang and Motohashi (2015), which uses the breadth of knowledge for each patent claim. Our measure for the patent quality is based on the change in the expected firm value.

References

Aw, B. Y., M. J. Roberts, and D. Y. Xu (2011). R&D investment, exporting, and productivity dynamics. *American Economic Review 101*(4), 1312–44. 1, 3.3, 6.1, A

Brandt, L., J. V. Biesebroeck, and Y. Zhang (2012). Creative accounting or creative destruction? firm-level productivity growth in Chinese manufacturing. *Journal of Development Economics 97*(2), 339 – 351. 3.1

Chen, Z., Z. Liu, J. C. S. Serrato, and D. Y. Xu (2017). Notching R&D investment with corporate income tax cuts in China. *mimeo, Duke University*. 3.3

Chen, Z. and J. Zhang (2019, aug). Types of patents and driving forces behind the patent growth in China. *Economic Modelling 80*, 294–302. 13

Chen, Z., J. Zhang, and W. Zheng (2017). Imports and innovation: Evidence from Chinese firms. *European Economic Review, 94*, 205–220. 4

Crépon, B., E. Duguet, and J. Mairesse (1998). Research, innovation and productivity: An econometric analysis at the firm level. *Economics of Innovation and new Technology 7*(2), 115–158. 1

Dang, J. and K. Motohashi (2015). Patent statistics: A good indicator for innovation in China? Patent subsidy program impacts on patent quality. *China Economic Review 35*, 137–155. 20, C.4, A.3

Doraszelski, U. and J. Jaumandreu (2013). R&D and productivity: Estimating endogenous productivity. *The Review of Economic Studies 80*(4), 1338–1383. 1, A

Eberhardt, M., C. Helmers, and Z. Yu (2016). What can explain the chinese patent explosion? *Oxford Economic Papers 69*(1), 239–262. 13

Farmer, L. E. and A. A. Toda (2017). Discretizing nonlinear, non-gaussian markov processes with exact conditional moments. *Quantitative Economics 8*(2), 651–683. 4.3, C.1

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *The Bell Journal of Economics*, 92–116. 1

Hall, B. and F. Hayashi (1989). Research and development as an investment. Technical report, National Bureau of Economic Research. 1, 30

Hall, B. H., Z. Griliches, and J. A. Hausman (1986). Patents and R and D: Is there a lag? *International Economic Review 27*(2), 265–283. 3.3

Hall, B. H., J. Mairesse, and P. Mohnen (2010). Measuring the returns to R&D. *Handbook of the Economics of Innovation 2*, 1033–1082. 1, 3.3, 30

He, Z. L., T. W. Tong, Y. Zhang, and W. He (2018). A database linking Chinese patents to China's census firms. *Scientific Data 5*, 1–16. 3.1, 11

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing TFP in China and India. *The Quarterly Journal of Economics 124*(4), 1403–1448. 3.1

Hu, A. G. and G. H. Jefferson (2009). A great wall of patents: What is behind China's recent patent explosion? *Journal of Development Economics 90*(1), 57–68. 4, 13, 4.2

Hu, A. G., P. Zhang, and L. Zhao (2017). China as number one? evidence from China's most recent patenting surge. *Journal of Development Economics 124*, 107–119. 4, 13

Klette, T. J. (1996). R&D, scope economies, and plant performance. *The RAND Journal of Economics*, 502–522. 1

Lach, S. and M. Schankerman (1989). Dynamics of R&D and Investment in the Scientific Sector. *Journal of Political Economy 97*(4), 880–904. 3.3

Mairesse, J., P. Mohnen, and E. Kremp (2005). The importance of R&D and innovation for productivity: A reexamination in light of the French innovation survey. *Annales d'Economie et de Statistique*, 487–527. 1, 3.3

Pakes, A. (1986). Patents as options: Some estimates of the value of holding European patent stocks. *Econometrica 54*(4), 755–784. 1, 2.2

Peters, B., M. J. Roberts, and V. A. Vuong (2016). Dynamic R&D choice and the impact of the firm's financial strength. *Economics of Innovation and New Technology*, 1–16. 18, 29, A

Peters, B., M. J. Roberts, V. A. Vuong, and H. Fryges (2017). Estimating dynamic R&D choice: an analysis of costs and long-run benefits. *The RAND Journal of Economics 48*(2), 409–437. 1, 3.3, 6.1, A

Raymond, W., J. Mairesse, P. Mohnen, and F. Palm (2015). Dynamic models of R&D, innovation and productivity: Panel data evidence for Dutch and French manufacturing. *European Economic Review 78*, 285–306. 1

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica: Journal of the Econometric Society*, 999–1033. 4.3

Song, Z., K. Storesletten, and F. Zilibotti (2011). Growing like China. *The American Economic Review 101*(1), 196–233. 3.1, 4.3